Analyzing Scaffolding Needs for Industrial Construction Sites Using Historical Data

Lingzi WU¹, Yasser MOHAMED¹, Hosein TAGHADDOS², and Rick HERMANN²

¹ Department of Civil & Environmental Engineering, University of Alberta, 5-080 Markin CNRL Natural Resources Engineering Facility, Edmonton, AB, T6G 2W2; PH (780) 492-8096; emails: lingzi1@ualberta.ca, yaly@ualberta.ca
² PCL Industrial, 5402-99 Street, Edmonton, AB, T6E 3P4; PH (780) 733-5500; emails: HTaghaddos@pcl.com, RHHermann@pcl.com

ABSTRACT

Industrial construction includes a wide range of construction projects such as petroleum refineries and chemical plants, which involve several trades such as civil, mechanical, and electrical. Different trades carry out different tasks on these projects, and often depend on scaffolds to access their work areas. Quantification of scaffold requirements of large projects is difficult due to variability in work area heights and congestion, and the multiple trades that need to be serviced by the scaffold system. Traditional estimating methods rely on percentages of direct trade hours or volume of work area and usually result in significant deviation from real scaffold costs. The study presented in this paper aims to develop better understanding and estimates of scaffold needs for industrial construction sites, based on analysis of data collected from a mega project over the course of two and a half years by a major contractor. The study seeks to discover patterns and reliable correlations that may exist between required scaffold hours and other work attributes that can allow for development of a reliable estimation model. The paper presents the results of initial analysis and exploration of data mining experiments, in addition to the challenges faced, and future research recommendations.

INTRODUCTION

Industrial construction projects are usually large-scale, involve complex structures and utilize a large number of workers from different trades. On site, different trades might have similar or completely different requirements for the scaffolding that provides access to their working areas. These different requirements contribute to large scaffolding systems that contribute significantly to project cost. Though indispensable, scaffolding systems often receive little attention, and are neither plotted on drawings, nor managed into work packages that can be integrated into the project schedule. Estimation of scaffolding work is treated as indirect work, and is usually calculated as a percentage of the total man-hours of direct work, and can be as high as 30%–40%. Scaffolding has potential for significant improvement in productivity and cost, if estimated, scheduled, and managed well.

Current scaffolding practice is mostly carried out ad-hoc, which depends on expert experience, historical data, and a company’s risk policy. For estimation, an envelope volume calculated by applying a certain percentage upon direct work is commonly used. Using this method, it is hard to be accurate and specific for different trades or different types of work. Unfortunately, no systematic or sophisticated
methods have been established to tackle this issue in the practical world, nor has academic research addressed this problem so far.

Scaffolds are usually treated as a temporary work estimate that is part of indirect costs. Temporary works play crucial roles in a project’s safety, profitability, and quality (Ratay, 1987). Sometimes, temporary works account for more than 60% of the total project cost (Illingworth, 1987), and are therefore of great importance to the success of a project. From the project planning and scheduling perspective, it is necessary and urgent for the project planner to properly estimate and plan temporary scaffolding work. However, it is common that temporary work costs are not properly tracked (Carr, 1989).

OBJECTIVES AND METHODOLOGY

This paper documents progress made towards building a mathematical model or group of models for scaffold estimation and planning purposes. Based on these models, an estimator can find the scaffold man-hours needed for future projects, for the whole project, or detailed down to each trade in each construction area. Construction areas were divided at the engineering design phase by the engineer, according to type of structures, their functions, and location. Figure 1 below graphically illustrates the inputs and outputs of the target models. The major technique used in this research is data mining. A set of historical scaffold data from a mega industrial construction project is explored to train the scaffold estimate model.

![Figure 1. A graphical representation of scaffold estimate tool](image)

The objective of this study is pursued through the following steps: (1) understand basic scaffold knowledge and current practices of scaffold estimation and planning; (2) understand, examine, and prepare the scaffold data for data mining use; (3) design a set of experiments to test different learning algorithms, different parameter settings of each learning algorithm, different input tables, and different subsets of attributes; (4) evaluate the performance of the models generated from these experiments; and (5) optimize the best performing model(s) to be used for future project use. In this paper, the first two steps are elaborately explained; the third and fourth steps are partially stated here, as far as the progress has been achieved.
Business Model

In order to achieve a better understanding of the scaffold request process on site, and how the scaffold request database was built, a series of interviews were held with an experienced project manager, scaffold foreman, and scaffold coordinator, who built and maintain the scaffold request database, and a site visit, led by the scaffold foreman of a module assembly yard, was made.

