Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

WEARABLE SENSOR-BASED ACTIVITY RECOGNITION FOR DATA-DRIVEN SIMULATION OF CONSTRUCTION WORKERS' ACTIVITIES

Reza Akhavian

School of Engineering
California State University, East Bay
25800 Carlos Bee Blvd.
Hayward, CA 94542, USA

Amir Behzadan

Department of Technology and Construction Management
Missouri State University
901 S. National Ave.
Springfield, MO 65897, USA

ABSTRACT

Wearable technologies are becoming the main interface between human and surrounding environment for a variety of context-aware and autonomous applications. Ubiquitous, small-size, and low-cost smartphones carried by everyone nowadays are equipped with a host of embedded sensors that provide groundbreaking opportunities to collect and use multimodal data in data-driven decision support systems. Simulation models are one of the most widely used decision support tools in project management that can highly benefit from the integration of contextual knowledge with the model design. In this paper, a discrete event simulation (DES) model of construction operations involving human activities is designed, enriched with wearable sensor data using smartphones, and validated. The model parameters are defined using 1) a data-driven activity recognition and 2) a static engineering estimation method for comparison. Results show that the output of the data-driven simulation model is in a closer agreement with the values observed in the real system.

1 INTRODUCTION

The "high volume, high velocity, and high variety information assets" or Big Data (Gartner 2014) are finding their niche in coordinating urban technologies. Smart cities and internet of things, as two recent sensor-centric phenomena have been emerged to offer solutions for critical urban problems pertaining to energy efficiency, transportation planning, and risk mitigation. However, deployment of sensors, as the backbone of these innovative phenomena in design, planning, and construction of the smart cities' infrastructure has yet to be investigated by researchers in academia and industry (Suzumura and Kanezashi 2014).

Wearable sensors have been emerged during the recent years in order to collect data from processes in which humans are dominantly involved. Most of recent research studies aim at collecting data from human activities and behavior for medical, sport, and security applications (Cheng et al. 2010; Jia 2009). In all such studies, the goal is the recognition and classification of basic human activities. For instance, it is important for patients with heart disease or obesity to follow an exercise routine that can be readily identified using wearable motion detection sensors (Jia 2009). In a broader scheme, there are other attributes pertaining to human activities that may be of significance. For example, an important attribute (knowledge) is the time it takes for an individual to carry out a certain task. This piece of knowledge can be effectively extracted from data collected by wearable sensing devices providing that the sensory data collected from that individual is rich enough to help distinguish that task from its preceding and succeeding tasks.

In designing simulation models, the attributes of model elements should be defined in the model prior to running it. In discrete event simulation (DES), activities and resources (that travel between queues and activities) are considered the key model elements. Any change in their attributes can affect the overall performance of the model in representing the real system. Particular to highly complex, dynamic, and ever-changing operations such as those constituting most construction projects, activity durations are unique to each project and thus, rough estimations and predictions are not reliable enough in determining their values. Moreover, it is highly probable that attributes such as activity durations change over the course of a project with a hardly predictable pattern. Therefore, a systematic approach in extracting process knowledge from pervasive context-aware sensors and integrating the extracted knowledge with simulation model design can substantially help in validation and verification of the model.

The research presented in this paper is built upon the need for construction pervasive computing as outlined above and takes advantage of wearable technologies to extract contextual knowledge pertaining to the attributes of simulation model elements in construction projects. In particular, a DES model of a construction operation was designed in which construction workers were in constant interaction with each other. Different activities were carried out by these workers in an experimental setting. Each worker wore an armband that carried a smartphone using which data was collected. Streaming data from smartphone built-in accelerometer and gyroscope sensors was used to recognize construction workers' activities. As discussed in the next Section, the inertial measurement unit (IMU) that consists of accelerometer, gyroscope, and magnetometer is widely used to design human activity recognition systems for various applications within the field of computer science (Shoaib et al. 2015). Next, machine learning classifiers were employed to recognize activities and eventually extract activity durations. Probability distributions were then fit to the extracted duration values of multiple instances of each activity and used in the simulation model script. For result comparison purposes, another set of probability distributions were defined based on the engineering assumptions and estimations for those activities. A DES model was once designed based on the knowledge-based activity durations and then using estimated values. The output of the simulation models were then compared to the observed values in the real system to assess the model accuracy and improvement achieved by integrating sensory data into the simulation model.

