



**HAL**  
open science

# A Proposal of Data-Driven and Multi-scale Modeling Approach for Material Flow Simulation

Satoshi Nagahara, Toshiya Kaihara, Nobutada Fujii, Daisuke Kokuryo

► **To cite this version:**

Satoshi Nagahara, Toshiya Kaihara, Nobutada Fujii, Daisuke Kokuryo. A Proposal of Data-Driven and Multi-scale Modeling Approach for Material Flow Simulation. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2022, Gyeongju, South Korea. pp.207-215, 10.1007/978-3-031-16411-8\_26 . hal-04768957

**HAL Id: hal-04768957**

**<https://inria.hal.science/hal-04768957v1>**

Submitted on 6 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# A proposal of data-driven and multi-scale modeling approach for material flow simulation

Satoshi Nagahara<sup>1,2</sup>, Toshiya Kaihara<sup>2[0000-0001-7653-1236]</sup>, Nobutada Fujii<sup>2</sup>  
and Daisuke Kokuryo<sup>2[0000-0001-9365-0704]</sup>

<sup>1</sup> Industry Automation Research Dept., Hitachi, Ltd., Yokohama-shi, Kanagawa, Japan

<sup>2</sup> Graduate School of System Informatics, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe-shi, Hyogo, 657-8501 Japan  
satoshi.nagahara.eb@hitachi.com

**Abstract.** Material flow simulation is a powerful tool to realize efficient operation in complicated production systems such as high-mix and low-volume production. However, it takes great efforts and expertise to construct accurate simulation models. On the other hand, in recent years, IoT and machine learning techniques that collect and utilize field data are advancing rapidly. In this research, we propose a data-driven and multi-scale modeling approach which constructs accurate simulation models semi-automatically. The proposed approach aims to optimize the configuration of simulation model by combining deductive models such as queue model and inductive model such as machine learning model to maximize accuracy. In this article, we introduce the concept of the proposed method and experimental results on a simple production system.

**Keywords:** Material flow simulation, queueing system, machine learning.

## 1 Introduction

In recent years, the diversification of market needs makes production systems more complicated such as high-mix low-volume production. In this situation, Cyber-Physical Production System (CPPS) and Digital Twin are attracting attention as concepts that realize efficient operation in such complicated systems [1]. In these concepts, material flow simulation has an important role to predict the future behavior of production systems [2]. However, it takes a lot of efforts and expertise to make accurate simulation models. This difficulty is one of the barriers to realize CPPS/Digital Twin concept in practice. On the other hand, in recent years, more data can be obtained from production system thanks to the advance of IoT devices. In addition, Machine Learning (ML) techniques are remarkably advancing. These technologies enable us to automatically identify system behavior from data. Additionally, ML techniques are capable to make models with different scales such as production process and whole system. This capability will be useful because it is difficult to completely mimic real system and we need to abstract the system according to the available data. Therefore, in this research, we propose a data-driven and multi-scale modeling approach which semi-automatically constructs simulation models by utilizing the data in production system and ML.

## 2 Challenges in material flow simulation modeling

### 2.1 General modeling process for material flow simulation

A general modeling process for material flow simulation is as follows:

**Step 1:** Modelers define the scope and granularity of simulation model based on their experience and knowledge so that the purpose of simulation use is achieved.

**Step 2:** Many types of data required to express the system behavior are collected and created. Data such as Bill of Materials and process routing often can be obtained from IT systems such as ERP and MES. However, data rarely managed in ERP/MES such as operational control rules need to be defined through analysis and interview.

**Step 3:** The developed model is implemented on simulation software using the data.

**Step 4:** The simulation accuracy is evaluated by the comparison of simulation results with production log. The model is reviewed mostly by trial and error until the required accuracy is archived. The parameters and such as operation time and dispatching rules are mainly adjusted, but the model configuration is also changed if necessary.

### 2.2 Related works and challenge

Most research for automatic modeling regards to data models and system architectures to generate simulation program from data in IT systems [3]. For instance, Kirchhof et al. proposed a method in which simulation objects representing production process are defined in advance, then a simulation program is generated by combining the objects based on MES data [4]. These studies contribute to automate Step 3 in Section 2.1.

