www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

A SIMULATION-BASED STATISTICAL METHOD FOR PLANNING MODULAR CONSTRUCTION MANUFACTURING

SUBMITTED: May 2021
REVISED: September 2021
PUBLISHED: February 2022
EDITOR: Esther Obonyo
DOI: 10.36680/j.itcon.2022.007

Angat Pal Singh Bhatia, Ph.D. Student
Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada
angatpalsingh.bhatia@concordia.ca

SangHyeok Han, Associate Professor
Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM),
Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada
sanghyeok.han@concordia.ca

Osama Moselhi, Professor
Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM),
Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada
osama.moselhi@concordia.ca

SUMMARY: *Modular construction is a promising alternative to conventional construction; offering improved productivity, quality, and safety. To realize these benefits, sequencing the module fabrication process in a manner that ensures efficient allocation of labor resources is essential. However, the varying sizes and design specifications of modules lead to high variation in process times at workstations, ineffective utilization of resources, and imbalanced production lines. To address these challenges, this paper proposes a simulation-based statistical method to plan the sequencing of module fabrication and the allocation of workers at workstations for such that productivity and control are improved. The method consists of four processes: (i) data collection to obtain historical and near real-time data; (ii) identification of significant impact factors affecting process times at workstations along the production line; (iii) development of a predictive model for forecasting process times at workstations using statistical analysis and probability distribution function; and (iv) planning the sequencing of module fabrication in a manner that ensures efficient labor allocation (i.e., crew size). The developed method is validated using data captured from a light gauge steel wall panel production line operated by a modular fabricator in Edmonton, Canada. The industrial partner produces both interior and exterior light gauge steel wall panels on a production line consisting of multiple workstations. First, five significant impact factors for each workstation among the design factors that highly influence the process times were identified in order to develop cycle time formula as a predictive model. The simulation model developed and implemented in conjunction with cycle time formula (CTF) in this case study was deemed to be a reliable predictive model (i.e. 89.39% accuracy), which can be used to improve productivity. The method is shown to be capable of assisting in decision-making by enabling production managers to better understand the effects of proposed changes to the production line prior to implementation. In this way, production managers can plan effectively and thereby reducing non-productive idle time.*

KEYWORDS: *Modular construction manufacturing, Planning, Productivity, Statistical analysis, Probability distribution, Simulation*

REFERENCE: *Angat Pal Singh Bhatia, SangHyeok Han, Osama Moselhi (2022). A simulation-based statistical method for planning modular construction manufacturing. Journal of Information Technology in Construction (ITcon), Vol. 27, pg. 130-144, DOI: 10.36680/j.itcon.2022.007*

1. INTRODUCTION

Modular construction has grown in popularity in recent years, given its potential to enhance productivity. According to the Modular Building Institute (MBI, 2018), modular construction can be used to construct condominiums, dormitories, and duplex homes in about half the time of conventional construction. Previous research (Bertram et al., 2019) has defined modular construction in terms of the following features: (i) repetitive nature of manufacturing activities; (ii) standardized operating procedures (SOPs) on a production line; and (iii) a streamlined construction process due to simultaneous on-site and off-site activities. Based on these features, MBI reported an increase in gross revenues from $3.3 billion to $3.97 billion in 2017 (MBI, 2018). However, modular construction manufacturing (MCM) still has various challenges, such as: (i) variation in design specifications of modules, leading to varying production rates and imbalanced production lines; and (ii) queuing and waiting due to imbalanced production, in turn further increasing the idle time of workers at workstations (Hsu et al., 2019). To overcome these challenges, robust production planning that encompasses careful sequencing of fabrication processes and efficient allocation of workers at workstations is critical in efforts to improve productivity and control.

Effective planning plays an essential role in construction projects, as it helps to better allocate workers (minimizes their idle time), leading to timely and efficient delivery of projects. However, the planning methods used in conventional construction projects, such as the critical path method, are not suitable for MCM due to the lack of crew work continuity leading to low productivity (Kavanagh, 1985). Consideration of crew work continuity helps to prevent idle time during repetitive tasks by planning the advancing from one task to another in accordance with the task demands (El-Rayes and Moselhi, 1998). In this respect, the linear scheduling method (LSM) addresses resource continuity in repetitive projects. This method provides information about productivity and duration of activities in an easy to understand format. For example, Salama et al. (2018) integrated critical chain project management with LSM to accelerate schedules while maintaining resource continuity. However, this method does not (i) account for decision variables such as module dimensions and wall openings in their sequencing and labor allocation, which highly affects the cycle time of the production line or (ii) consider multiple scenarios of resources and sequences of modules which is critical for planning effectively and improving production performance of MCM lines (Altaf et al., 2014; Liu et al., 2015).

To address the above limitations, this paper proposes a simulation-based statistical method for production line planning that supports production planning and labor allocation based on historical and near real-time data. The developed method encompasses: (i) a C-track application to collect historical and near real-time data from workstations; (ii) identification of significant impact factors (SIFs) influencing fabrication process times at workstations; (iii) development of predictive models for forecasting process times at workstations; and (iv) development of a simulation model facilitating the generation of multiple scenarios based on sequences of modules and allocation of workers at workstations.

2. BACKGROUND

To improve planning in modular construction, researchers have proposed planning methods based on various techniques such as lean principles and simulation (Yu et al., 2009). The implementation of lean principles enhances substantially process improvement efforts by identifying and eliminating non-value-added activities in the production line. For example, Yu et al. (2009) used a value stream mapping (VSM) tool to determine steady production flow for productivity improvement by analyzing the production process and controlling the fluctuation of resources. Zhang (2017) integrated the production line breakdown structure with VSM to assess the status of the production line, identify issues, and propose solutions for future implementation. However, based on these techniques, designing plans without validating and understating their effects on the production line can be costly and time-consuming. To address this, computer simulation can be employed as a validation tool for future planning by imitating production line processes as a way of assessing the effect of proposed solutions prior to incurring the cost and disruption of actual implementation (Han et al., 2012).

Computer simulation is commonly used for planning and decision-making in the construction industry, and several simulation tools have been developed. Cyclone (Halpin, 1997) can be deployed to model and analyze the construction process. Hajjar and AbouRizk (1999) introduced Simphony.NET as a discrete-event simulation system to evaluate production line scenarios before implementing them in actual production. Other studies have advanced production line planning in MCM in particular using simulation platforms. For instance, Altaf et al.

(2018) used radio frequency identification (RFID) technology and simulation-based optimization for planning and controlling the production line. Bhatia et al. (2019) created a simulation model in Simphony.NET for predicting the productivity of the production line. Azimi et al. (2011) proposed a data acquisition and simulation-based framework as a decision support tool on the basis of which production managers can take proactive corrective actions. However, what these methods lack is: (i) a systematic way to identify the SIFs affecting process times at workstations, which provides critical information that can improve the accuracy of the simulation results; (ii) analysis of the allocation of workers to workstations for the purpose of productivity improvement; and (iii) a data collection system that offers ease of use and cost efficiency without sacrificing efficiency and accuracy. Due to these deficiencies, previous studies have relied on a manual data collection approach that is error-prone and time-consuming. Although some studies have introduced RFID-based data collection (Azimi et al., 2011; Altaf et al., 2018), this entails upfront and operational costs for the RFID infrastructure that the small- and medium-sized construction enterprises—which still dominate the Canadian construction market (KSBS, 2019) cannot afford. Moreover, depending on the material type (e.g., light gauge steel), some modular construction companies may prefer not to apply this advanced technology due to the interference of the RFID tag detection caused by the material.