2 RESEARCH BACKGROUND

DES has been used in various research fields such as material flow and supply chain management (Tannock et al. 2007; Wohlgemuth et al. 2006). Recently, due to the dynamic and complex nature of construction projects, some research studies attempted to design more realistic simulation models by collecting data from construction entities (Song and Eldin 2012; Vahdatikhaki et al. 2013; Zhang et al. 2013). However, such studies are mostly limited in scope as they have only targeted specific operations and used a single mode of data to extract contextual knowledge. The authors have previously investigated the design and implementation of a multi-modal data-driven simulation system for construction engineering and management (CEM) applications (Akhavian and Behzadan 2013, 2014b). Previously, the applicability and feasibility of a knowledge-based simulation model was demonstrated through the use of a wireless sensors network (WSN) to collect multiple modes of data such as positional, orientation, and payload. This study takes a further step in design and implementation of data-driven simulation models within the CEM domain using a more ubiquitous data collection scheme. In the designed methodology, smartphones as standalone self-sufficient data collection, storage, and transmission nodes are used to provide data for activity recognition. Such data acquisition setting is not vulnerable to ambient factors and challenges often present in real jobsites such as dust and weather conditions that require frequent maintenance and calibration of the sensors. In addition to their value to human activity recognition, smartphones can be also placed in heavy equipment cabins for construction equipment activity recognition (Akhavian and Behzadan 2014a).

A great deal of research in pervasive computing within the domain of computer science aims at recognizing daily human activities using smartphone built-in sensors. Miluzzo et al. (2008) conducted one of the earliest studies in this context and discussed important design decisions to resolve corresponding limitations. More recent studies adopted methodologies that are, in essence, similar by employing smartphone sensors to classify human activities using machine learning classification algorithms (Bayat et al. 2014; Khan et al. 2014; Martín et al. 2013; Thiemjarus et al. 2013). In all such studies, however, data collection was performed for classification of routine daily activities such as walking, standing, jogging, running, climbing up- and down-stairs, and biking. However, the process of activity recognition in a construction jobsite is relatively more challenging. That is essentially due to the more degrees of freedom and discretion each worker has while performing his or her tasks. Furthermore, workers' interactions with each other, material, and equipment coupled with the underlying complexity of field tasks make the recognition process even more complicated. In this study, three different processes are investigated each containing activities with different natures to evaluate the performance of an activity recognition system for activity duration extraction towards more accurate simulation input modeling.

3 THE OPERATION EXPERIMENT DESIGN

The goal of the operation that was replicated in this research was to prepare, transport, and install wood sections in a full-scale outdoor experimental setting that resembled a real construction jobsite. Figure 1 shows a snapshot of the experiment. As shown in this Figure, the cyclic operation starts with a worker, *W1*, who saws lumber inside an imaginary *wood workshop* and prepare wood sections of proper sizes and shapes. These sections are then transported to the *installation area* by two other workers, *W2* and *W3* who are tasked with loading the sections into wheelbarrows, pushing wheelbarrows to the installation area, and dumping the sections where an installer worker, *W4*, is waiting to receive the sections and install them in their positions. Also, Figure 2 shows a snapshot of the accelerometer and gyroscope data on the smartphone's data collection and logging application that was used in the experiment.

Figure 1: Snapshot of the operation showing four workers performing the experiment.

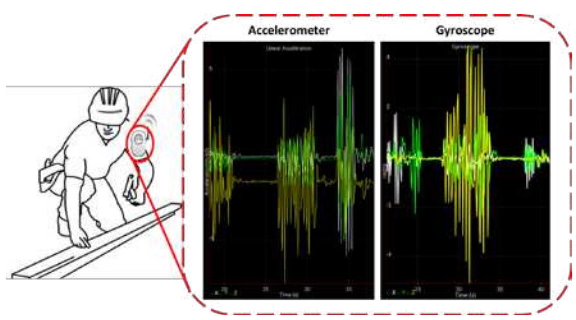

Figure 2: Different body motions create distinctive patterns in accelerometer and gyroscope data.

Each process involves one or more activities assigned to different workers. Here, the intention is to explore the generalizability of the developed framework by evaluating the accuracy of recognizing activities with different movement patterns. In the *wood workshop*, the process of cutting the lumber pieces consists of only one activity, *sawing*, carried out by worker *W1*. The transportation process involves four activities namely putting sections into the wheelbarrow or *loading*, pushing a loaded wheelbarrow or *pushing*, dumping the sections in the installation area or *unloading*, and returning the empty wheelbarrow or *returning*. Workers *W2* and *W3* are responsible for the transportation process. Finally, worker *W4* is tasked with the installation process which involves the activities *hammering* and *turning the wrench*. All the aforementioned processes, activities, and tasked workers are summarized in Table 1. The *Loading* and *unloading* activities follow underlying operational rules that are enforced in the experiment and later in the simulation models. These rules are as follows:

1. The *loading* activity will not be executed until there are at least two wood sections available for transportation. Therefore, when there are less than two sections prepared by worker *W1*, and either or both workers *W2* and *W3* are available, they will wait in a queue until at least two sections are ready for loading. With the same token, if either or both workers *W2* and *W3* are available, one section should wait until there is at least one more section prepared by *W1* so that both sections can be loaded.
2. For *unloading* activity, it is assumed that the space available for unloaded sections is enough only for two sections and *unloading* activity should be executed in only one instance, meaning that if there is any section waiting to be processed by worker *W4*, the available workers *W2* or *W3* should wait until there is no section awaiting installation process.
3. For *loading* and *unloading* activities, only one instance of each activity can be performed at any given time, meaning that simultaneous execution of either *loading* or *unloading* activity is not allowed.