In addition, there are some researches about the improvement of simulation accuracy. For example, Karnok et al. proposed a method which estimates process route and operation time from production log [5]. Popovics et al. proposed a method to construct the response model of equipment in a conveyor system from PLC program [6]. In addition, Nagahara et al. proposed a method which identifies dispatching rules from production log by ML techniques [7]. They also proposed a method which adjusts parameters in simulation model to improve accuracy [8]. These studies aim to extract the information required for simulation from the data. They contribute to automate the data collection process and improve simulation accuracy.

Furthermore, ML-based methods which directly express the input-output relationship of the system from production log have been proposed. For example, Lingitz et al. proposed a method that predicts production lead time in semiconductor production systems using ANN (Artificial Neural Network), Random Forest, and so on [9].

The above modeling methods enable us to express production system by various scales of model such as each process and whole system. However, it is not clear what kind of model configuration and modeling method are suitable for a specific system. Thus, the automation and optimization of model configuration is still a challenging issue. Therefore, in this research, we propose a novel modeling approach that optimizes model configuration to improve simulation accuracy. The proposed method aims to automatically generate various model configurations by combining multiple models with different scale and modeling method and derive the optimal one.

### 3 Proposed method

#### 3.1 Classification of modeling methods

Prior to the explanation of the proposed method, we first summarize and classify various modeling methods. Most of production systems consists of multiple production processes, and production process consists of various activities such as setup, machining, resource allocation, and so on. In this research, therefore, we classify modeling methods in three scales: activity scale, process scale and system scale.

In the activity scale, we consider two types of models: “Activity White Box Model (WBM)” and “Activity Black Box Model (BBM)”. Activity WBM is a model in which the activity behavior is explicitly and deductively described based on the background knowledge. On the other hand, Activity BBM is a ML based model which inductively expresses the input-output relationship of the activity from activity log.

In the process scale, we consider “Process WBM” and “Process BBM” as well as the activity level. The former is a kind of queue model in which all activities is expressed by WBM and the connection between activities is also explicitly and deductively described. The latter is a ML based model that predicts the process output from the input. In addition, we can consider “Process Gray Box Model (GBM)” as a kind of queue model in which some or all activities are expressed by BBM.

In the system scale, we consider “System WBM”, “System BBM” and “System GBM”. System WBM is a model in which all process is expressed by WBM. Most of conventional models constructed using commercial software corresponds to System WBM. System BBM is a ML model that predicts system output from system input such as the method by Lingitz et al.[9]. Furthermore, we consider two types of models as System GBM. One is a model in which some or all processes is expressed by GBM or BBM. Another is a model that any sub-system, which is composed with some processes, in the entire system is expressed by System BBM.

#### 3.2 Data-driven and multi-scale modeling for material flow simulation

Based on the classification in Section 3.1, we propose a novel modeling approach called as data-driven and multi-scale simulation modeling. Fig.1 shows a schematic view of the proposed method. The proposed method aims to derive an accurate simulation model by combining various modeling methods. W/G/B in Fig.1 represent WBM/GBM/BBM in arbitrary scale (activity/process/system) respectively. As shown in Model (a), (b) and (c) in Fig.1, we can consider various configurations of simulation model for a certain target system. In this method, multiple model configurations are generated by integrating and dividing the processes and activities in target system (Fig.1(1)). As an example, Model (a), (b) and (c) represent System WBM, GBM and BBM respectively. System WBM is generated from data in ERP/MES using appropriate data interface of simulation software. The data representing background knowledge such as activity information is also utilized for model generation. On the other hand, Activity/Process/System BBM in Model (b) and (c) are generated from production activity logs by ML techniques. Then, according to the input-output requirements for the simulation use, the parameters in each model are calibrated to accurately predict

changes in the output with respect to changes in the input (Fig.1(2)). Let us consider a case that the requirement for simulation is to predict job completion time from job arrival time in the system. In this case, for instance, we can consider a simulation model in which each process is described by Process WBM, GBM or BBM and they predict process completion time of each job in their corresponding process from completion time in the previous processes. Then, the parameters in each model can be calibrated by optimization techniques to minimize the prediction error on job completion time of the overall model. Finally, the most accurate model is selected (Fig.1(3)). This method is a data-driven method that selects the optimal model configuration based on the prediction accuracy to the target system. Furthermore, this method is a multi-scale modeling method that combines models with different scale (activity, process and system level) and modeling method (white/black/gray box model) to express the entire system.

