ASCE

Distributed Simulation–Based Analytics Approach for Enhancing Safety Management Systems in Industrial Construction

Estacio Pereira[1]; Mostafa Ali[2]; Lingzi Wu[3]; and Simaan Abourizk, M.ASCE[4]

Abstract: Although methods for assessing and simulating the influence of safety-related measures on safety performance have been proposed, practical applications remain limited. Data required by these methods are dispersed across departments, necessitating the development or redesign of data warehouses. This research proposes a simulation-based analytics approach to enhance safety management system (SMS) decision making using distributed simulation to overcome limitations associated with previous approaches. This distributed simulation approach is used to (1) integrate historical data without modifying data-warehouse structures (i.e., data fusion component), (2) link data to an artificial neural network–based analysis component for determining the influence of safety-related measures on incident levels, (3) connect data and analysis components to existing simulation components, and (4) combine the outputs, resulting in a comprehensive safety performance evaluation system to examine incident levels. Results demonstrate that this approach successfully fuses and integrates data from several sources with analysis and simulation components in a cost-, labor-, and time-efficient manner. A distributed simulation–based analytics approach represents a considerable opportunity for industrial construction companies to more effectively use historical data, analysis tools, and simulation models. **DOI:** 10.1061/(ASCE)CO.1943-7862.0001732. *© 2019 American Society of Civil Engineers.*

Author keywords: Safety management system; Safety-related measures; High-level architecture (HLA); Distributed simulation.

Introduction

The safety performance of contractors, measured as the ratio of recordable incidents to the number of working hours, has become one of the major factors affecting a company's decision to award a contract. Accordingly, safety has become increasingly prioritized in the construction industry, especially in the industrial construction domain (Jannadi and Bu-Khamsin 2002; Salas and Hallowell 2016; Saunders et al. 2017). Although there has been a notable improvement in safety performance due to the implementation of safety practices (e.g., mandatory safety training and accident investigations) and increased resource allocation by organizations (e.g., safety managers on-site and use of personal protective equipment) (Hinze et al. 2013a), accidents continue to occur on construction worksites (US Department of Labour 2017).

The occurrence of incidents may be a result of the intricate nature of construction projects. Projects are composed of simple and complex processes requiring a variety of resources and resulting in an intricate, dynamic environment (Marle and Vidal 2016). The control of hazards remains challenging in an environment where decisions taken on several levels and by different decision makers can inadvertently interact in a manner that leads to the occurrence of an incident. To manage this complex environment, and consequently mitigate hazards, many organizations have implemented a safety management system (SMS). According to American National Standards Institute (ANSI) and the American Industrial Hygiene and Association (AIHA) ANSI/AIHA Z10:2005 (ANSI 2012), a SMS is "broadly characterized as a set of institutionalized, interrelated, and interacting elements strategically designed to establish and achieve occupational health and safety."

The implementation of a traditional SMS by organizations in practice carries some drawbacks. First, a SMS is usually assessed using reactive indicators (e.g., total recordable incident rate or days away from work) (Esmaeili and Hallowell 2013), which measure safety performance immediately following the occurrence of an incident. Second, a SMS relies heavily on expert (e.g., safety manager) experience. Due to the complex environment of construction projects, safety managers may have difficulties perceiving how all potential factors interact with each other to contribute to an incident.

In an attempt to address these challenges, several safety-related measures and methods capable of proactively assessing safety performance have been proposed in the literature (Guo et al. 2016; Hinze et al. 2013b; Podgórski 2015; Poh et al. 2018). In practice, however, these measures are collected by different project agents and are not freely shared within the organization for the purpose of sophisticated decision making, thereby limiting the practical implementation of these assessment methods. Even on

[1]Postdoctoral Fellow, Dept. of Civil and Environmental Engineering, Univ. of Alberta, 5-080 Markin/Canadian Natural Resources Ltd., Natural Resources Engineering Facility, 9105 116 St. NW, Edmonton, AB, Canada T6G 2W2. ORCID: https://orcid.org/0000-0002-1654-1827. Email: estacio@ualberta.ca

[2]Postdoctoral Fellow, Dept. of Civil and Environmental Engineering, Univ. of Alberta, 5-080 Markin/Canadian Natural Resources Ltd., Natural Resources Engineering Facility, 9105 116 St. NW, Edmonton, AB, Canada T6G 2W2. ORCID: https://orcid.org/0000-0003-3571-3070. Email: mostafaali@ualberta.ca

[3]Ph.D. Student, Dept. of Civil and Environmental Engineering, Univ. of Alberta, 5-080 Markin/Canadian Natural Resources Ltd., Natural Resources Engineering Facility, 9105 116 St. NW, Edmonton, AB, Canada T6G 2W2. Email: Lingzi1@ualberta.ca

[4]Professor, Dept. of Civil and Environmental Engineering, Univ. of Alberta, 5-080 Markin/Canadian Natural Resources Ltd., Natural Resources Engineering Facility, 9105 116 St. NW, Edmonton, AB, Canada T6G 2W2 (corresponding author). ORCID: https://orcid.org/0000-0002-4788-9121. Email: abourizk@ualberta.ca

Note. This manuscript was submitted on December 3, 2018; approved on May 2, 2019; published online on October 31, 2019. Discussion period open until March 31, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, © ASCE, ISSN 0733-9364.

multi-billion-dollar megaindustrial construction projects, it is not uncommon for organizations to collect and store site information in data warehouses that were designed to serve one agent's interests alone. Indeed, the structures, as well as the contents, of these data warehouses vary drastically from one contractor to another. These protected and self-serving data-warehouse structures create significant obstacles when it comes to the analysis of data across departments for the purpose of supporting higher-level decisions (Husted and Snejina 2002). To avoid the significant cost of redesigning and reconstructing an integrated data warehouse for all departments, the most time- and cost-efficient practice is to develop a decision-support approach that is capable of automatically retrieving required information from various organizational data sources. Importantly, the ability of this approach to enable communication between existing data analysis tools and/or simulation components would allow practitioners to capitalize on technologies currently available in literature and within an organization.

This research proposes a distributed simulation communication framework approach capable of fusing fragmented data from several sources and of integrating various data-analysis and simulation components to create a cohesive safety system in a cost-, labor-, and time-efficient manner. To demonstrate the functionality of the proposed approach, the resulting safety system was tested using actual project data. By facilitating data fusion and component integration, the proposed approach was found capable of removing certain practical challenges currently limiting the application of proactive safety assessment methodologies and tools proposed in literature, in turn increasing accessibility of industrial construction organizations to state-of-the-art technologies and improving safety-associated decision making in practice.

Research Background

Safety Management System and Safety-Related Measures

According to Wu et al. (2015), safety performance evaluation is a crucial element of a SMS because it provides information on the quality of the system and demonstrates the effectiveness and efficiency of the practices and policies being developed and implemented within an organization. Due to the complexity of construction projects, where different processes are continually interacting with each other (Dao et al. 2017), and because accidents are caused by many factors (Saleh et al. 2013), a SMS is required to define measures that can be used to proactively evaluate safety performance.

In this context, Hale (2009) stated that safety-related measures could be used to proactively manage a SMS because they are expected to provide early warnings of accidents and motivate people to take corrective actions. Although many safety-related measures have been proposed by academics and practitioners (Table 1), these indicators are usually measured and stored by different agents within an organization. In addition, the measures available can vary due to different cultures, practices, and management standards implemented across organizations, which may lead to variations in indicators used across the industry. A combination of safety-related measures should be used to increase model accuracy and to more reliably assess SMS performance. Indeed, managers can experience difficulties discerning the relationship between an indicator and project safety performance due to low accuracy when using only one indicator for predictions (Pereira et al. 2017).

Use of Historical Data to Enhance SMS

Although most companies collect significant amounts of data as a result of testing and monitoring systems (Alavi and Gandomi 2017), "they do not make good use of the generated or collected data to improve production system efficiency" (Dean 2014). The predominant approach to proactive management of industrial construction in practice relies heavily upon the knowledge and experience of construction professionals, who are often limited in their ability to capture the intricate details of various processes, resources, and uncertainties associated with the execution of a project (Manyika et al. 2011). Collected data are generally incomplete and often not mapped to the level of detail required for effective decision making (Pereira et al. 2017). For instance, although some indicators are stored at a worker level (e.g., payroll information), others are only stored at a project level (e.g., inspection rate).