Table 1: List of the processes involved in the operation and activities within each process.

Process	**Activity**	**Worker**
Cutting Lumber	Sawing	W1
Transportation	Loading	W2 & W3
	Pushing	
	Unloading	
	Returning	
Installation	Hammering	W4
	Turning the Wrench	

4 SIMULATION MODEL OF THE OPERATION

The operation described in Section 3 was carefully modeled in Stroboscope (STate and ResOurce Based Simulation of COnstruction ProcEsses), a DES scripting environment based on Activity Cycle Diagrams (ACDs) that is designed for the simulation of processes common to construction engineering (Martinez 1996). Simulation models created in Stroboscope are based on a network of interconnected modeling elements described in a script containing programming statements that give the elements unique behavior and control the simulation (Martinez and Ioannou 1994). This network of the interconnected elements (a.k.a. the ACD) is designed to be similar in appearance and function to CYCLONE simulation platform, which was the first system developed specifically for construction operations (Halpin 1977). The ACD of the operation described in Section 3 is shown in Figure 3. In this Figure, resources move from each node to the succeeding node in the direction shown by the connection link. A circle with a slash in the bottom right corner is a *Queue* that serves as the storage location for the resources. A rectangle with a cut-off in the top-left corner is called a *Combi* and a regular rectangle is called a *Normal*. These two nodes represent two different types of activities and hold the resources for the amount of time determined by activity durations. In particular, a *Combi* is always preceded by a *Queue* while a *Normal* activity cannot be preceded by a *Queue*. In Figure 3, `LumbersWait` holds lumber pieces before they are taken by worker *W1* for activity `Sawing`. The `WorkerW1Wait` Queue populated with 1 entity (i.e. 1 worker) ensures that only one instance of the `Sawing` activity is carried out in any point of time. Upon being sawed, sections wait in `SectionsWaitI` Queue to be loaded for transportation. This Queue satisfies the first operational rule described in Section 3. The `WorkersW2&W3WaitII` Queue is where Workers W2 and W3 are drawn from one by one to load only two sections, if available. Similar to `SectionsWaitI,` this Queue also contributes to satisfying the first operational rule. When enough sections and transportation workers are available, the `Loading` Combi is activated, lasts for its assigned duration, and then releases the captured resource (i.e. worker) to the `Hauling` Normal. Again, this activity will hold the resource for the amount of time determined by its corresponding duration. Next, according to the second operational rule in Section 3, Workers *W2* and/or *W3* wait in the `WorkersW2&W3WaitI` Queue before the space is available for activation of the `Unloading` Combi. Finally, the `SectionsWaitII` Queue is where at most two sections are being held before they can proceed to the `Hammering` Combi. It must be noted that the `Hammering` Combi will not be activated if either of the `SectionsWaitII` or `WorkerW4Wait` Queues does not have available resources. Such situation happens for example if worker W4 is captured by the `TurningtheWrench` Normal.