According to all the information gathered through those interviews, the site visit, and meetings, a business model using Integrated Definition (IDEF0) diagram was created. Within this business model, the most central part is the scaffold request database, and all the other processes directly or indirectly interact with this database. The main activities involved in the business model include: foreman submits scaffold request; superintendent meeting; check and allocate scaffold request (scaffold coordinator checks scaffold requests; scaffold superintendent checks existing scaffold, approve and prioritize scaffold requests; scaffold foreman site visit and allocation); erect/modify approved scaffold request; use of scaffold by trade and weekly inspection; dismantle scaffold.

The scaffold erection cycle starts with the foreman talking with his/her crew, and checking the construction schedule and drawings for scaffolding requirements (usually a few days or weeks in advance). Then he/she documents the information into a foreman’s log. Based on the information in the foreman’s log, as well as his/her experience, the foreman submits the scaffolding request. All the scaffolding requests will be stored in a central database, called scaffold request database.

At the same time, trade foremen from all trades, the scaffold superintendent, the scaffold foreman and the scaffold coordinator attend a superintendent meeting every morning. In terms of scaffold planning, this is a high-level meeting, dealing with material availability, labor availability, and setting a general priority, which will be the high-level control of the entire scaffold planning and scheduling.

A major process after the requests have been submitted is to check and allocate all the scaffold requests. This process is subdivided into the following three steps:

1. Scaffold coordinators input all the requests into the scaffold request database, and then sort them by required date and check for duplication.
2. Scaffold superintendent gets the request list for a specific date (present or next day; scaffolding planning is short-term). Then the superintendent checks each request, considering ground condition, site density, area platform, existing scaffolding, etc. He/she makes sure that the request is necessary and confirms that there is enough material, and that the schedule requirement is met. Depending on the overall situation, some requests become modifications on existing scaffolds, some have to be erected from scratch, and others might be cancelled. Then, all the approved requests will be prioritized based on the current schedule. While checking scaffold requests, the scaffold superintendent checks existing scaffolding, and considers the overall situation as well, to decide whether dismantlement of some used scaffold is needed or not. All this information will be updated in the scaffold request database immediately.
3. Next, the scaffold foreman goes through the approved scaffolding requests and assigns a scaffolding crew to each request. The approved scaffolding requests will be used to update the scaffolding database.

After scaffold crews get the approved scaffolding request from the scaffolding database, they erect scaffolding according to the requirement. After completing the scaffold, a safety inspection will be done; if the scaffold is safe, it gets tagged. The erection information will be updated in the scaffolding database. Next, the tagged scaffolds will be used by the originally requesting trades and be inspected for safety weekly. Scaffold crews also get dismantle requirements for certain existing scaffolds; they tear down this scaffolding, and release the material. Used scaffolds become a site condition and will be considered when the scaffold superintendent checks the site.

Data Preparation

The original data set obtained from the scaffold request database included some inconsistency and errors. First, effort was made to understand the database structure, and remove the outliers. Corrections were made under the supervision of the scaffold coordinator who collected the data to address omissions and mistakes. Then, according to understanding of the database, as well as consideration of the need for the following computer learning experiments, certain changes were made to convert the database into a more organized format for data mining investigation purposes. For instance, columns were added into the main data table to add more consistency and organization to the table.

Next, a series of general analysis was done to explore different aspects of the data set. Consequently, a better understanding, and an overall idea of this set of data was gained. For example, a set of charts plot the scaffold request count and sum of scaffold man-hours on lead time between “Required Date” and “Request Date.” This series indicates the running status of this scaffold request system. These charts are made based on day-shift, night-shift, and both day-shift and night-shift. An example is shown in Figure 2. Similarly, a set of charts are built to show scaffold request count and sum of scaffold man-hours on time difference between “Required Date” and “Completion Date,” which reflect the accomplishment information in terms of time.

![Figure 2. Scaffold man-hours and count of lead time between required date and request date of both day shift and night shift](image)

1 All numbers related to man-hours throughout the paper scaled for confidentiality reasons
A group of charts and tables was organized to show scaffold features on different construction areas, and furthermore, different trades on each construction area. For instance, Figure 3 below shows scaffold request count and sum of scaffold man-hours in different construction areas; Figure 4 is a box plot of quotient of scaffold man-hours over direct trade man-hours. These charts and tables directly indicate that different trades have different requirements for scaffolds, and different construction areas contain different features, showing different scaffold needs. This conclusion is in agreement with the scaffold foreman, scaffold coordinator, and project manager’s experience.