Difficulties using historical data to proactively assess safety performance may arise from the often fragmented nature of construction data. In construction, data are usually stored in relational databases, which are unsuitable for use by different parts of a project due to specialized storage and processing needs (Bilal et al. 2016). For instance, safety-related measures may be stored in different warehouses, including, but not limited to, safety, design information [e.g., drawings, specifications, and three-dimensional (3D) models], project control (e.g., schedule, cost, and quality), and human resources. Based on the size of the contractor as well as the project, these warehouses could be as simple as one spreadsheet or as complex and sophisticated as a professionally designed system. Additionally, diverse format files in these warehouses add complexity to the data integration problem (Bilal et al. 2016). An approach able to integrate many data sources within an organization without modifying current practices for data collection and storage, therefore, could facilitate the identification and use of safety-related measures to enhance decision making.

Table 1. Examples of safety-related measures

References	Safety-related measures
Lee et al. (2012)	Temperature; workers' age; work days on current site; amount of safety training; type of building; work process rate
Hinze et al. (2013b)	Percentage of jobsite toolbox meetings attended by jobsite supervisors; number of close class reported per 200,000 h of worker exposure; inspection rate; worker behavior observation rate
Guo and Yiu (2017)	Working hours per day; percentage of workers with certificates to operate equipment; frequency with which safety planning is conducted before performing tasks; percent of subcontractors selected based on historical safety performance
Lingard et al. (2017)	Rate of nonsafety compliance; rate of toolbox meeting; rate of audits; rate of drug tests; rate of site inductions
Podgórski (2015)	Percentage of workers with defined occupational health and safety (OHS) practices in job description; percentage of workers trained on emergency procedures; percentage of complete corrective actions
Poh et al. (2018)	Project days delayed; average worker power at organizational level for the month; current progress of project; scaffold rate; weighted safety inspection score
Mohammadi et al. (2018)	Worker wage; reward and penalty; peer pressure; rule compliance; competence; safety experience; worker age; safety budget; safety personnel; schedule delay

High-Level Architecture Standard for Data Integration

Researchers have traditionally overcome the challenge of disparate data by first integrating all data into a single source and then implementing a system (Guo et al. 2016; Lee et al. 2012; Salas and Hallowell 2016); this approach, however, presents certain drawbacks. For example, a slight modification in the data structure by any agent, particularly through the addition of new data, can (1) make it difficult to update the data structure, (2) complicate system training, or (3) even render the system dysfunctional. Approaches capable of communicating with various data sources can overcome such limitations while providing each agent the freedom to store their data in any format of their choosing. Distributed simulation allows for the creation of larger systems by combining various components and simulations and has been shown to be capable of achieving these functionalities. Standards for achieving distributed simulation have resulted in the creation of an international standard, IEEE 1516 (IEEE Computer Society 2010), also known as the high-level architecture (HLA) standard, which has since been used by a variety of sectors for data, simulation, and analysis component integration (Azimi et al. 2011).

Proposed Simulation-Based Analytics Approach

Many factors contribute to the occurrence of an incident. To develop SMS that are effective, it is crucial to define policies to proactively control project safety performance and, synchronously, understand how various managers' combined decisions impact safety management. Although a variety of analytical methods and models have been developed to assist practitioners with the complex safety assessments inherent to large-scale industrial construction projects, completion of the data fusion, analysis, and simulation steps required remains a complicated, labor-intensive process.

The simulation-based analytics approach proposed in this research aims to overcome these limitations by capitalizing on advances in computing sciences to (1) integrate historical data without modifying the structure of data warehouses (i.e., data fusion component), (2) link fused data to a data analysis component capable of determining the influence of safety-related measures on safety-incident levels, (3) connect data and the analysis component with an existing simulation component to replicate the behavior of project performance, and (4) combine the outputs of the various components to create a comprehensive safety evaluation system. The approach was used to develop a safety system for an industrial construction organization as outlined in the following sections.

System Integration and Interoperability

Fig. 1 summarizes the four components of the proposed system: (1) the data warehouse responsible for updating and publishing the information of each data source; (2) the safety performance analysis module, which queries all information provided by the data warehouse, selects the critical safety-related measures, trains and tests the safety assessement model using a neural network algorithm, and is used during the simulation to assess the safety performance of various scenarios; (3) the simulation model, which replicates the various scenarios (i.e., decisions or actions) to predict their effect on current and future performance; and (4) the decision support interface, which is responsible for updating user scenarios, initiating the simulation model, sampling values from distributions (e.g., behavior-based observation cards), and allowing users to visualize how various policies affect incident levels.

The HLA of the system plays an essential role, allowing the various components to communicate with one another. It is used because of its capacity to exchange information remotely, bring together complementary data and information from different locations, and make these data available to all components. The development of the proposed safety system uses different programming languages, namely C# and Java. Notably, the interoperability between the components is one of the advantages of using the HLA approach. For this purpose, Portico RTI version 2.1 (Portico RTI 2018) was used in this research.

Fig. 2 illustrates the functionality of the system developed using the proposed approach. The system is composed of two phases: initialization and simulation. In the initialization phase, the HLA environment is created, historical project performance information is updated, and the safety performance analysis component trains and validates the safety assessment model. In the simulation phase, the behavior of the project to be evaluated (as defined by the user) is replicated by the simulation model, with outputs assessed at defined intervals. The decision support interface then reports these values to the user. The simulation model is terminated once the evaluation period (as defined by the user) is achieved.

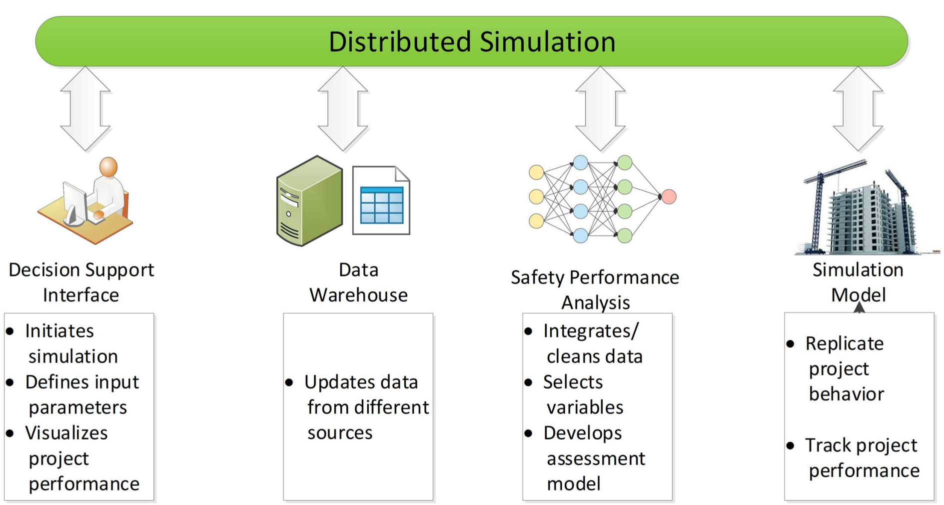

Fig. 1. Proposed simulation-based analytics approach for safety management.

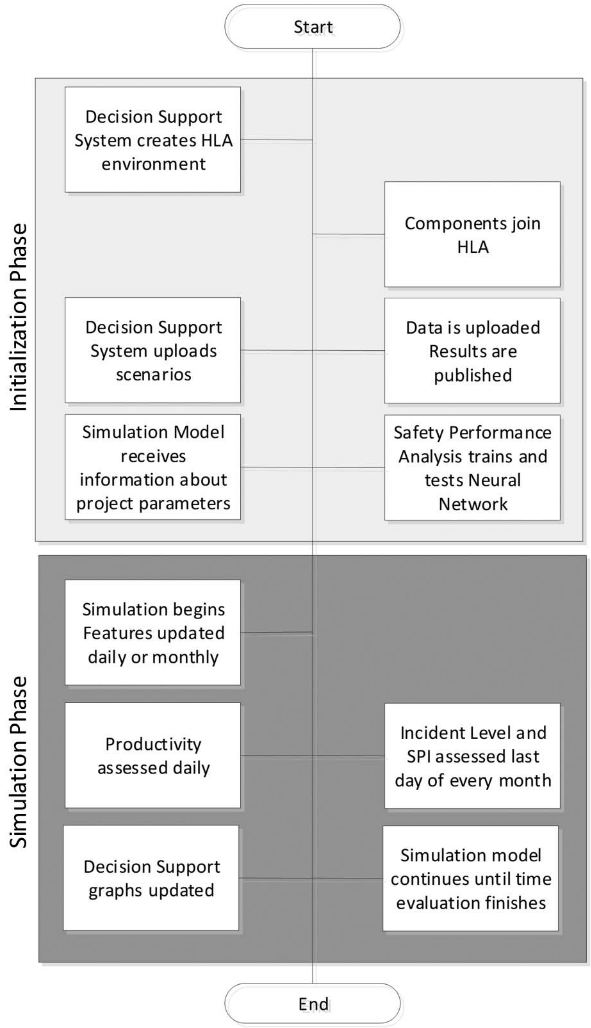

Fig. 2. Simulation-based analytics phases.