Accurate prediction of workstation process times based on a module's design specifications is essential for production planning in MCM. In this respect, identifying the key features affecting the process times and developing regression-type predictive models that use historical data is vital to support intelligent decision making in production planning that promote productivity improvement. Feature selection methods in data analysis are used to select important features from an available subset of variables in order to develop efficient and accurate predictive models (Guyon and Elisseeff, 2003). According to Mohsenijam and Lu (2019), a predictive model with key input features significantly reduces the collinearity between input variables and overfitting issues. Various feature selection methods, such as correlation matrix, principal component analysis, and *t*-test, have been used to identify key features (e.g., work durations, worker's skill, profit margin, module type and workplace design parameters) in order to improve understanding of the underlying processes and overall performance of projects (Xu et al., 2017; Xie et al., 2018; Zaalouk and Han, 2021). For instance, Chanmeka et al. (2012) carried out a correlation analysis and statistical test of significance to determine critical factors related to the performance of oil & gas projects in Alberta, Canada.

Moreover, multiple linear regression has been implemented to develop data-driven prediction metrics in various construction and fabrication projects. This approach relies on several input variables to predict an output variable as a way of gaining insights into underlying patterns. For example, Mohsenijam and Lu (2019) used multiple linear regression techniques to develop a predictive model for structural steel fabrication. In their study, stepwise regression was implemented to identify key design features such as rebar, bolts, nuts, etc. The project's labor hours were predicted using a regression model created using these smaller subsets of design features. Despite the appropriate use of feature selection methods and regression-type predictive models in the aforementioned studies, though, they cannot be effectively applied to MCM because of its process-oriented nature. In the MCM approach, the entire production process is divided into sequences of smaller repetitive processes, such that the productivity of the manufacturing production line can be improved using lean concepts such as elimination of wastes and continuous improvement. As such, in the case of MCM, an analytical framework is required in order to identify significant module design factors and establish a model to accurately predict the workstation process times. In consideration of these characteristics, the research presented in this paper seeks to identify which feature selection methods are most effective for production planning in MCM.

3. DEVELOPED METHOD

Fig. 1 depicts the components of the developed method and its simulation-based planning process for MCM lines using historical and near real-time data. The process encompasses three phases: (i) data collection; (ii) data analysis; and (iii) simulation-based planning. Input parameters such as process times of workstations, design specifications of modules, and the number of workers at workstations are housed in a central database. The criteria are workflow of the production line, availability of resources, working hours, and the capacity of workstations in terms of the module length they can accommodate (e.g., the capacity of a framing workstation may be a length of 20 ft). In the data collection phase, work and time studies are performed in order to gain understanding of the SOPs at workstations as well as collect historical and near real-time workstation production data. The data collected includes the start and finish times of modules, design specifications, and the number of workers assigned to various workstations. This data is stored in a database via a cloud-based time-track application called "C-track". The database is used in: (i) the data analysis phase to identify SIFs, develop the probability distribution functions and cycle time formula (CTF) using statistical techniques, and select the best predictive model by comparing the performance of the cycle time formula with that of the probability distribution functions; and (ii) the simulation model of the production line developed in Simphony.NET. Both cycle time formula (CTF) and probability distribution functions are used in the simulation model as a way of capturing the unique nature of production lines (e.g., facility layout and number of workstations). In this respect, the outputs (productivity) from the simulation model using cycle time formula (CTF) and probability distribution functions are compared with historical productivity data in order to determine which model is most accurate. The output, it should be noted, lists the sequence of modules and allocation of workers at workstations, and can be used to determine the total duration for completing a module in the multiple scenarios generated by the simulation model. The simulation model also helps to balance the production line by assigning an optimal sequence of modules and allocation of workers.

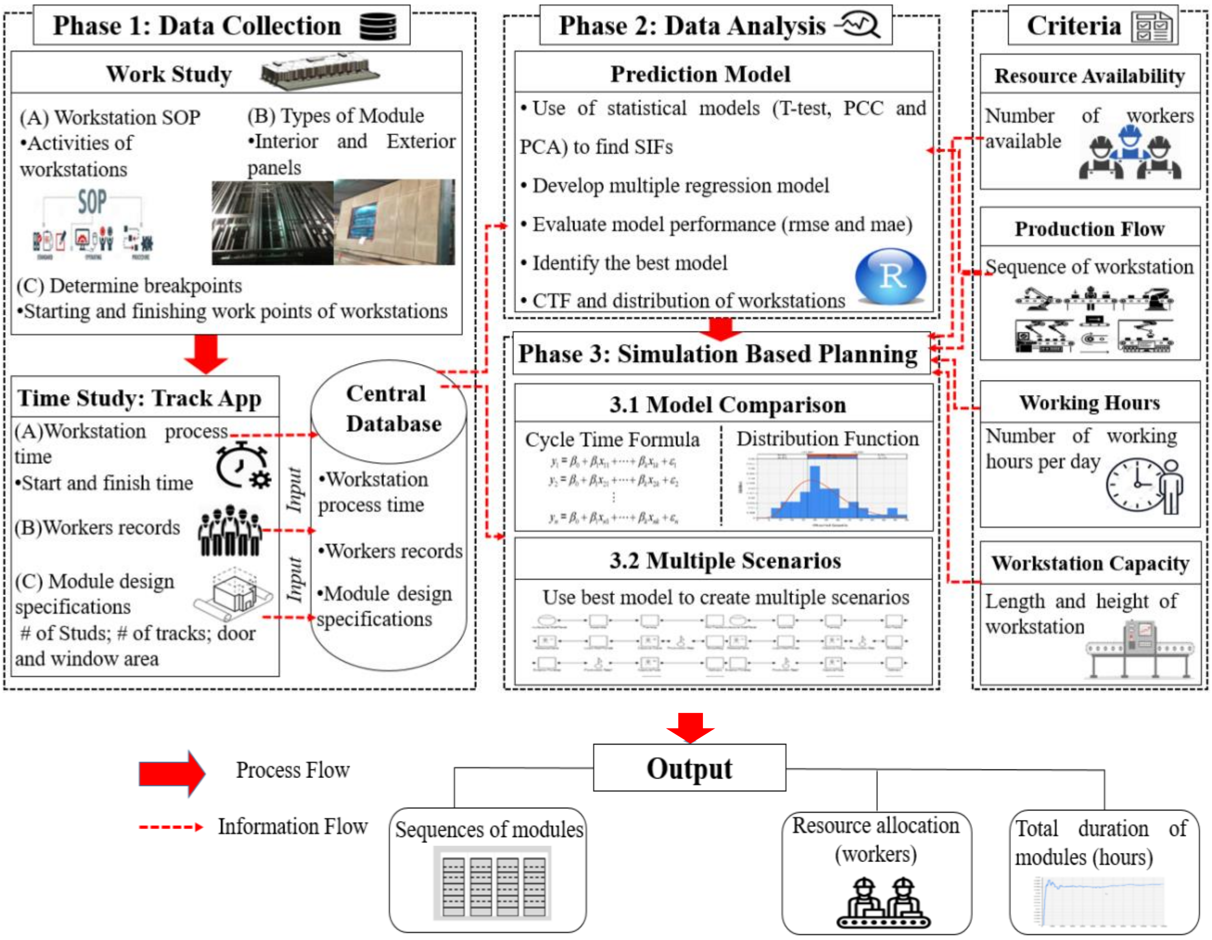

FIG. 1: Main components of the developed method.