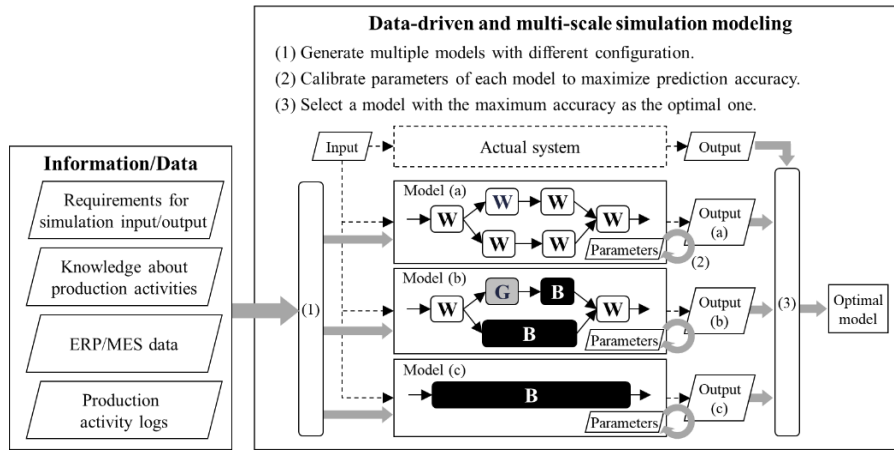


Fig. 1. Schematic view of the proposed modeling approach.

## 4 Computational experiments

### 4.1 Purpose of experiments

The proposed method combines WBM/BBM/GBM to achieve high accuracy. In other words, the proposed method assumes that a certain modeling method is not always the best and the superiority of modeling method depends on the target system. The superiority of modeling method is closely related to the available data and background knowledge. For example, if we know all activities in the target system and the behavior of them, System WBM will achieve high accuracy. If the activities and/or their activity are partially unknown, the accuracy of the System WBM would be low. On the other hand, if the activity log of some activities is available, it is possible to construct Activity BBM of them. From this point of view, we conducted a computational experiment to compare multiple modeling methods and find out the relationship between the availability of data and the superiority of those modeling methods.

## 4.2 Experimental conditions

In this experiment, the virtual production systems A and B in Fig. 2 are considered as target production systems. Both systems A and B are one-equipment and mixed flow systems. The system A has dispatching and machining activities, and the system B has transportation and sequence-dependent setup activities in addition. The dispatching rule is first-in first-out (FIFO), and the transportation time  $l_p$  and machining time  $d_p$  differ for product type  $p$ . The setup time  $s_{p,q}$  differs for the combination of previous product type  $p$  and successive product type  $q$ . The values of  $l_p, d_p, s_{p,q}$  are randomly set, and the number of product types  $M$  is 5 and the number of jobs  $N$  is 50 in each scenario. We created 1,000 scenarios, in which the arrival date-time  $t_i^{in}$  and product type index  $pt_i$  of job  $i$  are randomly set, for model construction and evaluation respectively. The completion date-time  $t_i^{out}$  and lead time  $lt_i$  of job  $i$  are calculated for each scenario from the simulation results of the Process WBM of system A and B.

In this experiment, Process WBMs of system A and B are assumed as actual systems, and the simulation results of the Process WBMs are assumed as actual activity log. In this experiment, it is assumed that the purpose of simulation use is to predict the lead time of each job from the arrival date-time and product type of each job.

Here, since the systems A and B are deterministic systems, it is obvious that Process WBM is the same as the target system if the activities and their behavior are known. In addition, even in the case that the behavior of each activity is unknown, if the activity log, i.e., start and completion date-time of transportation, setup and machining activities, is available, we can obtain accurate Process GBM by constructing Activity BBM for each activity. In this experiment, therefore, we consider the case in which the behavior of some activities is unknown and the activity log of them is unavailable. Specifically, the time information  $l_p, s_{p,q}, d_p$  is unknown, and only the information described in ‘‘Input’’ and ‘‘Output’’ part in Fig.1 is available as the activity log.