Data Warehouse

The data warehouse component integrates and publishes information stored in different departmental databases or files such as spreadsheets, drawings, and text files. Because data collected by the various departments are often stored in disconnected database systems and in different formats, each data source is uploaded into the system as one independent component of the model.

In this research, data from the human resources department, project control, environment conditions, and the health and safety department were used to test the suggested approach. The information of each data source was available as a CSV file. Table 2 contains a list of all safety-related measures considered in each data warehouse. The output collected from the data warehouse is the safety incident rate [calculated using Eq. (1)]. Eq. (1) considers first aid, lost time incidents, major first aid, and fatalities to determine the safety incident rate. The methodology used to identify each leading indicator has been detailed by Pereira et al. (2018a)

$$\text{Safety incident rate}_n = \frac{\text{Quantity of incidents}_n \times 200{,}000}{\text{Working hours}_n} \tag{1}$$

where safety incident rate = total incident rate; quantity of incidents$_n$ = number of incidents; working hours$_n$ = quantity of working hours; subscript n = during time interval n; and 200,000 approximates the number of hours worked by 100 workers per year (i.e., at 40 h per week for 50 weeks).

In addition to uploading the information from the previous project, the environmental conditions data warehouse is also responsible for sampling values from historical weather based on the project location and for publishing weather parameters (e.g., temperature and wind speed average) during the simulation.

Table 2. Safety-related measures available in the data-warehouse component

Data warehouse	Safety-related measures available
Health and safety	Project number
	Year
	Month
	Behavior-based observation rate (BBO)
	Inspection rate
	Total audited pre-job safety inspection (PSI) rate
	Near miss rate
	Investigation time average
	Training hours average
	Safety incident rate
Human resources	Project number
	Year
	Month
	Average working hours per worker
	Average worker experience on the site
	Average foreman experience on the site
	Percentage of operators
	New workers rate
	Percentage of workers <30 and <50 years old
Environmental conditions	Project number
	Year
	Month
	Average temperature
	Average wind speed (km/h)
Project control	Project number
	Year
	Month
	Crew size
	Project ramp-up/ramp-down
	Scaffolding hours rate
	Schedule performance index (SPI)
	Project evaluation

Safety Performance Data Analysis Module

The behavior of this component is divided into four phases: (1) query the data, (2) clean the data, (3) select the appropriate safety-related measures, and (4) train and test the assessment module to predict the safety performance level. First, the analysis component combines all information from the data warehouse and classifies them according to project number, year, and month. Second, the final data set is preprocessed by cleaning and removing lines with incomplete information and safety incident rate values that exceed the 95th percentile. Here, the library jOOL version 0.9.12 (jOOL 2018) was used to perform the queries. Because the objective of this particular safety system is not to predict safety performance, but rather identify how management decisions affect safety output, all safety-related measures in the final data set are standardized from 0 to 1 using feature scaling. To avoid any misunderstanding by users attempting to predict the “number” of accidents, the Safety Incident Rate is transformed into an Incident Level using Eq. (2).

J. Constr. Eng. Manage., 2020, 146(1): 04019091

Table 3. Description of safety-related measures considered in the final assessment model

Indicator	Description
BBO rate	Number of BBO cards[a] filled out in a given month, divided by the number of working hours in the same month, multiplied by 200,000[b] working hours per year
Crew size	Ratio of the number of workers to foreperson in a given month
Project elevation	Average project elevation (m)
Temperature	Average temperature (Celsius) on a site in a given month
Schedule performance index (SPI)	Actual working hours divided by the planned working hours
Ramp-up ramp-down	Increase of workers on the site compared with the number of workers in the previous month. Sudden increases in workers may indicate site congestion and a greater number of workers with reduced knowledge of the organization's safety practices
Worker's age	Percentage of workers younger than 30 and over 50 years old
Percentage of operators	Ratio of the number of equipment operators to workers
Near-miss rate	Total near misses, divided by the working hours in the same month, multiplied by 200,000[b] working hours per year

[a]BBO is a safety observation used to assess the behavior of a coworker. The assessment is anonymous, and workers being evaluated are not identified. During the assessment, a card is filled out.
[b]200,000 = 100 workers working 40 h per week for 50 weeks per year.

$$\text{Incident level}_n = \frac{\text{Safety Incident Rate}_n - \text{Safety Incident Rate min}}{\text{Safety Incident Rate max} - \text{Safety Incident Rate min}} \quad (2)$$

where incident level_n = incident level in month n; safety incident rate_n = saftey incident rate in month n; safety incident rate min = minimum safety incident rate; and safety incident rate max = maximum safety incident rate.

Then, the number of safety-related measures is reduced using a heuristic method, as detailed by Pereira et al. (2018b). Nine measures were considered in the final safety assessment model, namely behavior-based observation (BBO) rate, crew size, project elevation, temperature, schedule performance index (SPI), ramp-up ramp-down (RURD), workers' age (as percentage of workers under 30 and over 50 years old), percentage of operators, and near miss rate. Safety-related measures are detailed in Table 3.

In the fourth phase, data sets are used to develop the incident level assessment method. An artificial neural network (ANN), as detailed by Goh and Chua (2013), was used in this study due to its ability to (1) detect implicit and complex nonlinear relationships between dependent and independent variables, and (2) detect interactions or interrelationships among all input methods through hidden layers (Ayhan and Tokdemir 2019). Indeed, ANNs have been used in the construction industry (Kulkarani et al. 2017)—particularly in the safety field—to forecast safety climate (Patel and Jha 2014) and safety performance (Goh and Chua 2013). Although new machine-learning algorithms have been proposed, El-Abbasy et al. (2014) have contended that there is no specific machine-learning algorithm that outperforms others across all measurable aspects. Therefore, the ANN was considered a suitable method for this research.

In this research, a multilayered perceptron (MLP) was implemented, and a back-propagation algorithm was used to train the neural network. According to Hinton et al. (2006), a MLP can learn complex representations and perform any linear or multivariate arbitrary nonlinear computation in a manner capable of achieving desired accuracy. The neural network is composed of an input layer, multiple hidden layers, and an output layer (Fig. S2). The input layer consists of neurons that represent the safety-related measures previously selected. The number of hidden layers and neurons in each of these layers are defined based on trial and error (Goh and Chua 2013; Soibelman and Kim 2002). In this case study, the minimum root-means square method is used to select the optimum combination. The output layer is composed of a single neuron that represents the incident level. Based on the measures selected in Phase 3, the data set is divided into two groups, the training set (70% of the data) and validation set (30%). With this data set, two hidden layers consisting of eight neutrons each are selected. The final neural network has four layers, namely one input, two hidden layers, and one output layer. The neural network is further validated using a validation set, and the model's ability to determine safety management trends is verified using correlation analysis. Here, the Neuroph Library version 2.7 (Neuroph 2018) was used to build the MLP neural network.

As a result of the interoperability of the HLA approach, users can easily change the machine-learning algorithm in the safety performance analyses component to their preferred algorithm (e.g., use of case-based reasoning or linear regression models) without changing the structure of the other components.