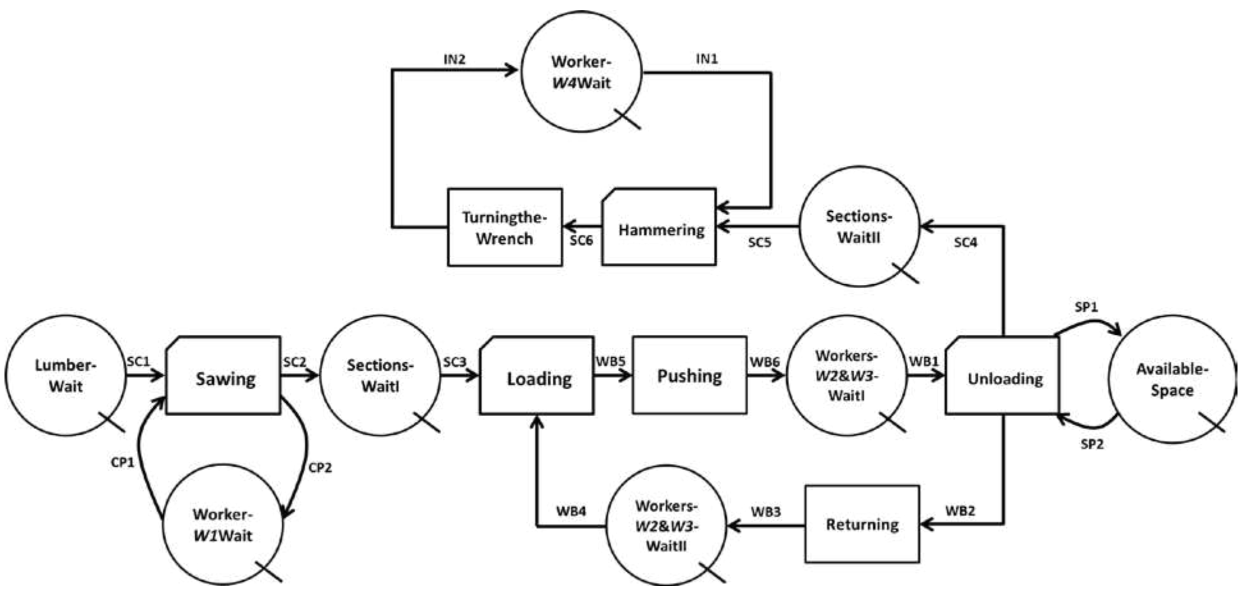

Figure 3: The ACD of the operation for modeling in Stroboscope.

While the ACD shown in Figure 3 provides a high level representation of the simulated operation, more specific operational details are incorporated in the script of the model. This is where attributes of the queues and activities as well as the model parameters are assigned. Such attributes define how model parameters behave. For example, the following sample lines from the model script show how some of the network elements inducing activities (Normal and Combi), Queues, and links are defined in Stroboscope:

```
/* Definition of Network Elements
QUEUE  SectionsWaitI       Sections;
COMBI  Loading;
NORMAL Pushing;
```

```
QUEUE  WorkersW2&W3WaitI   Workers;
COMBI  Unloading;
LINK   CP1    WorkerW1Wait     Sawing;
LINK   CP2    Sawing           WorkerW1Wait;
LINK   SC1    LumberWait       Sawing;
LINK   SC2    Sawing           SectionsWaitI;
```

Another key attribute is the durations of Combi and Normal activities. Activity durations are sampled from the specified probability distributions. In the next Section, the activity recognition framework developed in this research in order to extract realistic activity durations is described.

5 DURATION EXTRACTION THROUGH ACTIVITY RECOGNITION

In this study, data are collected using mobile phone accelerometer and gyroscope sensors. Collected raw sensory data are segmented into windows containing certain number of data points. Next, key statistical features are calculated within each window. Furthermore, each segment is labeled based on the corresponding activity class performed at the time identified by the timestamp of the collected data. In order to train a predictive model, supervised classifiers were used to recognize activities performed in the experiment. Details of data collection and preprocessing configurations are presented in Table 2.

Table 2: Sensory data collection configurations used for activity recognition.

Configurations	Mechanism or Values Used
Sampling Frequency	100 Hz for both accelerometer and gyroscope
Data Preparation	Interpolating missing data and removing data with close timestamp
Window Size	128 data points with 50% overlap
Extracted Features	Statistical time- and frequency domain using fast Fourier transform

Five supervised machine learning classifiers, namely, neural network, decision tree, k-nearest neighbor (KNN), logistic regression, and support vector machine (SVM) were trained and tested using 10-fold stratified cross validation. The specifications of these machine learning algorithms and classification details is outlined in a previous study by the authors to recognize construction equipment activities and can be found in (Akhavian and Behzadan 2015). Table 3 shows the classification accuracy results for individual classifiers for each of the processes involved in the experiment operations.

Table 3: Classification accuracy (%) of activities in each of the three processes using the five classifiers.

	Neural Network	Decision Tree	KNN	Logistic Regression	SVM
Cutting Lumber	96.27	95.58	96.22	96.54	96.64
Transportation	88.17	85.62	87.68	85.84	78.34
Installation	87.78	78.57	87.73	82.23	82.18

As tabulated in Table 3, neural network outperforms the other four classifiers in terms of overall classification accuracy, while KNN is closely following neural network in all three categories. Therefore, in order to incorporate both classifiers for potential improvements in classification accuracies, an ensemble methodology is also adopted. Bootstrap aggregation or *Bagging* is the ensemble algorithm used in this research. Using this algorithm, T training data subsets each containing m training examples are selected randomly with replacement from the original training set of m examples. The classification result of the ensemble is determined through plurality voting (Lin et al. 2003). Here, the number of training dataset is T

= 20. Classification was performed on the activity level within each process, meaning that the result of the classification in terms of accuracy in correctly predicting the activities within each process is reported. It is worth mentioning that within each class, an extra activity is included as the *idling* state in which the worker is not contributing to any of the assigned activities within the process.