3.1 Data collection

The purpose of the data collection phase is to understand the performance of production lines by conducting work and time studies. The work study involves: (i) reviewing the SOPs of workstations in order to gain a high-level understanding of the sequences of activities at workstations prior to direct observation; (ii) classifying the types of modules (e.g., interior and exterior wall panels) and their components (e.g., number of window openings per module) using the shop drawings; and (iii) determining the breakpoints, i.e., the start and finish points for the various work processes at the workstations, in order to ensure that the process times are collected efficiently and accurately. In a typical time study, a series of time data in the production line is collected and recorded manually using a stopwatch and timesheets, but this approach is time-consuming and inaccurate. Therefore, to achieve more accurate results in a more efficient manner, a C-track app is used in the present study in order to: (i) improve the accuracy of the data collection process by recording near real-time data of modules at workstations; and (ii) improve communication by providing efficient information transfer between the production line and the production planning department. The C-track app consists of a desktop-based production management system (used in the production planning department) and an iPad-based system (installed at the workstations). The key users in the desktop-based system are the project manager and the project coordinator. The desktop-based system encompasses: (i) representation of the production line for the purpose of defining the workflow; (ii) management of the assignment of modules to workstations (i.e., sequence); and (iii) data organization (e.g., productivity of specific workstations and of the production line overall) based on the historical and near real-time recorded by the iPad-based system. The iPad-based system receives the information related to modules and their workflow from the desktop-based system. The workers at workstations follow the information concerning sequencing of modules assigned under the list of module names in order to record the process times. For collecting the time records, the 'Track App' features start, pause, and finish buttons. The worker first selects a 'module name' from the list, then presses 'start' to record the process start-time. In case of a disruption due to an error in the drawings or a work stoppage for a scheduled break, the worker uses the 'pause' button to stop the time record. After completing a module, the worker selects the 'finish' button. The timestamp is then recorded and transferred automatically in the 'production management system' and the database. Along with this, the 'module name' of the finished module is updated automatically in the 'product item' list of the next workstation. Although workers are still required to manually start/stop the time recordings, this semi-automated data collection application efficiently and accurately tracks modules at workstations and allows the productivity of the production line to be monitored in real-time. Based on the app's data organization feature, production managers can monitor bottlenecks on the production line in near real-time by comparing current and historical production rates.

3.2 Data analysis

Fig. 2 provides a flowchart of the data analysis used to develop the models for predicting process times. To address the effect of missing data and outliers on the accuracy of the predictive models, the first step of the data analysis is to identify any missing values in the dataset due to human error (e.g., incorrect keystrokes on the app.) or technical issues such as temporary internet disconnection. The most common method to deal with missing values in the data is substitution through linear interpolation or regression (Piryonesi and El-Diraby, 2019). Furthermore, outliers, a set of data points that follow a pattern inconsistent with the rest of the data points, must be removed to improve the accuracy of the predictive model. Outliers can be identified and removed either by drawing scatter plots or by using residuals, which measure differences between observed and estimated durations. Residual analysis proceeds with: (i) fitting a regression model to the dataset; (ii) finding the estimated durations; and (iii) calculating the residuals and adjusted residuals. In this research, the standardized residuals (individual residual divided by the standard deviation of residuals) are found to fall within the range of ±1.64, where data points outside of this range are considered outliers, as per Cottrell (2006). After detecting the outliers, normalization is implemented using Eq. (1) to: (i) reduce the sizes of variables and thereby reduce computation time for calculating the process times; and (ii) improve the accuracy of the model.

$$V' = (V - min_A)/(max_A - min_A) \qquad \text{Equation (1)}$$

where min_A and max_A are the minimum and maximum values, respectively, of the independent variable, A, and V represents the original value of A.

The dataset having been cleaned, statistical techniques such as PCA, *t*-test, and PCC are applied in order to identify the SIFs affecting the process times. These feature selection techniques enhance the performance of the predictive model and provide deeper understanding of the underlying process (Guyona and Elisseeff., 2003; Chanmeka et al., 2012). These techniques, it should be noted, are selected for the present study by virtue of (i) their wide use in investigating prediction and variable selection problems in manufacturing and pipe fabrication; and (ii) their simplicity, empirical accuracy, and generic applicability. To avoid overfitting, machine learning algorithms such as random forest and decision tree, as alternative methods for variable selection, are not considered because of the small size of the dataset, i.e., less than 1,000 observations. According to Makridakis et al. (2018), the use of such methods for small datasets can yield a "black box solution", which may not be acceptable to industry practitioners. Below is a brief description of statistical techniques used:

1) *t*-test: The sample *t*-test is a statistical analysis technique used to determine the probability value of rejecting the hypothesis, which, in this research, is the statistical significance of the selected factors that affect the process

times at MCM workstations. In this research, a sample *t*-test (Gerald, 2018) is deployed based on Eq. (2), where *n* is the sample size, $\bar{x}$ is the mean of the sample data, μ is the population mean, and σ is the standard deviation. The SIFs are selected by evaluating the p-value—defined as p-value ≤ 0.01 in this research.

$$t = (\bar{x} - \mu)/(\sigma/\sqrt{n}) \quad \text{Equation (2)}$$

2) Pearson correlation coefficient (PCC): PCC (Yu and Liu, 2003) is represented by Eq. (3), where $\overline{x_i^j}$ is the mean of independent variable x^j and $\overline{y_i}$ is the mean of dependent variable y (duration). x_i^j and y_i represent the original values of variables x_i and y. The range of R is ±1, where values close to 1 are indicative of a strong correlation between the independent variables, while values close to 0 signal a weak correlation between independent variables.

$$R = \sum\left(x_i^j - \overline{x_i^j}\right)(y_i - \overline{y_i}) \,/\, \sqrt{\sum\left(x_i^j - \overline{x_i^j}\right)^2 \sum\left(y_i^j - \overline{y_i^j}\right)^2} \quad \text{Equation (3)}$$

where *i* is the number of observations and *j* is the number of variables.

3) Principal component analysis (PCA): PCA is a widely used data reduction technique that extracts the small set of variables that accounts for maximum variance in the original dataset. In this method, typically the principal components (PCs) accounting for 90% of the dataset's total variation are selected for further analysis (Rocchi et al., 2004). These components represent a linear combination of original independent variables, where the first PC has the most significant variance, and the succeeding PCs are built by reducing the variances of the preceding PCs.