Simulation Model

The simulation component simulates the actual performance of the construction project. A simulation model, based on that detailed by Pereira et al (2018b), was developed. Briefly, a list of project characteristics is received by the simulation model, which includes planned working hours, number of workers per month, and working hours per day at the beginning of each period (in this study, monthly). For each simulation time (i.e., each day), 100 randomly generated producitivty factors are sampled by the simulation model. Because the productivity of a project is affected by a variety of factors (e.g., weather conditions, site congestion, and material availability) and is not constant, productivity factors are randomly sampled from a predefined distribution that is defined by the user in the user interface. Specifically, 100 randomly generated productivity factors are sampled by the simulation model for each simulation time (i.e., each day). For each simulation period, 100 actual working-hour samples are considered. Actual working hours are determined by Eq. (3)

$$\begin{aligned} \text{Actual working hours} = {} & \text{Number of workers per month} \\ & \times \text{Working hours per day} \\ & \times \text{Productivity factor} \end{aligned} \quad (3)$$

At the end of the chosen period, two indices, SPI and RURD, are generated. Based on the 100 simulated actual working hours, the SPI and RURD are calculated using Eqs. (4) and (5), respectively, and then published

$$\text{SPI} = \frac{\text{Actual working hours}}{\text{Planned working hours}} \tag{4}$$

$$\text{RURD} = \frac{\text{Number of workers current month} - \text{Number of workers previous month}}{\text{Number of workers previous month}} \times 100 \tag{5}$$

Decision Support Interface

The decision support interface component acts as an interface that allows the simulator to control the simulation process. There are three main parts of this component: (1) input control, (2) results control, and (3) time management control. The input control, shown in Fig. 3, allows the simulator to define different parameters that will be used as input by the other simulation components, such as productivity factors and BBO rates. Because many of these inputs will be based on previous projects, the interface allows users to define parameters as probability distribution functions (PDFs). These probability distributions are then sampled and published to other simulation components at each time step.

The result control, shown in Fig. 4, collects results from other system components and displays as graphs the outputs versus time. Results include SPI, cost performance index (CPI), cumulative actual working hours, and planned working hours. Because the inputs are indeterministic, each graphs reports minimums, 25th, 50th, and 75th percentiles, as well as maximums.

The third component, time management control, is responsible for controlling the simulation time. It allows the simulator to advance the simulation time by a specified unit value (corresponding here to 1 real day), allowing the simulator to obtain live results and change input parameters throughout the simulation cycle to update project performance and test various scenarios.

Framework Results and Discussion

To demonstrate the functionality of the proposed framework, a case study was conducted using data collected by a construction organization for four industrial construction projects (Table 4).

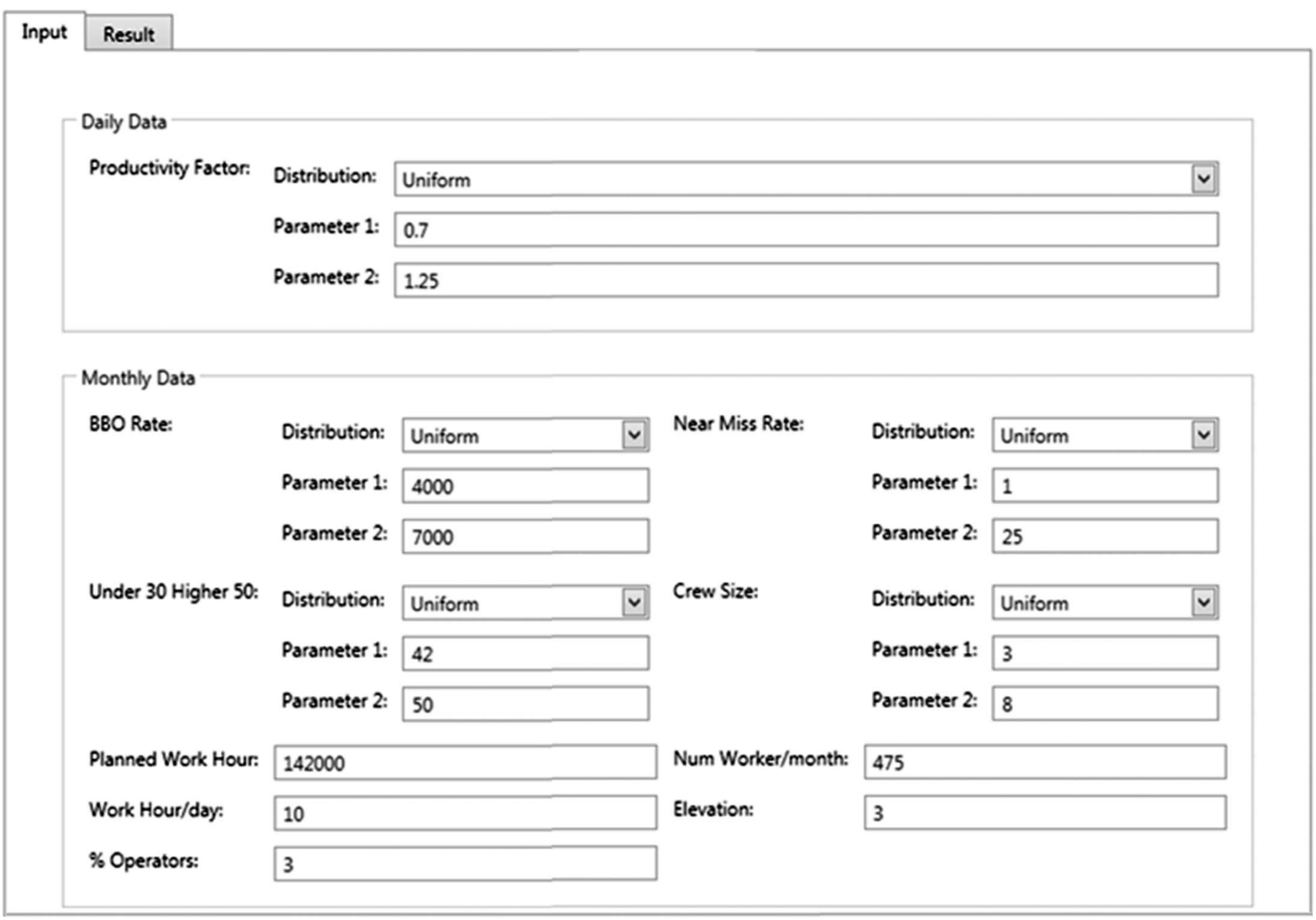

Fig. 3. Decision support interface input control.

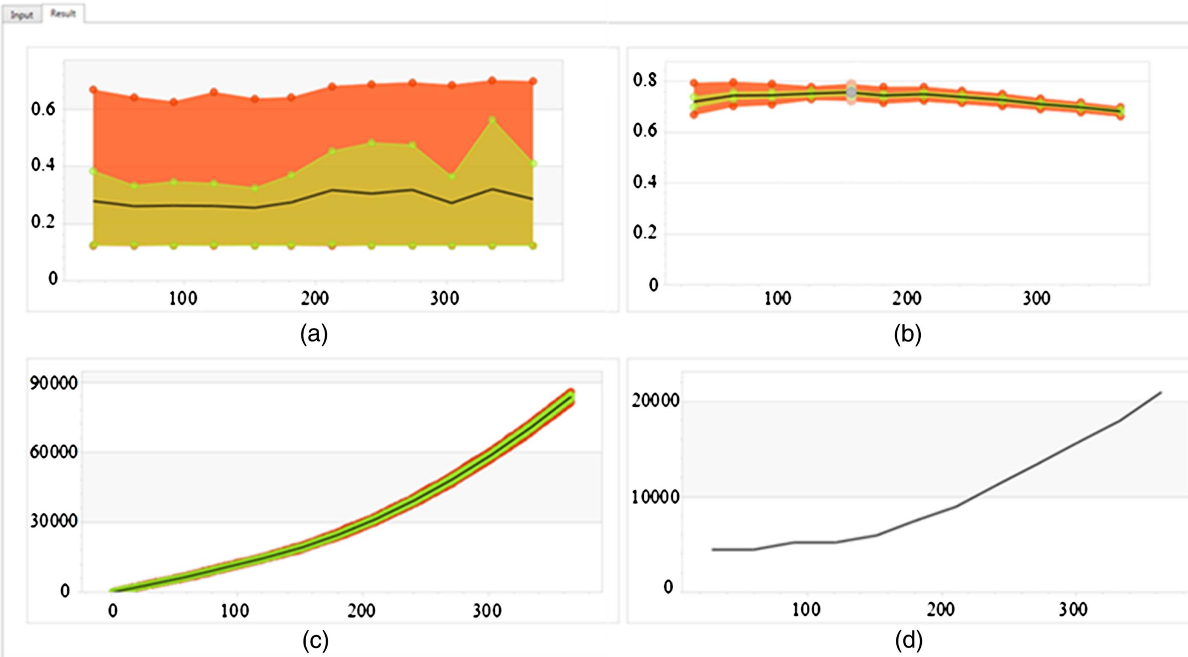

Fig. 4. Result control of the decision support interface for (a) SPI; (b) cost performance index; (c) cumulative actual working hours; and (d) planned working hours.