Table 4 shows the Bagging ensemble result of classification accuracies for each process. As seen in Table 4, the classification accuracy results have been improved using the Bagging ensemble model. The accuracy of activity recognition for the first process involving the *sawing* activity is almost perfect and it is expected that it closely matches the observed activity durations. However, activities that comprise the other two processes, namely transportation and installation have not been classified as accurately, although more than 90% accuracy was achieved. Therefore, the durations extracted from these activities are expected not to be as close to the observed durations as the first process. However, it should be noted that the similarity of the extracted durations to the observed values does not necessarily conform the same accuracy as their associated activity recognition accuracy. In other words, although it is expected that the durations of activities within the cutting lumber process is predicted with the highest accuracy of all, the accuracy of predicting activity durations for the transportation and installation processes may not follow the same results in terms of relative accuracies. This is due to the fact that extracting activity durations follows a heuristic algorithm according to which many of the misclassified instances are ignored. In essence, the algorithm first replaces instances of any different classes that are appeared within a large number of detected instances of the same class. For example, few instances of class C_2 classified after many instances of class C_1 followed by other instances of class C_1 are considered as class C_1. The exact numbers followed by this heuristic algorithm depends on the sampling frequency, window size, and rough approximation of the activity durations. Here with sampling frequency of 100 Hz, window sizes of 128 data points with 50% overlap that amounts to 0.64 seconds of data, any two instances of an activity that normally takes more than 20 seconds but are separated out to less than 12 seconds are merged. Such heuristics result in improved accuracy for activity duration extraction. It should be noted that these result are derived from testing the model with the data that was collected from the same jobsite from which training data was collected. Therefore, further experiments are required to obtain subject-independent results.

Table 4: Bagging classification accuracy (%) for recognizing activities within each process.

Process	**Accuracy (%)**
Cutting Lumber	99.28
Transportation	90.09
Installation	92.97

6 SIMULATION INPUT MODELING

In this Section, the process of input modeling of the operation simulation is described. Simulation input modeling includes fitting probability distribution functions to the activity durations and has a high impact on the accuracy of the model. First, observed activity durations using the recorded videotape of the experiment are compared to those extracted through the activity recognition system. This step serves to guarantee that extracted activity durations are not statistically significantly different from those that actually took place in the real experiment. If there is a statistically significant difference between the two sets of duration values, then it cannot be expected from the data-driven simulation model to output values close to the actual ones observed in the experiment.

In order to compare observed and extracted activity durations, the *student t-test* is used to evaluate the null hypothesis of no considerable difference between the expected and sample distributions. Table 5 shows the result of the t-test for activities within each process. As shown in this Table, the null hypothesis for none of the activities was rejected through comparison of the observed and extracted activity durations with 5%

significance level. This confirms that the two sets of activity durations are not statistically significantly different.

Table 5: Comparison of the observed and extracted activity durations using student t-test.

Process	**Activity**	**Observed Duration (sec.)**		**Extracted Duration (sec.)**		**p-value**	**Null Hypothesis**
		Mean	**SD**	**Mean**	**SD**		
Cutting Lumber	Sawing	27.95	6.50	27.97	6.57	0.78	Not rejected
Transportation	Loading	8.96	1.40	9.24	1.75	0.27	Not rejected
	Pushing	14.02	2.43	14.14	2.81	0.63	Not rejected
	Unloading	13.18	1.96	13.53	2.01	0.08	Not rejected
	Returning	11.33	2.14	11.39	2.29	0.78	Not rejected
Installation	Hammering	17.05	2.48	17.59	2.46	0.09	Not rejected
	Turning the Wrench	13.39	3.42	13.44	3.35	0.75	Not rejected

The objective of creating simulation models of the operation experiment is to compare the results of the simulation created based on the extracted activity durations (data-driven model) to the one created according to the estimated activity durations (static model). To this end, estimated activity durations were defined by taking into account the [minimum, maximum], or three-point estimation [minimum, mode, maximum] durations for each activity which is a common practice in creating construction simulation models or project management schedules using project evaluation and review technique (PERT) (Halpin and Riggs 1992). These two schemes are in essence equivalent to sampling from uniform and triangular distributions. Therefore, these two probability distributions were considered for activity durations inside the static model. The parameters of the two probability distributions however were estimated according to two heuristics; the instructions given to the workers performing the activities, and engineering assumptions of the variance for such durations considering the nature of each activity. For example, worker *W1* was asked to saw each piece of lumber for about 25 to 30 seconds. Therefore, the probability distribution considered for this activity was a uniform distribution with a minimum of 22 and maximum of 33 to account for 3 seconds of variations from the extrema. For the extracted durations, Kolmogorov–Smirnov and Chi-Square goodness-of-fit (GoF) tests were used to find the best distribution fit to durations of instances for each activity according to the both test statistics. More details about the GoF tests and their applications in data-driven simulation can be found in (Akhavian and Behzadan 2014b). Table 6 shows the probability distributions fitted to the extracted activity durations along with those estimated for each activity. The following sample lines show how extracted activity durations are defined inside Stroboscope:

```
DURATION  Sawing             'Triangular[12,31.6,40]';
DURATION  Loading            'Triangular[6,8.1,13]';
DURATION  Pushing            '9 + Gamma[1.63, 3.16]';
DURATION  Unloading          '9 + 8 * Beta[1.83, 1.41]';
DURATION  Returning          '6 + Gamma[1.26, 4.28]';
DURATION  Hammering          'Normal[17.8,2]';
```

7 PERFORMANCE OF THE DATA-DRIVEN VS. STATIC SIMULATION MODEL

Using the two sets of probability distributions shown in Table 6, two identical simulation models are created based on the ACD introduced in Section 4. The only difference between the two simulation models is in the activity durations defined in the input script of each model.