![Figure 3. Scaffold man-hours and count in each construction area](image1)

![Figure 4. Box plot of scaffold proportion to direct trade man-hours on each trade – excluding three outliers](image2)

Another data preparation step involved connecting scaffold request data with as built schedule and payroll data to add more details about actual work and costs associated with different requests. Both sources have records for scaffold man-hours; thus, a comparison between them and the scaffold request database was made to verify the reliability of information sources. In the schedule, a bulk number shows the total scaffold man-hours. In payroll, the sum of all workers’ man-hours for scaffolding is calculated. These two numbers are very close (schedule is 2.5% higher...
than payroll). However, a certain gap between them and the sum of scaffold man-hours from the scaffold request database was discovered.

Figure 5 shows a specific comparison that was made: first, in comparing the same time period, the scaffold database missed the starting point of this project, which is from August 2007 to August 2008; second, it excluded the scaffold foremen’s and scaffold superintendents’ man-hours, which were not recorded in the scaffold database. Still, almost a 30% difference of total scaffold man-hours was found between payroll and the scaffold request database. After some interviews and meetings, the reasons were identified: (1) some of the preparation time is not considered in the scaffold request database, which sometimes adds a considerable amount of time. This preparation time includes material handling time – loading, unloading; cleaning up the site; moving equipment; etc.; (2) some minor modifications, or some urgent scaffold requests, may not be updated in the request database; and (3) some human errors happened in the scaffold request database.

![Figure 5. Man-hours based on payroll sheet and scaffold request database](image)

Input Table

Selection of input, learning method—algorithm, and output is vital for this step. Due to the fast track feature of industrial construction projects, information at early stages of a project is incomplete and limited. The information available for this scaffold estimation tool will be rough, approximate, and high level. Given this reality, the information level of the existing scaffold database is too detailed, which wouldn’t train the model as needed. Coordinating with the schedule, the database was organized and aggregated to a higher level. Descriptive statistics of attribute values of a group of detailed request records were used as attributes for higher level group records. Thus, information of direct man-hours for each trade from the schedule was integrated into the input. Although efforts were made to try to gain information from the payroll sheet, the design and structure of it failed to provide useful information on the same level.

As a result of preparation of the data, a general input table – Trade_Area was built. A set of assumptions were made. (1) All scaffold information including height and volume came from day-shift records, because night-shift records in the database used a different ID number system, which cannot track construction areas and some other information needed to build this table. (2) “Trade Mhrs,” and “Trade Mhr Distribution” are from the as built schedule of the project (not from payroll, because
it was impossible to separate records in payroll into work packages and work areas).

(3) Winter efficiency decline is offset by use quotient of scaffold man-hours by direct trade man-hours as class.