Table 4. Project data used in the case study

Project type	Duration (months)
Utility plant for heavy oil extraction facility	29
Construction of new steam-assisted gravity drainage facility	40
Expansion of existing potash mine and mill	42
Mechanical tank and pipe rack installation	24

All projects were performed in Canada, and each calendar month was considered a data point. Incidents levels were calculated using Eqs. (1) and (2). The first 5 months of execution of each project were excluded due to the low number of on-site workers. A total of 111 data points were collected.

After information was uploaded by each data warehouse component, as defined in Fig. 3, data for each project and month were grouped by the safety performance analysis component, which also examined the data for outliers. Similar to other studies in construction, the presented study identified points exceeding the 95th as outliers (Soleimanifar et al. 2014; Shrestha et al. 2014; Wibowo 2008). Using this method, six data points were considered outliers and were removed. Data points and the corresponding safety incident performance are presented in Fig. S1.

Khanzode et al. (2012) suggest that outliers may be due to unknown factors that influence the occurrence of an accident or that cannot be controlled (e.g., luck) (Mitropoulos et al. 2005). In this scenario, an incident did not occur where the risk was assessed as high on a construction site resulting in an atypical data point. Following the removal of the six outliers, the remaining data set ($n = 105$) was divided for building/testing [$n = 74$ (70%)] and validating [$n = 31$ (30%)] the ANN model.

Although many ANN models are able to predict a range of topics with comparatively high accuracy, this degree of accuracy is often not achieved with models designed to predict safety risk. However, statistical tests have been used to demonstrate that machine-learning algorithms are able to successfully predict safety performance trends (Patel and Jha 2014; Salas and Hallowell 2016; Tixier et al. 2016). As suggested by Salas et al. (2016) and Patel and Jha (2014), a correlation test was used in this study to validate the proposed ANN model. In social or behavioral sciences, a correlation coefficient value of 0.30–0.49 is typically interpreted as moderate to substantial evidence of an association, and 0.50–0.69 is interpreted as substantial to very strong evidence (De Vaus 2002). According to Salas and Hallowell (2016), if an analysis provides a strong relationship with the model, then it can be used for predictive purposes. Predicted versus actual incident levels are illustrated in Fig. 5.

The final ANN model resulted in a correlation value of 0.59 between the predicted and actual incident levels. Based on this result, the ability of the neural network to satisfactorily identify trends at the incident level was confirmed.

In addition to the correlation test, a sensitivity analysis was conducted to visualize the behavior of each leading indicator. According to Sterman (1988), a sensitivity analysis can be used to assess how reasonable changes in some uncertain or adjustable assumptions can affect the final output. The test procedure was as follows: all indicators were targeted at the 50th percentile while one indicator would change between the 15th, 25th, 50th, 75th, and 85th percentiles. Results from the sensitivity analysis are presented in Fig. 6.

Results obtained from the sensitivity analysis of the simulation-based analytics framework demonstrate that variations in the SPI had the greatest impact on the incident level. Several authors have discussed the influence of schedule performance on safety and, in particular, how delays can influence workers' ability to recognize (e.g., increase overtime working hours to mitigate delays) or to overlook (e.g., pressure to take shortcuts to enhance productivity) hazards (Han et al. 2014; Jiang et al. 2015; Mitropoulos et al. 2005). This result suggests that design of the project schedule should occur in consideration of a schedule's potential impact

J. Constr. Eng. Manage., 2020, 146(1): 04019091

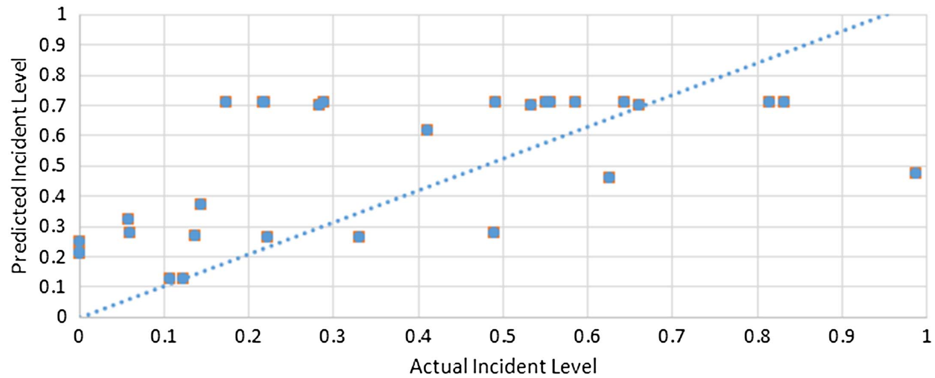

Fig. 5. Validation of neural network.

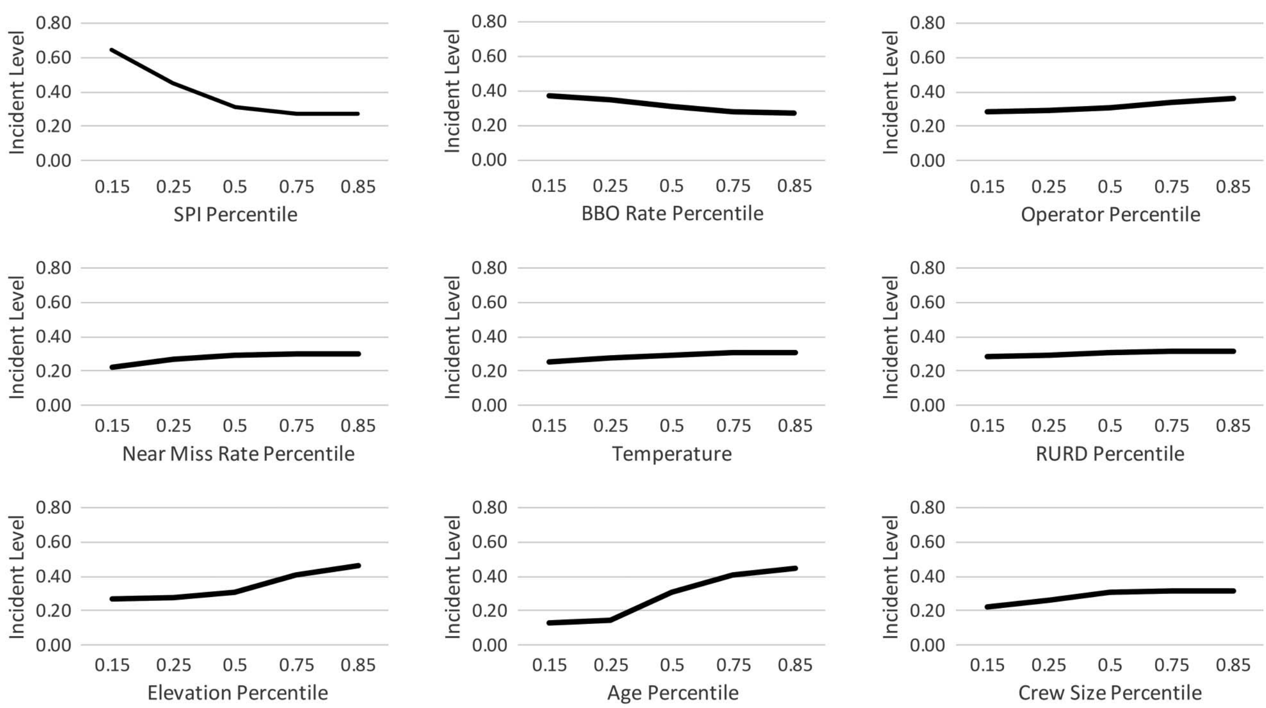

Fig. 6. Sensitivity analysis.

on safety. Indeed, involving safety managers in the project scheduling process may enhance the effectiveness of an organization's SMS.