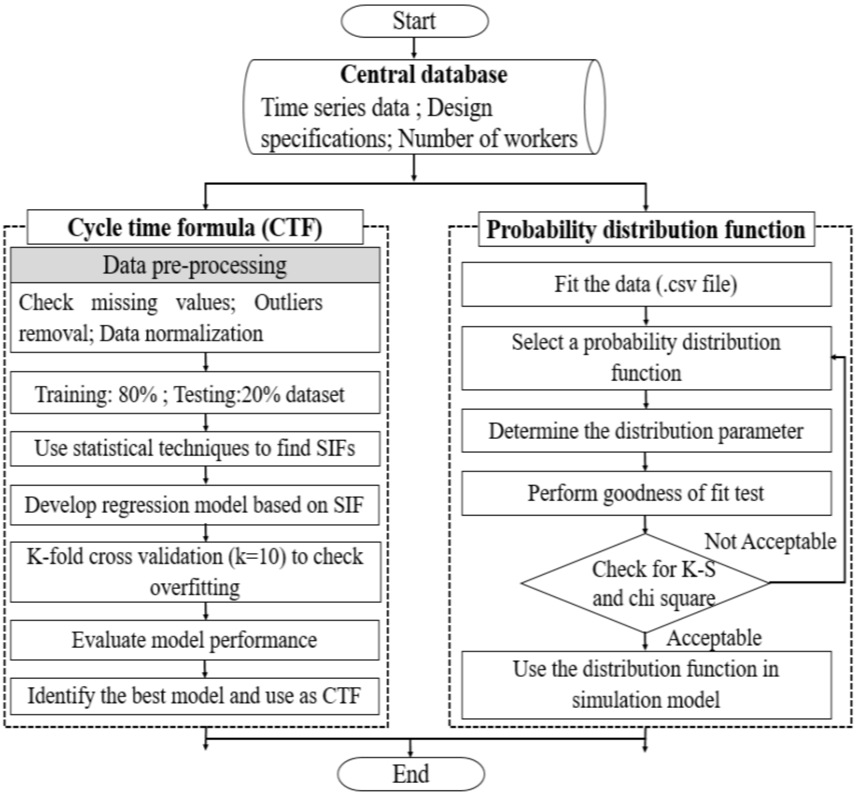

FIG. 2: Flowchart of data analysis.

Before developing the regression model, the dataset is divided into training (80%) and testing (20%) subsets. Based on the SIFs and the training subset, multiple linear regression (MLR) models are developed to predict the process times at workstations. To identify and mitigate overfitting, which diminishes the generalizability and accuracy of predictive models, cross-validation, as a technique that identifies overfitting by testing the accuracy of the predictive model (Mahmood and Khan, 2009), is applied. The present study adopts K-fold cross-validation, as this is one of the most robust approaches for validating and testing the performance of a model, given that it utilizes unseen data. The dataset is divided into K folds, where, in turn, (K−1) folds are used to train the predictive model, and the remaining fold is used to test the accuracy. This process is repeated for K iterations, where the accuracy of the predictive model is determined by calculating the average of these K iterations.

Based on the results of the *t*-test, PCC, and PCA using different lists of SIFs, there are a few different MLR models capable of predicting the process times. There are four evaluation indices that can be used to determine which predictive model is most accurate: (i) R-square (R^2); (ii) adjusted R-square, (adj. R^2); (iii) root-mean-square error (RMSE); and (iv) mean absolute error (MAE) (Willmott and Matsuura 2005, Elmousalami 2019). These evaluation indices have been shown to be reliable and are used widely in various disciplines such as manufacturing and construction (Martinez et al., 2020).

Based on the historical time data stored in the database, a probabilistic model is developed to predict the process times at workstations. Various functions, including Gamma, Weibull, Uniform, and Triangular, are fitted in order to generate the process times based on the parameters of the associated distributions, such as moment matching, maximum likelihood, and least squares. A given function and associated parameters having been determined, goodness-of-fit tests, including Kolmogorov–Smirnov (K–S) and chi-squared, are performed in order to assess the goodness-of-fit between the observed and expected distributions (AbouRizk et al., 2016). If the predicted and actual process times are in close correlation, the given function is accepted for use as an input in the simulation model.

3.3 Simulation-based planning

In practice, different types of modules, such as interior and exterior wall panels, may be produced in the same production line but following different SOPs at the workstations. In this respect, MCM typically follows a mixed-production line model, leading to varying process times and, in turn, imbalanced production flow and inefficient utilization of resources. Given this, it is important to reduce: (i) waiting time of modules between workstations; and (ii) idle time of workers at workstations. In this respect, continuous workflow can be achieved by planning efficient sequencing of modules and labor allocation through predictive modeling based on historical time data. In this regard, in the present study production planning is implemented using a simulation model developed in Simphony.NET. Fig. 3 illustrates the process flow of the simulation-based planning method, which consists of (i) developing a simulation model by mimicking the workflow of the production line based on the work study; (ii) importing inputs such as cycle time formula (CTFs), probability distribution functions, labor allocation, and module design specifications (e.g., heights of framing components and number of frames) from the database into the simulation model using structured query language (SQL); (iii) comparing the results (productivity) between the cycle time formula (CTFs) and probability distribution functions using historical productivity data in order to identify the most accurate predictive model ; (iv) using the most accurate predictive model to develop multiple scenarios in the simulation model by considering various sequences of modules and various labor allocation cases; and (v) selecting the best scenario, that is, the scenario with the minimum total duration to complete the project (i.e., to produce the given modules).

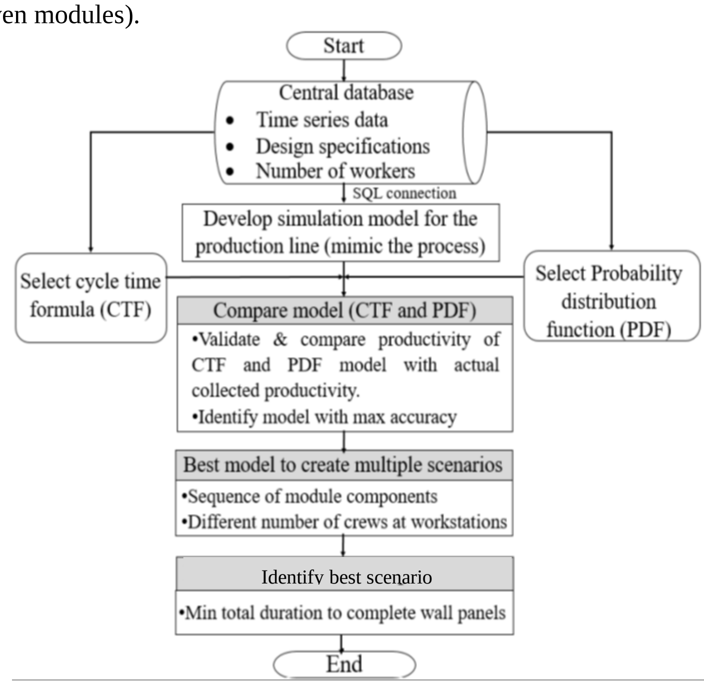

FIG. 3: Flowchart of simulation-based planning.

4. IMPLEMENTATION

The developed method was implemented on a wall panel production line of a modular fabricator in Edmonton, Canada. The industry partner produces both interior and exterior light gauge steel (LGS) wall panels on a production line consisting of five workstations: (i) as assembly station where wall components, such as types of studs and tracks and the number of studs and tracks, are prepared based on the shop drawings; (ii) a framing station where the wall components are fastened together to form wall panels; (iii) a rim track sub-station where track studs are installed on the panels to secure the wall studs; (iv) a sheathing station where drywall is installed on the exterior walls; and (v) a temporary storage area where wall panels are stored as they await delivery to sites. The industry partner produces various types of LGS wall panels for six-storey residential buildings, each comprising 120 units and more than 1,500 wall panels with varying design specifications. Workstation process times (200 series of time data) were collected and stored in a database using the C-track app. The names of wall panels, the number of workers at workstations, and module design specifications such as the number of studs and total area of windows and doors (in sq. ft) were stored manually into the database.