Table 6: Probability distributions used inside the two simulation models.

Activity	Probability Distributions Used for	
	Extracted Duration (Data-Driven Model)	**Estimated Durations (Static Model)**
Sawing	Triangular[12,31.6,40]	Uniform[22, 33]
Loading	Triangular[6,8.1,13]	Uniform[5, 7]
Pushing	9 + Gamma[1.63, 3.16]	Uniform[8, 12]
Unloading	9 + 8 × Beta[1.83, 1.41]	Uniform[7, 12]
Returning	6 + Gamma[1.26, 4.28]	Uniform[7, 10]
Hammering	Normal[17.8,2]	Triangular[13,15,17]
Turning the Wrench	7 + 14 × Beta[1.62, 1.78]	Triangular[13,15,17]

Each model was run for 50 replications by generating random numbers from the same seed, and five measures were collected for comparison of the outputs of the simulations to the real world observations. The measures include the average waiting times (in seconds) of the entities in the four Queues namely SectionsWaitI, SectionsWaitII, WorkersW2&W3WaitI, and WorkersW2&W3WaitII, as well as the total operation time (in minutes). Figure 4 shows the comparison between these five measures. For each measure shown in this Figure, the first bar from the top refers to the value of that measure observed in the real-world operation. The second bar with a slightly lighter color corresponds to the mean of the average waiting times resulted from the data-driven simulation after 50 replications. The error bar refers to the standard deviation of these 50 replications. The third bar with the lightest color is the result of the static simulation created based on the estimated activity durations with the error bar that refers to the standard deviation.

8 DISCUSSION OF THE RESULTS

According to Figure 4, the observed values for all five measures are within one standard deviation of the results obtained from the data-driven simulation model. This is while all the output measures obtained from the static simulation with estimated values are underestimating the waiting times and total duration of the operation. In fact, this is what happens most of the time in construction projects where simulation models created in the planning and pre-construction stage estimated significantly underestimate or overestimate the durations of the real world processes (Halpin and Riggs 1992). This is while uniform and triangular probability distributions (and not fixed values) were used for estimating activity durations inside the static model. It must be noted, however, that the underestimation observed in the output of the static simulation model in all five measures is particular to this specific example and cannot be generalized to other problems. More specifically, the measures obtained from the static model could have as well resulted in an overestimation. What is of outmost importance in interpreting the results is the noticeable difference between the outputs of the two simulation models and the fact that the result of the data-driven model is closer to real-world observations.

A considerable discrepancy can be seen in the result obtained from the static simulation and the observed value for the average waiting time in Queue `SectionsWaitI`. This can be explained as follows; since the `WorkersW2&W3WaitII` average waiting times in bar chart (d) are very close (considering the scale of this chart), the availability of workers should not have influenced the difference. Therefore, it can be explained through the difference in durations considered for the `Sawing` activity. It turns out that the data-driven simulation with the probability distribution of Triangular [12,31.6,40] for `Sawing` activity, samples from a lower range of numbers starting from 12 seconds, while the minimum value for the uniform distribution of the observed values is 23 seconds. This results in a much faster `Sawing` in reality which in turn provides more sections waiting in the `SectionsWaitI` Queue. Other than `Sawing`, most of the other activities have estimated distributions resulting in sampling of lower values.