Furthermore, variation in the percentage of workers younger than 30 or older than 50 (i.e., age) had the second greatest impact on incident levels. The result is supported by the findings of Bande and López-Mourelo (2015), who identified that workers older than 50 have a greater likelihood of being involved in incidents of high severity, whereas younger workers are associated with a higher frequency of lower-severity incidents. In addition, Lee et al. (2012) also identified that workers within this age range have a higher probability of being involved in an incident.

The sensitivity analysis results also demonstrated that (1) the influence of some of the indicators analyzed was reduced when the factors were within the lower (versus higher) 50th percentile, and (2) a holistic approach may allow for the generation of more comprehensive and representative results, in turn improving safety performance. For example, results demonstrated that increasing the BBO rate or controlling RURD after reaching the 50th percentile has little effect on safety performance. In this context, the sensitivity analysis may assist organizations in taking more efficient actions by reducing implementation costs and resources. Altogether, these findings can provide valuable insight, allowing practitioners to develop more effective organizational strategies capable of proactively controlling worksite incidents.

After the development and validation of the ANN assessment model, the simulation phase of the proposed approach was used to examine the predicted impact of various scenarios on the incident level. Three scenarios were established in consideration of the different decisions that project managers should consider during the project planning phase: (1) smoothing resource allocation during project planning (Scenarios 1 and 2), or (2) increasing resource allocation and the number of workers toward the end of the project (i.e., back-end loading; Scenario 3). Distributions for each indicator were established to identify how different strategies adopted by managers from different departments in an organization affected

Table 5. Project schedule characteristics of scenarios

	Worker quantity		Planned production		Operators	
Month	Scenarios 1 and 2	Scenario3	Scenarios 1 and 2	Scenario 3	Scenarios 1 and 2	Scenario 3
1	200	150	60,000	45,000	3	1
2	225	150	67,500	45,000	3	1
3	250	175	75,000	52,500	3	1
4	275	175	82,500	52,500	3	1
5	300	200	90,000	60,000	3	1
6	325	250	97,500	75,000	3	1
7	350	300	105,000	90,000	3	4
8	375	350	112,500	112,500	3	4
9	400	400	120,000	135,000	3	5
10	425	450	127,500	157,500	3	5
11	450	500	135,000	180,000	3	5
12	475	550	142,500	210,000	3	5

Table 6. Distributions of safety-related measures

Scenario	Workers' age	Productivity factor	BBO rate	Near-miss rate	Crew size
1	(42, 50)	(0.7–1.25)	(4,000–7,000)	(1–25)	(3–8)
2	(48, 58)	(0.5–0.9)	(500–3,000)	(10–50)	(4–10)
3	(42, 58)	(0.5–1)	(500–7,000)	(1–50)	(3–10)

the incident level. Tables 5 and 6 detail the characteristics of each scenario.

Project elevation was kept constant for all three scenarios, working hours per day was 10 h, and the project start date was set as September 1, 2018. The results for each scenario are presented in Fig. 7. The dark-gray area represents the 85th and 15th percentiles, the light-gray area represents the 75th and 25th percentiles, and the

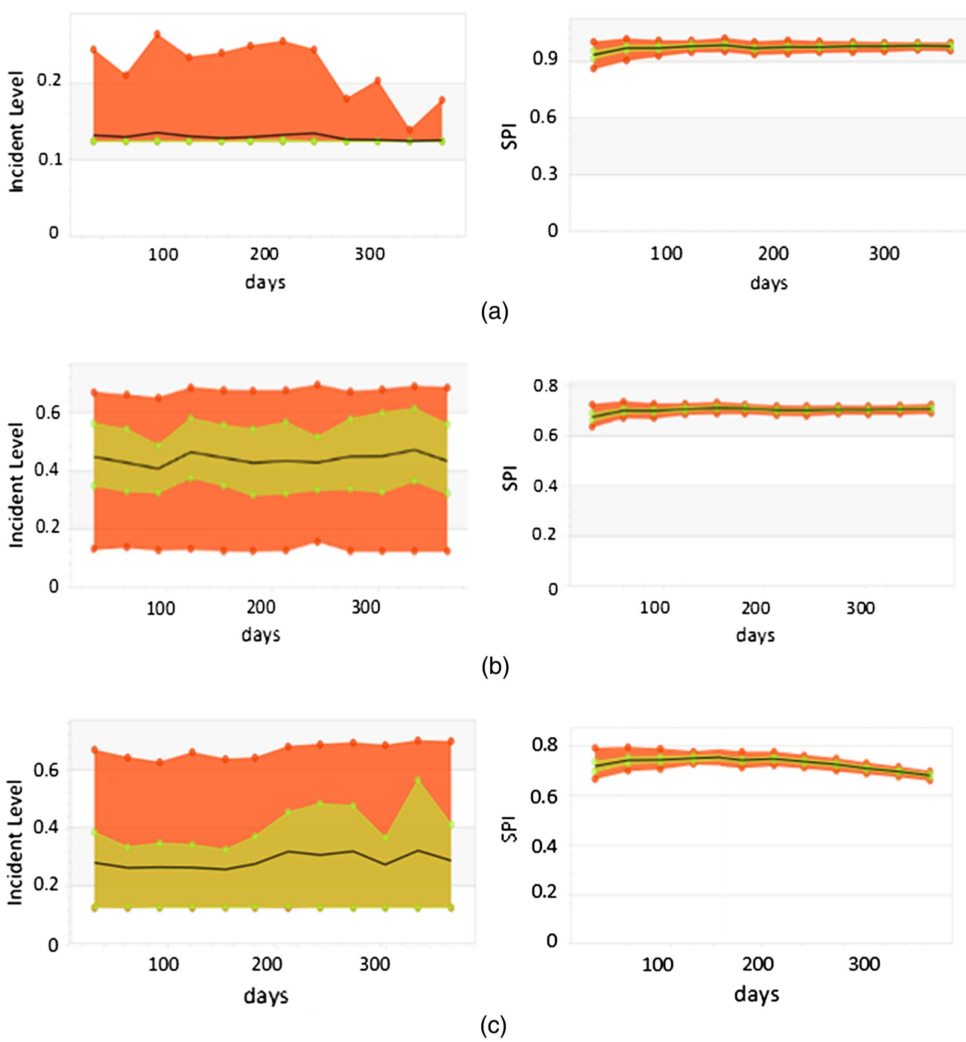

Fig. 7. Incident level and SPI results for (a) Scenario 1; (b) Scenario 2; and (c) Scenario 3.

solid line represents the average incident level for the scenario. As expected, input of smooth schedule planning and high control of the other indicators resulted in the lowest incident level (Scenario 1; average incident level ~0.1). Notably, the light-gray area is not visible because the variance between the average and the 15th or 25th percentile is minimal. Scenario 2 presented a safety incident level around 0.4 and an 85th percentile around 0.6, suggesting that controlling only the planning of the project schedule may not be sufficient to avoid incidents. As demonstrated by Han et al. (2014) and Lee et al. (2012), many safety attributes should be controlled concomitantly to improve safety performance. SPI in Scenario 2 was lower than in Scenario 1, which may influence worker behavior in practice.

Scenario 3 examined whether or not loading the back-end of a project schedule reduces the incident level. Although the average incident level was lower than that of Scenario 2 (due greater variance in indicators such as BBO rate and workers' age), Scenario 3 still resulted in an 85th percentile incident output that was similar to Scenario 2. This finding reinforces that the model can be used to test the impact of various scenarios, allowing managers to more effectively determine whether or not to take greater risks in the project (although, on average, there is a lower incident level than in Scenario 2).

Conclusion

Informed decision making is crucial for reducing unnecessary costs in construction. Increasing the likelihood of a success in a proactive approach to improve safety management could lead companies to reduce direct (e.g., site closure due to accident investigation) and indirect (e.g., costs associated with insurance and reputation of organization) project costs. Although use of historical data by organizations can assist in identifying safety measures and developing strategies to avoid the occurrence of accidents, the fragmented nature of data in the construction industry and the practical application of these data remains challenging. This study has developed a simulation-based analytics approach that allows construction organizations to make efficient use of existing databases, analysis tools, and simulation models to create comprehensive systems designed to enhance SMS practices and proactively control incident levels on construction projects. The application of a HLA approach was used to capture data from different sources and to allow communication between the ANN-based analysis module and the simulation component. The system was found capable of accurately forecasting incident-level trends, demonstrating that systems equipped with the distributed simulation–based approach can be used by the organization's practitioners to reliably test the impact of various scenarios on safety performance without the need for costly database redesigns.