The 200 series of time data and impact factors of each of the three main workstations (i.e., the assembly station, the framing station, and the sheathing station) were used for the data analysis. The time-series data collected, though it would be considered a relatively small dataset, provided useful insights about the production line, as illustrated below. Looking at the results of the data analysis, the workstation process times were found to vary depending on the wall panel design specifications, even when the allocation of tasks and workers at workstations did not change. For example, as shown in Fig. 4(a), the process times at the assembly workstation ranged from 7 to 99 minutes depending on the number of wall panel design factors (e.g., the number of studs, plates, and clips). The data in the figure also indicates that at this particular workstation all the listed wall panel design factors affected the process times. For example, the number of plates in the wall panels ranged from 5 to 80, and this factor, as the figure clearly shows, strongly affected the process times. However, as shown in Fig. 4(b) and Fig. 4(c), it was found that some design factors (e.g., number of tracks and window area) did not affect the process times at some workstations. In other words, these factors do not have a significant relationship with the workstation process times. For instance, as shown in Fig. 4(b), the number of tracks required was 2 to 3 for all types of wall panels; however, the process times were found to vary from 11 to 102 minutes. As illustrated in Fig. 4(d), there was found to be a high level of variance in the process times at workstations due to the influence of these design factors. For example, although there was an SOP at the assembly workstation, the process times ranged from 23 to 100 minutes, leading to reduced productivity due to an imbalanced production line. As described above, this imbalance can be reduced by planning more efficient sequencing of modules and allocation of labor based on historical production data. In this respect, the next critical step in the case study would be to identify the SIFs among the design factors that highly influence the process times.

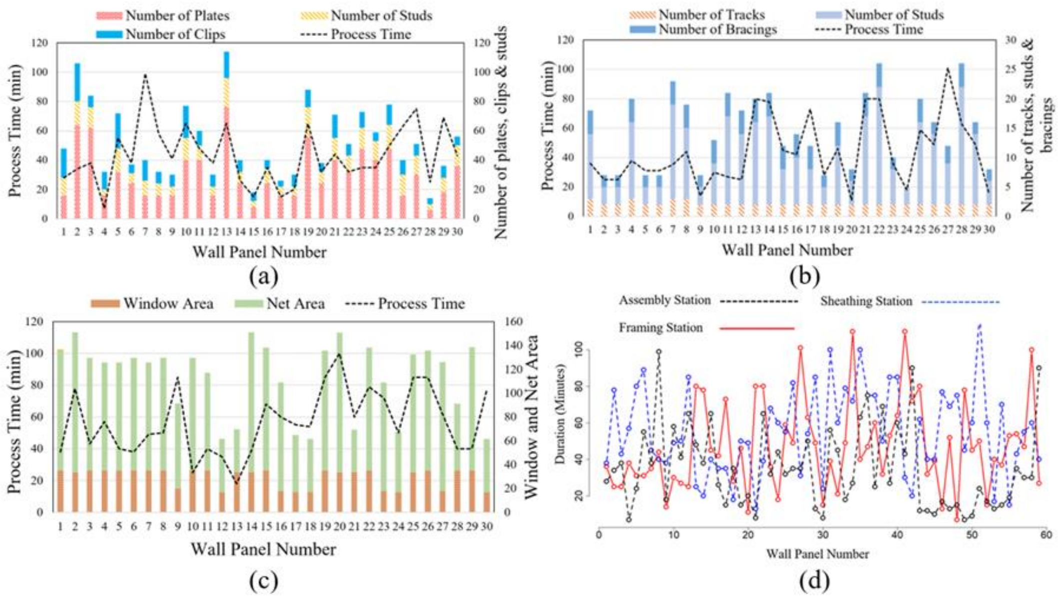

FIG. 4(a,b,c): Effect of design specifications on process time, and (d) average productivity of workstations.

Prior to identifying the SIFs, though, data pre-processing was implemented to identify and remove the outliers from the raw dataset based on the scatter plot and the standardized residual test. Based on the expert opinion of the industry partner's production manager, process times >80 minutes were removed from the scatter plot, given that such data points are indicative of a work disruption such as a delay due to errors in the shop drawings. After this, the standardized residuals test was implemented to identify hidden outliers. The data points with standardized residuals outside the range of ±1.64 were considered outliers and were removed from the dataset. As a result of the pre-processing task, the datasets numbered 178, 180, and 178 for the assembly, framing, and sheathing workstations, respectively. These data were normalized using a min–max normalization technique to transform the integer values into values ranging between 0 and 1. A *t*-test, PCC, and PCA were implemented to find the SIFs, with these, in turn, serving as the primary input in the development of the MLR models. The results of the *t*-test and PCC are represented in Table 1. With the significance level (p-value) defined as 1%, the different workstations were found to have different SIFs. For example, for the assembly workstation, six design factors (number of: header foam, studs, stud's foam, plates, clips, and openings) were identified as SIFs. However, the SIFs at the sheathing station were track length, number of angles, and window and door area. In terms of PCC, the design factors were selected by examining the correlation coefficient, which represents the relationship between the dependent variable (duration) and the independent variables. Design factors with a correlation coefficient >0.65 as defined by the experiments were deemed to be SIFs. Based on this, for the assembly workstation, four factors (number of: studs, studs foam, cripples, and openings) had a coefficient >0.65 and thus were identified as SIFs.

Table 1: Results of t-test and Pearson correlation coefficient

Assembly Station			Framing Station			Sheathing Station		
Factors	P value	Corr.	Factors	P value	Corr.	Factors	P value	Corr.
# of HeaderTrack	0.02977	0.57	# of HeaderTrack	2.08e-09	0.77	TrackLength	1.88e-11	0.71
# of SillTrack	0.05807	0.59	# of SillTrack	2.78e-12	0.58	# of Studs	0.697	0.73
# of HeaderFoam	0.00212	0.40	# of Studs	0.434	0.76	WindowArea	2.17e-07	0.42
# of Studs	1.40e-05	0.82	# of Bracings	0.007	0.71	Door Area	0.0002	0.05
# of StudsFoam	0.00141	0.74	Net Area	6.99e-05	0.18	Net Area	0.993	0.68
# of Plates	0.00393	0.62	# of Openings	0.69	0.70	# of Angles	1.61e-09	0.48
# of Clips	0.00086	0.36	-	-	-	# of Openings	0.941	0.39
# of Cripples	0.9381	0.66	-	-	-	-	-	-
# of Openings	3.53e-05	0.69	-	-	-	-	-	-

In addition to the *t*-test and PCC, PCA was applied. The first step in the PCA was to determine the percentage of variance for PCs, selecting for further analysis the set of PCs that cumulatively accounted for 90% of the dataset's total variation. For example, for the assembly station, the minimum set of PCs accounting for ≥90% of the cumulative variance was the set of PC1, PC2, PC3, and PC4, representing 61.5%, 14.1%, 9.56%, and 6.9%, respectively, of the variation, for a cumulative variation of 92.1%. These components were used as the basis for identifying SIFs, where PC1 was found to be highly correlated with the number of openings, PC2 with the number of clips, PC3 with the number of studs, and PC4 with the number of plates.

Based on the training dataset (80%) and the lists of SIFs generated by the different analyses, MLR models were developed and validated using a K-fold cross-validation. The training dataset was randomly split into 10 folds, with one fold used for the testing set and the remaining 9 folds used as a training set. The cross-validation process was repeated against all 10 folds in the dataset, and the average evaluation indices were calculated. The best predictive model was then selected based on four performance indices: R^2, adj. R^2, RMSEs, and MAE. As observed in Table 2, for the assembly station, the R^2, adj. R^2, RMSE, and MAE values in the training and testing datasets were 80.1%, 74.33%, 79.08%, 71.01%, 7.93 min, 9.83 min, 5.82 min, and 7.16 min, respectively, when the SIFs were the number of: header foam, studs, stud's foam, plates, clips, and openings identified by the *t*-test. The R^2 value for testing was 80.1%, meaning that the model was found to predict 80% of the outcomes. The RMSE value for testing depicted the deviation of 9.83 minutes between predicted and actual duration. However, for the framing station, the SIFs identified by the correlation test were selected instead of those identified in the *t*-test, since, for this station, the RMSE and MAE values generated by the correlation test were lower than those generated by the *t*-test.