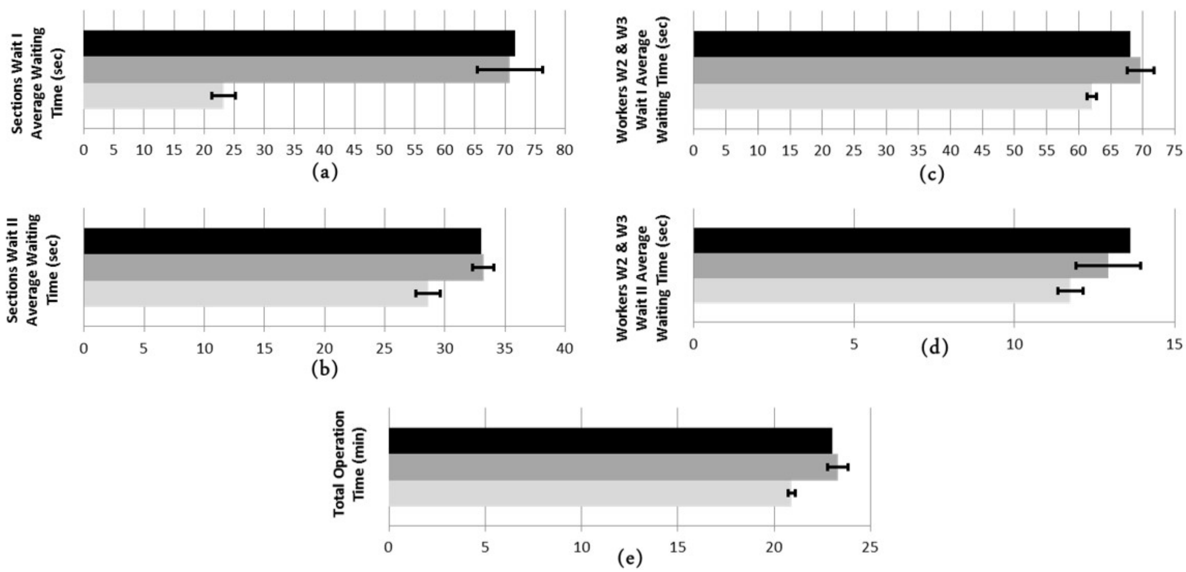

Figure 4: Comparison of results obtained from the real-world experiment, and the output of static and data-driven simulation models.

Regardless of the reasons for any discrepancy between the extracted and estimated activity durations, the very fact that any difference in activity durations can substantially change the simulation output statistics verifies the significance of having more realistic simulation models through data-driven input modeling.

9 SUMMARY AND CONCLUSIONS

In this paper, a complex operation involving multiple interactions between human workers performing construction activities was described and modeled in DES using process-level data collected from the crew in real time. Sensory data consisted of accelerometer and gyroscope data and were collected using smartphones affixed on workers' upper arms. Activities performed by the workers were then recognized and classified using the supervised machine learning classifiers. Following activity recognition, corresponding activity durations were extracted and probability distributions were fit to the extracted durations. Moreover, these durations were compared to the values observed in the real world experiment to confirm their fidelity. Extracted activity durations were then fused into a data-driven DES model created based on the experiment design in order to compare the results against those of a similar but static simulation model with estimated values for activity durations.

Analysis of the output obtained from the two simulation models with respect to five quantifiable measures (i.e. average waiting times of the entities in four Queues namely `SectionsWaitI`, `SectionsWaitII`, `WorkersW2&W3WaitI`, and `WorkersW2&W3WaitII`, as well as the total operation time) revealed that the data-driven simulation model created based on the knowledge (i.e. activity durations) extracted by the developed activity recognition framework outperforms the static simulation model created based on estimated activity durations. Considering the fact that often times the common practice in creating construction simulation models is using historical (secondary) information and subjective assumptions in designing model attributes, obtaining results in close agreement with reality reaffirms the significance of substituting this traditional approach in creating simulation models with a more robust and reliable data-driven and knowledge-based methodology that was described in this paper.

10 FUTURE WORK

Future work of this study includes incorporating positional data using smartphone built-in global positioning system (GPS) sensors to further improve the accuracy of activity recognition and enrich the extracted contextual knowledge. Another potential direction for future work in this area will be to explore whether the results achieved so far can be used for *automatically* extracting process knowledge such as activity durations and precedence logic for the purpose of ubiquitously updating and maintaining simulation models in true real time. Another branch of future work is automated identification and analysis of unsafe workers' postures in physically demanding construction activities. Work-related Musculoskeletal Disorder (WMSD), back, knee, and shoulder injuries are among the most common injuries that can be prevented or reduced by complying with Occupational Safety and Health Administration (OSHA) or the National Institute for Occupational Safety and Health (NIOSH) standards and rules (NIOSH 2015; OSHA 1990).

REFERENCES

Akhavian, R., and A. H. Behzadan. 2013. "Knowledge-Based Simulation Modeling of Construction Fleet Operations Using Multimodal-Process Data Mining." *Journal of Construction Engineering and Management* 139 (11):04013021.

Akhavian, R., and A. H. Behzadan. 2014a. "Construction Activity Recognition for Simulation Input Modeling Using Machine Learning Classifiers." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 3296-3307. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Akhavian, R., and A. H. Behzadan. 2014b. "Evaluation of Queuing Systems for Knowledge-Based Simulation of Construction Processes." *Automation in Construction* 47:37-49.

Akhavian, R., and A. H. Behzadan. 2015. "Construction Equipment Activity Recognition for Simulation Input Modeling Using Mobile Sensors and Machine Learning Classifiers." *Advanced Engineering Informatics*.

Bayat, A., M. Pomplun, and D. A. Tran. 2014. "A Study on Human Activity Recognition Using Accelerometer Data from Smartphones." *Procedia Computer Science* 34:450-457.

Cheng, J., O. Amft, and P. Lukowicz. 2010. "Active Capacitive Sensing: Exploring a New Wearable Sensing Modality for Activity Recognition." *Pervasive Computing*, 319-336. Springer.

Gartner. 2014. "It Glossary." Accessed April 30. http://www.gartner.com/it-glossary/big-data/.