Because many factors can contribute to the occurrence of a safety incident, of which many cannot be controlled or anticipated (Khanzode et al. 2012; Mitropoulos et al. 2005), it is difficult to predict safety performance of construction projects with the degree of accuracy that is typical of other situations. It is important to stress that this particular safety model is designed to predict the level of risk associated with particular project conditions—not the occurrence of an incident. As detailed by Khanzode et al (2012) and Hallowell et al. (2017), an incident is more likely to occur in high-risk conditions (e.g., a worksite with many new workers unfamiliar with the safety practices of the company under strict project schedule deadlines in severe weather) versus low risk conditions (e.g., a worksite with long-term employees and a low worker to foreperson ratio ahead of project schedule); however, an incident may not necessarily occur when the risk level is high.

The final ANN model resulted in a correlation value of 0.59 ($p < 0.001$) between the predicted and actual incident levels. Based on this result, the ability of the neural network to satisfactorily identify trends at the incident level was confirmed. As mentioned previously, although high-risk conditions may be present, an incident may not necessarily occur. This is consistent with the results observed in Fig. 5, where many of the data points were associated with a predicted incident level that exceeded the actual incident level.

The contribution of this research is twofold. From a practical perspective, the data-fusion capabilities provided by the proposed approach allow construction organizations executing projects with large data sets, such as industrial construction projects, to make more effective use of the vast quantity of existing historical data, data analysis tools, and simulation models available in literature and within their organization in a time-, cost-, and labor-effective manner. Specifically, results generated using the simulation based-analytics approach can be used to more efficiently allocate and prioritize safety resources. For instance, in the case study presented here, worker age (either younger or older workers) was associated with increased incident levels. Because age level was identified as an important factor in the current study, the company could focus on increasing safety training and awareness of new (i.e., young) workers and retraining for long-term (i.e., senior) workers. Alternatively, they may also adopt other risk-mitigation strategies, such as increasing the number of BBO cards filled, decreasing actual working hours (to reduce SPI), or reducing crew size to mitigate risks. The model can also be used (1) as a training tool for new practitioners to better understand how managerial decisions may affect safety performance, and (2) to foster a culture of safety consciousness across project agents by demonstrating that agents' decisions may influence the incident level.

From an academic perspective, this research has demonstrated the ability of a simulation-based analytics approach to efficiently integrate fragmented data. The approach was able to assess data in different locations without interfering with existing data-warehouse structures. Moreover, the approach enables the integration of additional simulation models to more comprehensively replicate project performance. The research results also demonstrate the ability of HLA to efficiently and successfully combine multiple simulation models and machine-learning algorithms. Furthermore, this approach allows for the intercommunication of different programming languages, which facilitates the future development and integration of additional components into the system.

Although the simulation-based analytics approach was able to successfully assess the incident level on an industrial construction project, the findings of this study should be applied in consideration of their limitations. First, the features used in this research were established from a single organization and, therefore, were limited to the data available in each department. Although the features used in this research were limited to those collected by the organization, results of the correlation test ($r = 0.59$ and $p < 0.001$ between the actual and predicted safety level) demonstrate that variables described in Table 2 can accurately predict the safety trends of the project. Nevertheless, features such as cost performance index and quality performance can be explored in further studies to supplement the final model. It is also important to consider that the focus of this study was on industrial construction; construction projects whose data-warehouse structures differ from those presented here may require model adjustments.

Second, although the ANN model was able to produce accurate results, increasing the number of data points and adding features are expected to enhance model predictability, consequently improving the decision-making process in the organization. Third, further research should be conducted to develop a deeper understanding of the holistic effect of each indicator on project performance (e.g., examine how features such as temperature or workers' age affect project productivity).

Based on the results of this research, several themes of future work for enhancing the simulation-based analytics framework can be performed. First, a new component can be developed to incorporate the 3D model design of the project. In this context, the manager will be able to visualize construction progress and analyze potential overlap between activities that can generate new hazards in the project. Second, the simulation-based analytics approach can be further expanded using data commonly associated with industrial construction projects to consider other aspects of project performance such as cost and quality. Third, a simulation model representing the behavior of each worker on the field (e.g., agent-based simulation model) can be developed to assess the incident level for each individual. Finally, further development of the model could allow practitioners to estimate the cost of changing safety policies in practice.

Data Availability Statement

Data analyzed during the study were provided by a third party. Requests for data should be directed to the provider indicated in the acknowledgments.

Acknowledgments

The authors are thankful for the support of PCL Industrial Management Inc. This research was made possible by the financial support of a Collaborative Research and Development Grant (CRDPJ 492657) from the Natural Sciences and Engineering Research Council (NSERC) of Canada. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSERC.

Supplemental Data

Figs. S1 and S2 are available online in the ASCE Library (www.ascelibrary.org).

References

Alavi, A. H., and A. H. Gandomi. 2017. "Big data in civil engineering." *Autom. Constr.* 79 (Jul): 1–2. https://doi.org/10.1016/j.autcon.2016.12.008.

ANSI (American National Standards Institute). 2012. *American national standard for occupational health and safety management systems*. ANSI Z10. Washington, DC: ANSI.

Ayhan, B. U., and O. B. Tokdemir. 2019. "Predicting the outcome of construction incidents." *Saf. Sci.* 113 (Mar): 91–104. https://doi.org/10.1016/j.ssci.2018.11.001.

Azimi, R., S. Lee, S. M. AbouRizk, and A. Alvanchi. 2011. "A framework for an automated and integrated project monitoring and control system for steel fabrication projects." *Autom. Constr.* 20 (1): 88–97. https://doi.org/10.1016/j.autcon.2010.07.001.

Bande, R., and E. López-Mourelo. 2015. "The impact of worker's age on the consequences of occupational accidents: Empirical evidence using Spanish data." *J. Labor Res.* 36 (2): 129–174. https://doi.org/10.1007/s12122-015-9199-7.

Bilal, M., L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, H. A. Owolabi, H. A. Alaka, and M. Pasha. 2016. "Big data in the construction industry: A review of present status, opportunities, and future trends." *Adv. Eng. Inf.* 30 (3): 500–521. https://doi.org/10.1016/j.aei.2016.07.001.

Dao, B., S. Kermanshachi, J. Shane, S. Anderson, and E. Hare. 2017. "Exploring and assessing project complexity." *J. Constr. Eng. Manage.* 143 (3): 04016126 https://doi.org/10.1061/(ASCE)CO.1943-7862.0001275.

Dean, J. 2014. *Big data, data mining, and machine learning: Value reaction for business leaders and practitioners*. Hoboken, NJ: Wiley.

De Vaus, D. A. 2002. *Analyzing social data science*. London: SAGE.

El-Abbasy, M. S., A. Senouci, T. Zayed, F. Mirahadi, and L. Parvizsedghy. 2014. "Artificial neural network models for predicting condition of offshore oil and gas pipelines." *Autom. Constr.* 45 (Sep): 50–65. https://doi.org/10.1016/j.autcon.2014.05.003.

Esmaeili, B., and M. Hallowell. 2013. "Integration of safety risk data with highway construction schedules." *Constr. Manage. Econ.* 31 (6): 528–541. https://doi.org/10.1080/01446193.2012.739288.

Goh, Y., and D. Chua. 2013. "Neural network analysis of construction safety management systems: A case study in Singapore." *Constr. Manage. Econ.* 31 (5): 460–470. https://doi.org/10.1080/01446193.2013.797095.

Guo, B. H. W., and T. W. Yiu. 2017. "Developing leading indicators to monitor the safety conditions of construction projects." *J. Manage. Eng.* 32 (1): 04015016. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000376.

Guo, B. H. W., T. W. Yiu, V. A. González, and Y. M. Goh. 2016. "Using a pressure-state-practice model to develop safety leading indicators for construction projects." *J. Constr. Eng. Manage.* 143 (2): 04016092. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001218.

Hale, A. 2009. "Why safety performance indicators?" *Saf. Sci.* 47 (4): 479–480. https://doi.org/10.1016/j.ssci.2008.07.018.