The “Trade_Area” table contains 20 columns, as show in the following figure, which shows screen shots of the input table; due to the large size of the table, the table is broken into three parts.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction Areas</td>
<td>Trades</td>
<td>Area</td>
<td>General</td>
<td>Elevation</td>
<td>Scaffold Type</td>
</tr>
<tr>
<td>A Charge Pumps</td>
<td>CIV</td>
<td>Pumps</td>
<td>Access</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Charge Pumps</td>
<td>EL</td>
<td>Pumps</td>
<td>Access</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Charge Pumps</td>
<td>FP</td>
<td>Pumps</td>
<td>Access</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Charge Pumps</td>
<td>INSTR</td>
<td>Pumps</td>
<td>Access</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Charge Pumps</td>
<td>INSUL</td>
<td>Pumps</td>
<td>Access</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>H_Mean</td>
<td>H_StDev</td>
<td>H_Max</td>
<td>H_Min</td>
<td>H_Mode</td>
<td>V_Mean</td>
<td>V_StDev</td>
<td>V_Max</td>
<td>V_Min</td>
<td>V_Mode</td>
</tr>
<tr>
<td>84.81139</td>
<td>6.080435</td>
<td>88.56</td>
<td>6.75</td>
<td>84.78</td>
<td>4.199259</td>
<td>7.329831</td>
<td>91.935</td>
<td>0.135</td>
<td>1.62</td>
</tr>
<tr>
<td>84.81139</td>
<td>6.080435</td>
<td>88.56</td>
<td>6.75</td>
<td>84.78</td>
<td>4.199259</td>
<td>7.329831</td>
<td>91.935</td>
<td>0.135</td>
<td>1.62</td>
</tr>
<tr>
<td>84.81139</td>
<td>6.080435</td>
<td>88.56</td>
<td>6.75</td>
<td>84.78</td>
<td>4.199259</td>
<td>7.329831</td>
<td>91.935</td>
<td>0.135</td>
<td>1.62</td>
</tr>
<tr>
<td>84.81139</td>
<td>6.080435</td>
<td>88.56</td>
<td>6.75</td>
<td>84.78</td>
<td>4.199259</td>
<td>7.329831</td>
<td>91.935</td>
<td>0.135</td>
<td>1.62</td>
</tr>
<tr>
<td>84.81139</td>
<td>6.080435</td>
<td>88.56</td>
<td>6.75</td>
<td>84.78</td>
<td>4.199259</td>
<td>7.329831</td>
<td>91.935</td>
<td>0.135</td>
<td>1.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q</th>
<th>R</th>
<th>S</th>
<th>T</th>
<th>U</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Scaffold Mhrs</td>
<td>Trade Mhrs</td>
<td>Scaffold Distribution</td>
<td>Trade Mhr Distribution</td>
<td>Scaffold Mhrs/Trade Mhrs</td>
</tr>
<tr>
<td>0.135</td>
<td>15.12</td>
<td>6131.751717</td>
<td>0.000562929</td>
<td>0.021505961</td>
<td>0.0033289</td>
</tr>
<tr>
<td>24.84</td>
<td>1059.615</td>
<td>10545.37269</td>
<td>0.039450288</td>
<td>0.036985903</td>
<td>0.01358504</td>
</tr>
<tr>
<td>4.725</td>
<td>293.6925</td>
<td>1369.871391</td>
<td>0.010934399</td>
<td>0.004804565</td>
<td>0.01356500</td>
</tr>
<tr>
<td>1.89</td>
<td>38.4075</td>
<td>3700.107236</td>
<td>0.001429941</td>
<td>0.012977427</td>
<td>0.00140131</td>
</tr>
<tr>
<td>6.075</td>
<td>160.8525</td>
<td>918.4421496</td>
<td>0.005988663</td>
<td>0.003221262</td>
<td>0.023643392</td>
</tr>
</tbody>
</table>

**Figure 6. Screen shots of Area_Trade input table**

Column A shows which construction area each instance belongs to; Column B shows what trade this record marks. There are eighteen construction areas, and eight trades, which are CIV – Civil, EL – Electrical, FP – Fireproof, INSTR – Instrumentation, INSUL – Insulation, IW – Structural Steel, and PF – Pipe Fitting.

Columns C–F show the general features of each construction area; Column C shows the description of area size; Column D shows the general information; Column E shows the general height description; Column F shows the overall or most possible scaffold type. This is additional information added to the original database, trying to imitate the real information available at the starting point of a project.

Columns G–K show elevation information; these include mean, standard deviation, maximum, minimum, and mode of elevation, respectively, for each record in this table. This information was generated from records in the scaffold database, which represents the elevation distribution of each trade in each construction area. The unit for elevation is meters. Columns L to P show volume information; these include mean, standard deviation, maximum, minimum, and mode of elevation, respectively, for each record in this table. The information is generated the same as for elevation. The unit for volume is m³.

Columns Q–S show scaffold information; scaffold information is collected from the database; number of scaffold requests, total man-hours of the scaffold
requests, and the percentage of scaffold man-hours of each trade in the construction area to the total scaffold man-hours of all trades in the construction area are calculated.

Column T provides trade information for each trade in each construction area. Column U shows the percentage of the direct man-hours of each trade in the construction area to the total direct trade man-hours of all trades in the construction area; the way it was calculated and its meaning are quite similar to Column S.

The last column – Column V is considered to be the class of this relation. For each instance, this is the quotient of Column R divided by Column T.

**Result of Initial Analysis**

The computer learning of the data set was performed using an open source data mining tool called WEKA, which has a collection of machine learning algorithms and data preprocessing tools. (Witten, Frank, & Hall, 2011). The scenario here is trying to exhaust all the possible combinations of four parameters: one is possible learning algorithms, one is settings of each possible learning algorithm, one is different input tables, and the last one is subset of attributes of the selected input table. To explore every possible combination is impossible if it is carried out manually. Thus, a systematic way of trying different possible combinations was planned.

As shown in the Trade_Area table, the class of this research is a quotient of “Scaffold Mhrs” by “Trade Mhrs,” which is numerical. When class is numerical, and most of the attributes are numeric as well, linear regression is a natural technique to try first. Linear Model is simple, concise, and frank, and shows good performance on simple problems. Thus, it was chosen here as the major algorithm. Since WEKA provides more sophisticated algorithms to treat numerical situations, Gaussian Process (M. Ebden, 2008) was tried in this research as a comparison.