Table 2: Results of multiple regression models at workstations

Model	R square (%)		Adj. R square (%)		RMSE (min)		MAE (min)	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
			Assembly Station					
t-test	80.1	74.33	79.08	71.01	7.93	9.83	5.82	7.16
Correlation	79.86	71.38	78.83	70.07	7.97	9.88	6.32	7.64
PCA	74.35	71.23	73.27	69.79	9.00	10.59	7.033	7.95
			Framing Station					
t-test	78.09	76.56	77.36	74.31	7.09	8.10	5.78	6.01
Correlation	79.13	74.55	78.33	73.48	6.89	6.99	5.56	5.89
PCA	64.74	57.01	63.87	51.64	8.37	10.20	6.58	8.46
			Sheathing Station					
t-test	74.90	73.35	72.37	71.26	9.04	9.54	7.56	8.19
Correlation	61.91	49.66	60.56	46.93	11.77	13.80	9.66	11.20
PCA	57.59	51.7	55.7	50.01	10.88	12.40	8.32	10.66

Table 3 represents the cycle time formula (CTF) of workstations, which involves the coefficient values of SIFs to predict the process times of workstations. As can be seen, some of the coefficients (Coef.) of SIFs had positive values while others were negative. For example, for the assembly station, the X_{HF}, X_S, and X_C coefficients indicate that the process times increased when the X_{HF}, X_S, and X_C in the wall panel increased. Similarly, the coefficient of X_W means that the process times decreased whenever additional workers were allocated to the station. To improve accuracy, probability distribution function-based models were developed based on the time data used in the statistical analysis. To identify the most suitable probability distribution functions, as shown in Table 3, a goodness-of-fit method, consisting of both a Kolmogorov–Smirnov (K–S) test and a chi-square test, was performed. Weibull, gamma, and triangular distributions were selected as the best fit for the assembly, framing, and sheathing stations, respectively. These distributions, along with cycle time formula (CTF), were used as inputs in the simulation model.

Table 3: Cycle time formula and summary of the goodness-of-fit test

Cycle time formula					
Assembly Station	Coef.	Sheathing Station	Coef.	Framing Station	Coef.
Number of Header Foam (X_{HF})	20.54	Track Length (X_{TL})	70.32	Number of Header Track (X_{HS})	15.14
Number of Studs (X_S)	38.36	Number of Angles (X_A)	15.63	Number of Studs (X_S)	20.78
Number of Plates (X_P)	−21.49	Window Area (X_{WA})	−36.47	Number of Bracings (X_B)	7.43
Number of Clips (X_C)	8.74	Door Area (X_D)	−20.72	Number of Openings (X_o)	0.05
Number of Workers (X_W)	−16.19	Number of Workers (X_W)	−7.27	Number of Workers (X_W)	−11.3
Number of Studs foam (X_{SF})	27.75	-	-	-	-

Goodness-of-fit test								
Assembly Station			Framing Station			Sheathing Station		
Distribution	K-S	X^2	Distribution	K-S	X^2	Distribution	K-S	X^2
Weibull	0.0818	45.47	Gamma	0.0543	26.80	Triangular	0.0752	27.59
Triangular	0.0915	64.26	Triangular	0.0558	30.83	Uniform	0.0849	16.29
Gamma	0.0916	41.97	Weibull	0.0571	27.35	Normal	0.0891	31.17
Normal	0.1052	64.26	Lognormal	0.0678	28.82	Weibull	0.0958	22.57
Lognormal	0.1063	60.21	Uniform	0.0701	35.04	Gamma	0.1074	42.47

Based on the process flow of the production line, the simulation model illustrated in Fig. 5 was built in Simphony.NET. The simulation model uses the ‘database’ element to update wall panel information from a central database using SQL. In the simulation model, ‘Database Create,’ ‘task,’ and ‘conditional branch,’ ‘resource’, and ‘destroy’ elements from the general template are used to mimic the process flow of the actual production line. Depending on the type of wall (interior or exterior), different process flows were required. In the model, the ‘Database Create’ element generates the simulation entities (wall panels) and passes them into the next task elements, which are the assembly station and framing station. After completing the tasks at the framing station, the wall panel components proceed to the ‘conditional branch’, where they are classified as components of either interior or exterior walls (since these types of walls require different tasks). From the ‘conditional branch element’ pertinent were related to types of wall panels, interior wall panel components are directed to ‘storage’ while exterior wall panel components are directed to the ‘sheathing station’. Once the process is complete at each workstation, the given entity is destroyed using the ‘destroy’ element. The simulation model was run using cycle time formula (CTFs) and probability distribution functions as inputs, and the results in terms of productivity were compared with the historical actual productivity data for validation purposes. It should be noted that 2 and 3 workers were allocated to the assembly and framing stations, respectively, for interior wall panels, while 3, 2, and 3 workers were allocated to the assembly, framing, and sheathing stations, respectively, for exterior wall panels.

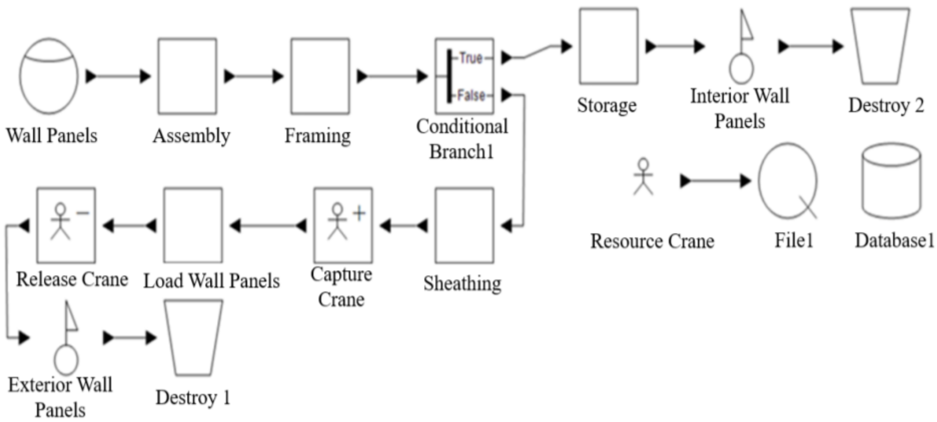

FIG. 5: Simulation model of a production line.

Fig. 6 shows a comparison of the actual(historical) and simulated (cycle time formula and probability distribution functions) cycle times for producing interior wall panels. As can be seen, the actual cumulative cycle time for the manufacture of 205 interior and exterior wall panels was 119 hr, compared to 106.38 hr for the cumulative cycle time formula and 90.1 hr for the cumulative probability distribution functions. It should be noted that cumulative cycle time is the total sum of the process times of all wall panels, while cumulative probability distribution functions is the total process time to complete the wall panels with probability distribution as an input in the simulation.