Hallowell, M. R., D. Alexander, and J. A. Gambatese. 2017. "Energy-based safety risk assessment: Does magnitude and intensity of energy predict injury severity?" *Constr. Manage. Econ.* 35 (1–2): 64–77. https://doi.org/10.1080/01446193.2016.1274418.

Han, S., F. Saba, S. Lee, Y. Mohamed, and F. Pena-Mora. 2014. "Toward an understanding of the impact of production pressure on safety performance in construction operations." *Accid. Anal. Prev.* 68 (7): 106–116. https://doi.org/10.1016/j.aap.2013.10.007.

Hinton, G., S. Osindero, and Y. W. Teh. 2006. "A fast learning algorithm for deep belief nets." *Neural Comput.* 18 (7): 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527.

Hinze, J., M. Hallowell, and K. Baud. 2013a. "Construction-safety best practices and relationships to safety performance." *J. Constr. Eng. Manage.* 139 (10): 04013006. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000751.

Hinze, J., S. Thurman, and A. Wehle. 2013b. "Leading indicators of construction safety performance." *Saf. Sci.* 51 (1): 23–28. https://doi.org/10.1016/j.ssci.2012.05.016.

Husted, K., and M. Snejina. 2002. "Diagnosing and fighting knowledge-sharing hostility." *Organizational Dyn.* 31 (1): 60–73. https://doi.org/10.1016/S0090-2616(02)00072-4.

IEEE Computer Society. 2010. *IEEE standard for modeling and simulation high level architecture (HLA): Framework and rules*. IEEE 1516. Piscataway, NJ: IEEE.

Jannadi, O. A., and M. S. Bu-Khamsin. 2002. "Safety factors considered by industrial contractors in Saudi Arabia." *Build. Environ.* 37 (5): 539–547. https://doi.org/10.1016/S0360-1323(01)00056-7.

Jiang, Z., D. Fang, and M. Zhang. 2015. "Understanding the causation of construction workers' unsafe behaviors based on system dynamics modeling." *J. Manage. Eng.* 31 (6): 04014099. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000350.

jOOL. 2018. "jOOλ." Accessed October 26, 2018. https://github.com/jOOQ/jOOL.

Khanzode, V. V., J. Maiti, and P. K. Ray. 2012. "Occupational injury and accident research: A comprehensive review." *Saf. Sci.* 50 (5): 1355–1367. https://doi.org/10.1016/j.ssci.2011.12.015.

Kulkarni, P.S., S. N. Londhe, and M. C. Deo. 2012. "Artificial neural networks for construction management: A review." *J. Soft Comput. Civ. Eng.* 1 (2): 70–88. https://doi.org/10.22115/scce.2017.49580.

Lee, H., H. Kim, M. Park, E. A. L. Teo, and K. Lee. 2012. "Construction risk assessment using site influence factors." *J. Comput. Civ. Eng.* 26 (3): 319–330. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000146.

Lingard, H., M. Hallowell, R. Salas, and P. Pirzadeh. 2017. "Leading or lagging? Temporal analysis of safety indicators on a large infrastructure construction project." *Saf. Sci.* 91 (Jan): 206–220. https://doi.org/10.1016/j.ssci.2016.08.020.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. 2011. *Big data: The next frontier for innovation, competition, and productivity.* New York: McKinsey & Company.

Marle, F., and L.-A. Vidal. 2016. *Managing complex, high risk projects: A guide to basic and advanced project management*. London: Springer.

Mitropoulos, P., T. S. Abdelhamid, and G. A. Howell. 2005. "Systems model of construction accident causation." *J. Constr. Eng. Manage.* 131 (7): 816–825. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:7(816).

Mohammadi, A., M. Tavakolan, and Y. Khosravi. 2018. "Factors influencing safety performance on construction projects: A review." *Saf. Sci.* 109 (2018): 382–397. https://doi.org/10.1016/j.ssci.2018.06.017.

Neuroph. 2018. "Neuroph: Java neural network framework." Accessed October 26, 2018. http://neuroph.sourceforge.net/index.html.

Patel, D. A., and K. N. Jha. 2014. "Neural network approach for safety climate prediction." *J. Manage. Eng.* 31 (6): 05014027. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000348.

Pereira, E., S. Han, and S. Abourizk. 2018a. "Integrating case-based reasoning and simulation modeling for testing strategies to control safety performance." *J. Comput. Civ. Eng.* 32 (6): 04018047. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000792.

Pereira, E., S. Han, S. Abourizk, and U. Hermann. 2017. "Empirical testing for use of safety related measures at the organizational level to assess and control the on-site risk level." *J. Constr. Eng. Manage.* 143 (6): 05017004. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001303.

Pereira, E., U. Hermann, S. Han, and S. Abourizk. 2018b. "Case-based reasoning approach for assessing safety performance using safety-related measures." *J. Constr. Eng. Manage.* 144 (9): 04018088. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001546.

Podgórski, D. 2015. "Measuring operational performance of OSH management system: A demonstration of AHP-based selection of leading key performance indicators." *Saf. Sci.* 73 (Mar): 146–166. https://doi.org/10.1016/j.ssci.2014.11.018.

Poh, C. Q. X., C. U. Ubeynarayana, and Y. M. Goh. 2018. "Safety leading indicators for construction sites: A machine learning approach." *Autom. Constr.* 93 (Sep): 375–386. https://doi.org/10.1016/j.autcon.2018.03.022.

Portico RTI. 2018. "The poRTIco project." Accessed October 26, 2018. http://www.porticoproject.org/.

Salas, R., and M. Hallowell. 2016. "Predictive validity of safety leading indicators: Empirical assessment in the oil and gas sector." *J. Constr. Eng. Manage.* 142 (10): 04016052. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001167.

Saleh, J. H., E. A. Saltmarsh, F. M. Favarò, and L. Brevault. 2013. "Accident precursors, near misses, and warning signs: Critical review and formal definitions within the framework of discrete event systems." *Reliab. Eng. Syst. Saf.* 114 (1): 148–154. https://doi.org/10.1016/j.ress.2013.01.006.

Saunders, L. W., B. M. Kleiner, A. P. McCoy, K. P. Ellis, T. Smith-Jackson, and C. Wernz. 2017. "Developing an inter-organizational safety climate instrument for the construction industry." *Saf. Sci.* 98 (Oct): 17–24. https://doi.org/10.1016/j.ssci.2017.04.003.

Shrestha, S., X. Fang, D. Wre, and W. C. Zech. 2014. "What should be the 95th percentile rainfall event depths?" *J. Irrig. Drain. Eng.* 140 (1): 06013002. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000658.

Soibelman, L., and H. Kim. 2002. "Data preparation process for construction knowledge generation through knowledge discovery in databases." *J. Comput. Civ. Eng.* 16 (1): 39–48. https://doi.org/10.1061/(ASCE)0887-3801(2002)16:1(39).

Soleimanifar, M., X. Shen, M. Lu, and I. Nikolaidis. 2014. "Applying received signal strength based methods for indoor positioning and tracking in construction applications." *Can. J. Civ. Eng.* 41 (8): 703–716. https://doi.org/10.1139/cjce-2013-0433.

Sterman, J. D. 1988. "Deterministic chaos in models of human behavior: Methodological issues and experimental results." *Syst. Dyn. Rev.* 4 (1–2): 148–178. https://doi.org/10.1002/sdr.4260040109.

Tixier, A. J.-P., M. R. Hallowell, B. Rajagopalan, and D. Bowman. 2016. "Application of machine learning to construction injury prediction." *Autom. Constr.* 69 (Sep): 102–114. https://doi.org/10.1016/j.autcon.2016.05.016.

US Department of Labor. 2017. "Commonly used statistics." Accessed October 30, 2018. https://www.osha.gov/oshstats/commonstats.html.

Wibowo, A. 2008. "Discussion of float consumption impact on cost and schedule in the construction industry." *J. Constr. Eng. Manage.* 134 (10): 828–829. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:10(828).

Wu, X., Q. Liu, L. Zhang, M. J. Skibniewski, and Y. Wang. 2015. "Prospective safety performance evaluation on construction sites." *Accid. Anal. Prev.* 78 (May): 58–72. https://doi.org/10.1016/j.aap.2015.02.003.