All the experiments have been done using the ten-fold cross validation test (Witten, Frank, & Hall, 2011), trying to find the best estimate of error. The table below shows some results of the completed experiments. From the results of these experiments, it is fair to say that simple Linear Model solves this problem well. However, a certain number of random experiments have to be run before any judgement can be made.

In order to improve model accuracy we tried to change the class (the output attribute) from the ratio of scaffold Mhrs/Trade mhrs to directly estimating scaffold mhrs. In addition, a decision tree (M5P tree) has been used with linear regression model; this means we will have many linear regression models, and depending on the tree branching and the attributes values, the suitable model will be used. The experimental data indicates that models that directly predict scaffold man-hours have better performance than the models that predict a ration of scaffold man-hours to direct trade man-hours. Figure 7 and 8 shows the developed model using M5P tree.

A simple linear regression model that directly estimated scaffold mhrs was built in order to compare it to the decision tree model. Two conclusions can be drawn from the comparison, first, no matter if the model is trained by linear regression or decision tree, the correlation coefficient increased by directly estimating scaffold
mhrs; second, the error evaluation decreased from the models trained by M5P trees, while the error stays at the same level from the models trained by linear regression.

**Table 1. Selected Performance Comparison of Initial Data Mining Models**

<table>
<thead>
<tr>
<th>Input Attributes</th>
<th>Algorithm</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Correlation Coefficient</td>
</tr>
<tr>
<td>11 attributes</td>
<td>Linear Reg. 0.7396</td>
<td>0.0804 0.1109 66.1%</td>
</tr>
<tr>
<td>9 attributes</td>
<td>Linear Reg. 0.7405</td>
<td>0.0796 0.1108 65.4%</td>
</tr>
<tr>
<td>6 attributes</td>
<td>Linear Reg. 0.7457</td>
<td>0.0793 0.1098 65.2%</td>
</tr>
<tr>
<td>11 attributes</td>
<td>Gaussian Process 0.7614</td>
<td>0.0761 0.1082 62.6%</td>
</tr>
</tbody>
</table>

**Figure 7. Structure of M5P tree model**

**CONCLUSION AND FUTURE WORK**

This research aimed at building a model for estimation of scaffold needs for industrial construction projects. The process of data collection, exploration and preparation has been discussed in the paper, in addition to some results of data mining experiments. Initial results show significant influence of trade type and construction area features on the scaffold needs in that area. Model training results show close performance of linear models and more complex models. Further analysis showed improvement with a regression model tree and smaller number of attributes. Future work will include other learning algorithms, performance comparison of the models resulting from these algorithms, and optimizing the best performers for day-to-day
use. Inclusion of data from other projects and contractors is also required, as current results cannot be generalized beyond the practice of the collaborating contractor.

```
Trade Mhrs <= 41975.081:
|   Trades=FP,INSUL,EL,PF <= 0.5:
|     |   Trade Mhrs <= 849.282 : LM1 (22/2.072%)  
|     |   Trade Mhrs >  849.282 : LM2 (25/5.807%)  
|   Trades=FP,INSUL,EL,PF >  0.5 : LM3 (44/12.386%)
Trade Mhrs >  41975.081:
|   Trade Mhrs <= 10146.394 : LM4 (18/17.235%)
|   Trade Mhrs >  10146.394 : LM5 (5/82.77%)
LM num: 1

Scaffold Mhrs =
- 53.7008 * Trades=FP,INSUL,EL,PF + 31.8098 * Trades=PF
  + 101.3334 * Area_Size=Medium,Large
  - 138.9984 * Area_Congestion Degree=Less Congested + 0.0421 * Trade Mhrs
  - 241.0761 * Trade Mhr Distribution + 221.8287
LM num: 2

Scaffold Mhrs =
- 53.7008 * Trades=FP,INSUL,EL,PF + 31.8098 * Trades=PF
  + 101.3334 * Area_Size=Medium,Large
  - 138.9984 * Area_Congestion Degree=Less Congested
  + 0.0418 * Trade Mhrs - 397.0958 * Trade Mhr Distribution
  - 18.9289 * Winter Factor + 360.5282
```

Figure 8. The first two regression models in the decision tree

ACKNOWLEDGEMENT
This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) research grant (CRDPJ 416362-11) in collaboration with PCL Industrial Management Constructors Inc. Edmonton, Alberta, Canada.

REFERENCES