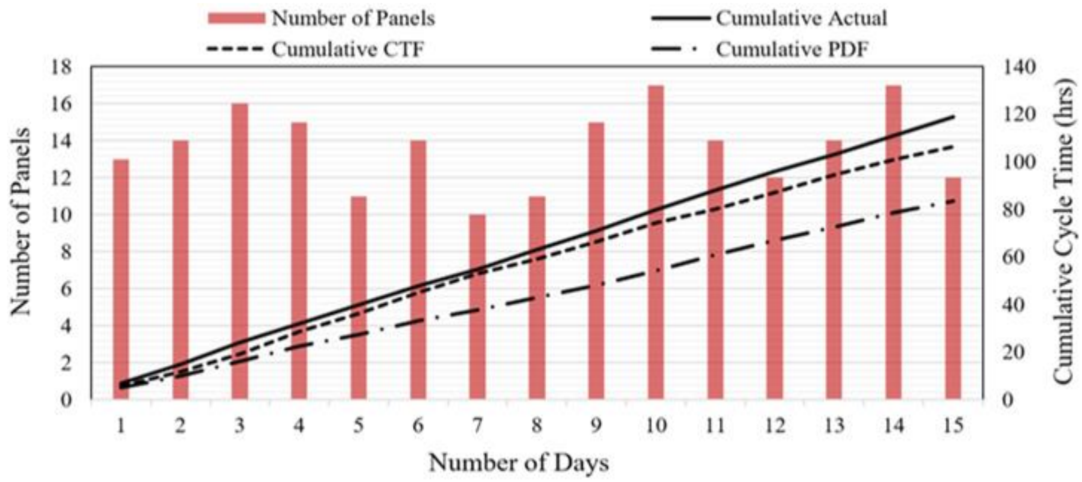

FIG. 6: Comparison between cumulative actual and simulated cycle time.

These results are indicative of general agreement between the actual and simulated cumulative process times generated by the CTF (i.e., 89.39% accuracy). In this respect, the simulation model developed and implemented in conjunction with cycle time formula (CTF) in this case study was deemed to be a reliable predictive model. After validating the production rate generated, multiple simulation-based scenarios were developed and analyzed using cycle time formula (CTF).

To compare the simulation results to the historical production data, the fabrication of 309 wall panels in five different panel-sequencing scenarios, each with the same labor allocations at the various workstations, was simulated. In the first scenario, the units of a residential project were selected in a clockwise direction and manufactured unit-by-unit, and their wall panels (exterior or interior) were prefabricated randomly in the production line. In the second scenario, the units of the residential project were again selected in a clockwise direction and manufactured unit-by-unit, but this time the exterior wall panels of a given unit were prefabricated first, followed by the corresponding interior walls for the unit. In the third scenario, the interior wall panels of a given unit were manufactured first, followed by the exterior wall panels. In the fourth scenario, the interior wall panels of all units were produced first, followed by the exterior wall panels of all units. In the fifth scenario, the exterior wall panels of all units were produced first, followed by the interior wall panels of all units. Based on these scenarios, the simulation model provided the cumulative duration to prefabricate all wall panels in each of the five scenarios, yielding cumulative durations of 93.52 hr, 51.92 hr, 69.76 hr, 59.84 hr, and 60.64 hr for the five respective scenarios. Thus, Scenario 2 was found to outperform the other scenarios.

In terms of production planning, finding the optimum allocation of labor to workstations is crucial in that it is a critical factor in (i) synchronizing process times; (ii) reducing waiting times along the production line, and (iii) increasing productivity. In this respect, as shown in Table 4, in the case study we experimented with different numbers of workers as an additional decision variable in the simulation model built based on Scenario 2. Workers were allocated to workstations according to the type of wall panel, and Scenario 2.3 was found to outperform the other labor allocation scenarios, with a cumulative duration of 44.42 hr. It is notable that the duration in Scenario 2.4 was found to be longer than that in Scenario 2.3 even though the number of workers at the framing and sheathing station was increased in Scenario 2.4. This may have been attributable to space congestion at the framing station disrupting the coordination between the workers completing the work. In this regard, Zhang et al. (2020) have demonstrated that space congestion interrupts the workflow in MCM, thereby reducing productivity at the workstation level. As such, this simulation-based statistical method for production planning in MCM is a significant tool for analyzing the effect of different crew sizes, especially in that it demonstrates that it is not always ideal to increase the number of workers at a workstation to reduce process times.

Table 4: Comparison of different crew sizes at workstations

Scenario	Duration (hours)	Number of workers				
		Interior Wall Panels		Exterior Wall Panels		
		Assembly	Framing	Assembly	Framing	Sheathing
2.1	51.92	2	3	3	2	3
2.2	45.44	2	3	3	3	3
2.3	44.42	2	4	3	2	2
2.4	45.42	2	4	3	3	3
2.5	51.04	2	2	3	2	3
2.6	51.04	2	2	3	3	3

The case study demonstrates that optimal sequencing of modules and allocation of workers is critical to improving productivity. The method described in this study is also capable of identifying the SIFs affecting fabrication process times, thereby removing the guesswork from production planning. However, it should be noted that the SIFs in this study are a function of the given product design specifications and tasks performed at the workstations of the case study production line. For other cases, practitioners would need to modify the data analysis phase based on the given design specifications in order to identify the SIFs. In modular construction, due to unpredictable demand, production managers must frequently alter their plans to accommodate change orders. As an alternative to this challenging and error-prone approach, the framework implemented in the case study can be deployed to devise different production line scenarios in terms of labor allocation and sequencing to streamline the MCM planning process.

5. CONCLUSIONS

In MCM, the modules vary in size and design specifications, meaning that the production line operates following a mixed-production line model. This poses a challenge for production planning, leading to inefficient utilization of resources and reduced productivity. In this respect, this paper proposes a simulation-based production line planning method that uses near real-time and historical data to assist production managers in achieving better productivity and control. This method integrates a C-track app, statistical analysis, and simulation for production planning in MCM. As an alternative to the experience-based approach implemented in traditional MCM production planning, the developed method collects historical and near real-time data using the C-track app to enhance decision making; identifies the SIFs affecting fabrication process times using statistical techniques that increase the accuracy of the predictive model; and improves productivity by using simulation as a production planning tool. It also introduces the concept of developing and evaluating multiple production sequencing and labor allocation scenarios using two types of input cycle time formula and probability distribution functions in the simulation model. The case study implementation of the developed method in an LGS wall panel production line demonstrates that this approach ensures improved productivity (i.e., reduced durations with the same labor input) and control. In particular, the case study results indicate a 44.42 hr duration to produce 309 wall panels. As demonstrated by the case study, this method can assist production managers in understanding the effects of proposed changes to the production line before implementing them in reality. In this way, production managers can plan effectively and reduce project costs. The introduced simulation-based planning method can assist production managers to evaluate the optimal scenario of (i) sequencing the modules and (ii) allocating resources along the production line, which can reduce idle time of workers and waiting time of modules at workstations.

However, the developed method does have some limitations, such as the need to manually test the simulation model scenarios. Therefore, future work will seek to enhance the performance of the developed method. This will include, for one, integrating simulation and advanced optimization techniques to identify sequencing of modules and labor allocation scenarios more efficiently and rapidly. In a simulation model, different decision variables can be analyzed and in the optimization model, heuristic rules can be developed in order to determine optimal solutions. Second, various key performance indicators (KPIs) for modular construction manufacturing can be developed in order to evaluate the performance of production line, which later can be integrated with simulation and optimization for re-scheduling (i.e. sequences of modules and allocation of resources). Third, simulation will be integrated with lean principles, where the former will be used for gaining a deeper understanding of the production line while the latter will be used for evaluating various production scenarios. It should also be noted that in some cases SOPs may not be available, depending on the company. In such cases, direct observation and interviews with plant managers can be used to identify the sequences of activities at workstations. Subsequently, integration of off-site, transportation and on-site schedule can be introduced in future work in order to develop an overall tracking and control system for modular construction.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) (Grant File No. CRDPJ 522018-17).