Halpin, D. W. 1977. "Cyclone - a Method for Modeling Job Site Processes." *ASCE Journal of Construction Division* 103 (3):489-499.

Halpin, D. W., and L. S. Riggs. 1992. *Planning and Analysis of Construction Operations*. New York: Wiley. http://www.loc.gov/catdir/description/wiley033/91032231.html
http://www.loc.gov/catdir/toc/onix04/91032231.html.

Jia, Y. 2009. "Diatetic and Exercise Therapy against Diabetes Mellitus." *Second International Conference on Intelligent Networks and Intelligent Systems, 2009. ICINIS'09.*

Khan, A. M., A. Tufail, A. M. Khattak, and T. H. Laine. 2014. "Activity Recognition on Smartphones Via Sensor-Fusion and Kda-Based Svms." *International Journal of Distributed Sensor Networks* 2014.

Lin, X., S. Yacoub, J. Burns, and S. Simske. 2003. "Performance Analysis of Pattern Classifier Combination by Plurality Voting." *Pattern Recognition Letters* 24 (12):1959-1969.

Martín, H., A. M. Bernardos, J. Iglesias, and J. R. Casar. 2013. "Activity Logging Using Lightweight Classification Techniques in Mobile Devices." *Personal and Ubiquitous Computing* 17 (4):675-695.

Martinez, J., and P. G. Ioannou. 1994. "General Purpose Simulation with Stroboscope." In *Proceedings of the 1994 Winter Simulation Conference*, edited by J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 1159-1166. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Martinez, J. C. 1996. "Stroboscope: State and Resource Based Simulation of Construction Processes." thesis, University of Michigan, Ann Arbour, MI.

Miluzzo, E., N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. 2008. "Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the Cenceme Application." In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*.

NIOSH. 2015. "Safety & Prevention." Accessed March 30, 2015. http://www.cdc.gov/niosh/topics/safety.html.

OSHA. 1990. "Excavation Final Rule." Safety and Health Administration.

Shoaib, M., S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga. 2015. "A Survey of Online Activity Recognition Using Mobile Phones." *Sensors* 15 (1):2059-2085.

Song, L., and N. N. Eldin. 2012. "Adaptive Real-Time Tracking and Simulation of Heavy Construction Operations for Look-Ahead Scheduling." *Automation in Construction* 27:32-39.

Suzumura, T., and H. Kanezashi. 2014. "Multi-Modal Traffic Simulation Platform on Parallel and Distributed Systems." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 769-780. Piscataway, NJ: IEEE, Inc.

Tannock, J., B. Cao, R. Farr, and M. Byrne. 2007. "Data-Driven Simulation of the Supply-Chain—Insights from the Aerospace Sector." *International Journal of Production Economics* 110 (1):70-84.

Thiemjarus, S., A. Henpraserttae, and S. Marukatat. 2013. "A Study on Instance-Based Learning with Reduced Training Prototypes for Device-Context-Independent Activity Recognition on a Mobile Phone." *2013 IEEE International Conference on Body Sensor Networks (BSN)*.

Vahdatikhaki, F., S. Setayeshgar, and A. Hammad. 2013. "Location-Aware Real-Time Simulation Framework for Earthmoving Projects Using Automated Machine Guidance." In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 3086-3097. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Wohlgemuth, V., B. Page, and W. Kreutzer. 2006. "Combining Discrete Event Simulation and Material Flow Analysis in a Component-Based Approach to Industrial Environmental Protection." *Environmental Modelling & Software* 21 (11):1607-1617.

Zhang, S., C. Du, W. Sa, C. Wang, and G. Wang. 2013. "Bayesian-Based Hybrid Simulation Approach to Project Completion Forecasting for Underground Construction." *Journal of Construction Engineering and Management* 140 (1).

AUTHOR BIOGRAPHIES

REZA AKHAVIAN is an Assistant Professor of Construction Management in the School of Engineering at the California State University, East Bay. He holds M.S. and Ph.D. degrees in Civil Engineering from the University of Central Florida. He also received his B.S. degree in Civil Engineering from University of Tehran (Iran). His current research interests include data-driven simulation, pervasive data collection, machine learning, and process mining and their applications in infrastructure construction and project management. He is a member of INFORMS, ACM-SIGSIM, and ASCE. His email address is reza.akhavian@csueastbay.edu.

AMIR H. BEHZADAN is an Associate Professor of Construction Management at the Missouri State University. He received his Ph.D. in Civil Engineering (Construction Engineering and Management) in 2008 and his Masters degree in Construction Engineering and Management in 2005 both from the University of Michigan, Ann Arbor. He also holds a B.Eng. degree in Civil Engineering from Sharif University of Technology (Tehran, Iran). His current research interests include intelligent and autonomous construction systems, data-driven simulation and visualization, and construction process data mining. He is a member of the American Society of Civil Engineers (ASCE) and serves on the editorial board of the ASCE Journal of Construction Engineering and Management. His email address is abehzadan@missouristate.edu.