REFERENCES

AbouRizk S., Hague S. and Ekyalimpa R. (2016). Construction simulation: An introduction using Simphony, Edmonton, AB, Canada: University of Alberta.

Altaf M.S., Al-Hussein M. and Yu H. (2014). Wood-frame wall panel sequencing based on discrete-event simulation and particle swarm optimization, *Proceedings of the International Symposium on Automation and Robotics in Construction,* Vol. 31, 1.

Altaf M.S., Bouferguene A., Liu H., Al-Hussein M. and Yu H. (2018). Integrated production planning and control system for a panelized home prefabrication facility using simulation and RFID, *Automation in Construction*, Vol. 85, 369–383.

Alvanchi A., Azimi R., Lee S., AbouRizk S.M. and Zubick P. (2011). Off-site construction planning using discrete event simulation, *Journal of Architectural Engineering*, Vol. 18, No. 2, 114–122.

Azimi R., Lee S., AbouRizk S.M. and Alvanchi, A. (2011). A framework for an automated and integrated project monitoring and control system for steel fabrication projects, *Automation in Construction*, Vol. 20, No. 1, 88–97.

Bertram N., Fuchs S., Mischke J., Palter R., Strube G. and Woetzel J. (2019). Modular construction: From projects to products, McKinsey & Company <modular.org/documents/document_publication/mckinsey-report-2019.pdf> (assessed August, 2019).

Bhatia A.P.S., Han S., Moselhi O., Lei Z. and Raimondi C. (2019). Simulation-based production rate forecast for a panelized construction production line, *7th International CSCE Construction Conference*.

Chanmeka A., Thomas S.R., Caldas C.H. and Mulva S.P. (2012). Assessing key factors impacting the performance and productivity of oil and gas projects in Alberta, *Canadian Journal of Civil Engineering*, Vol. 39, No. 3, 259–270.

Cottrell D.S. (2006). Contractor process improvement for enhancing construction productivity, *Journal of Construction Engineering and Management*, Vol. 132, No. 2, 189–196.

El-Rayes K. and Moselhi O. (1998). Resource-driven scheduling of repetitive activities, *Construction Management & Economics*, Vol. 16, No. 4, 433–446.

Elmousalami H.H. (2019). Artificial intelligence and parametric construction cost estimate modeling: State-of-the-art review, *Journal of Construction Engineering and Management*, Vol. 146, No. 1, 03119008.

Gerald B. (2018). A brief review of independent, dependent and one sample *t*-test, *International Journal of Applied Mathematics and Theoretical Physics*, Vol. 4, No. 2, 50.

Guyon I. and Elisseeff A. (2003). An introduction to variable and feature selection, *Journal of Machine Learning Research*, Vol. 3, 1157–1182.

Hajjar D. and AbouRizk S. (1999). Simphony: An environment for building special purpose construction simulation tools, *Proceedings of the 31st Winter Simulation Conference*, 998–1006.

Halpin D.W. (1977). Cyclone-method for modelling job site processes, *Journal of the Construction Division*, Vol. 103, No. 3, 489–499.

Han S.H., Al-Hussein M., Al-Jibouri S. and Yu H. (2012). Automated post-simulation visualization of modular building production assembly line, *Automation in Construction*, Vol. 21, 229–236.

Hong T., Cho K., Hyun C. and Han, S. (2011). Simulation-based schedule estimation model for ACS-based core wall construction of high-rise building, *Journal of Construction Engineering and Management*, Vol. 137, No. 6, 393–402.

Hsu P.Y., Aurisicchio M. and Angeloudis P. (2019). Risk-averse supply chain for modular construction projects, *Automation in Construction*, Vol. 106, 102898.

Kavanagh D.P. (1985). SIREN: A repetitive construction simulation model, *Journal of Construction Engineering and Management*, Vol. 111, No. 3, 308–323.

Key Small Business Statistics. <https://www.ic.gc.ca/eic/site/061.nsf/eng/h_03090.html#point1-1>, (accessed August, 2019).

Liu H., Altaf M.S., Lei Z., Lu M. and Al-Hussein M. (2015). Automated production planning in panelized construction enabled by integrating discrete-event simulation and BIM, *5th International/11th Construction Specialty Conference*, 8-10.

Mahmood Z. and Khan S. (2009). On the use of k-fold cross-validation to choose cutoff values and assess the performance of predictive models in stepwise regression, *The International Journal of Biostatistics*, Vol. 5, No. 1.

Makridakis S., Spiliotis E. and Assimakopoulos V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward, *PLoS ONE*, Vol. 13, No. 3, 0194889.

Martinez P., Mohamed E., Mohsen O., and Mohamed Y. (2020). Comparative study of data mining models for prediction of bridge future conditions, *Journal of Performance of Constructed Facilities*, Vol. 34, No. 1, 04019108.

Modular Building Institute, Permanent Modular Construction Report <https://www.deluxemodular.com/hubfs/Third%20Party%20Press%20Downloads/MBI-2018PermanentModularConstructionReport.pdf> (accessed July, 2019).

Mohsenijam A. and Lu M., (2019). Framework for developing labour-hour prediction models from project design features: Case study in structural steel fabrication, *Canadian Journal of Civil Engineering*, Vol. 46, No. 10, 871–880.

Piryonesi S.M. and El-Diraby T.E. (2019). Data analytics in asset management: Cost-effective prediction of the pavement condition index, *Journal of Infrastructure Systems*, Vol. 26, No. 1, 04019036.

Rocchi L., Chiari L. and Cappello A. (2004). Feature selection of stabilometric parameters based on principal component analysis, *Medical and Biological Engineering and Computing*, Vol. 42, No. 1, 71–79.

Salama T., Salah A. and Moselhi O. (2018). Integration of linear scheduling method and the critical chain project management, *Canadian Journal of Civil Engineering*, Vol. 45, No. 1, 30–40.

Willmott C.J. and Matsuura K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Climate Research*, Vol. 30, No. 1, 79–82.

Xie H., Chowdhury M.M., Issa R.R. and Shi W. (2018). Simulation of dynamic energy consumption in modular construction manufacturing processes, *Journal of Architectural Engineering*, Vol. 24, No. 1, 04017034.

Xu C., Zhang Y. and Xu B. (2017). Research of construction safety factors based on the factor analysis method. *Proceedings, 2016 International Conference on Construction and Real Estate Management: BIM Application and Off-Site Construction*, 1348–1354.

Yu L. and Liu H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, 856–863.

Yu H., Tweed T., Al-Hussein M., and Nasseri R. (2009). Development of lean model for house construction using value stream mapping, *Journal of Construction Engineering and Management*, Vol. 135, No. 8, 782–790.

Zaalouk A. and Han S. (2021). Parameterized Design Optimization Framework for Worker-Friendly Workplaces in Modular Construction. *Journal of Construction Engineering and Management*, Vol.147, No. 5, 04021030.

Zhang Y. (2017). A framework to improve modular construction manufacturing production line performance. MSc thesis, University of Alberta.