In the above problem setting, Process GBM and Process BBM are compared. For Process GBM, we build a queue model and calibrate the time information parameters to minimize the prediction error. Particle Swarm Optimization (PSO) is used to calibrate parameters. As Process GBM, we consider two models, GBM-A and GBM-B. GBM-A and GBM-B consists of the same activities with system A and system B respectively. As Process BBM, we build a ML model which predicts the lead time of each job from the product type and arrival date-time of each job. ANN is used for the ML model.

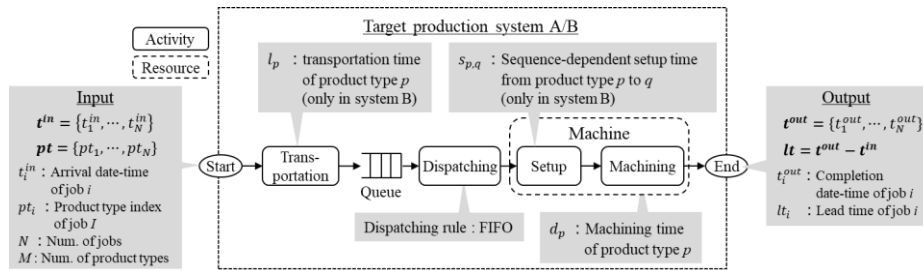


Fig. 2. Experimental conditions: Target production system.

### 4.3 Parameter calibration in Process GBM by PSO

In the parameter calibration in GBM, the parameters that minimize the objective function  $J^{GBM}$  in Eq. (1) are searched.

$$J^{GBM} = \|\mathbf{lt}^{ref} - \mathbf{lt}^{GBM}\| \quad (1)$$

Here,  $\mathbf{lt}^{ref}$  and  $\mathbf{lt}^{GBM}$  are vectors of the lead time in the target system and GBM respectively. Let  $\mathbf{X}$  be a vector whose elements are the values of the parameters  $(l_p, s_{p,q}, d_p)$ .  $\mathbf{lt}^{GBM}$  is the simulation result by GBM with  $\mathbf{X}$ .

In this experiment, PSO with R-best Model proposed by Choi et al. is used for parameter calibration [10]. In general, the vector representing the solution is called as position vector, and its value is updated by Eq. (2) and (3).

$$\mathbf{X}_i^{k+1} = \mathbf{X}_i^k + \mathbf{V}_i^{k+1} \quad (2)$$

$$\begin{aligned} \mathbf{V}_i^{k+1} = w \times \mathbf{V}_i^k + c_1 \times rand() \times (\mathbf{gbest} - \mathbf{X}_i^k) \\ + c_2 \times rand() \times (\mathbf{pbest}_i - \mathbf{X}_i^k) \end{aligned} \quad (3)$$

Here,  $\mathbf{X}_i^k$  and  $\mathbf{V}_i^k$  are the position vector and velocity vector of the individual  $i$  at the  $k$ -th generation, respectively,  $\mathbf{gbest}$  is the position vector of the best solution during the search process, and  $\mathbf{pbest}_i$  is the position vector of the best solution of the individual  $i$  during the search process.  $w, c_1, c_2$  are weight coefficients.

### 4.4 ANN model for Process BBM

For a Process BBM, we use an ANN model with  $N$  inputs and 1 output. The inputs of the ANN are the features related to an arbitrary job, and the output of it is the predicted lead time of that job. The loss function  $J^{BBM}$  is shown in Eq. (4).

$$J^{BBM} = \|\mathbf{lt}^{ref} - \mathbf{lt}^{BBM}\| \quad (4)$$

Here,  $\mathbf{lt}^{BBM}$  is a vector of the predicted lead time by BBM. Since the lead time of job  $i$  depends on not only the product type and arrival date-time of job  $i$  but also the jobs which arrive before and after job  $i$ , the inputs of ANN should be designed to include such information. Therefore, we define the features of job  $i$  by Eq. (5), (6), and (7).

$$x_l = \begin{cases} 1 & \text{if } pt_i = l \\ 0 & \text{otherwise} \end{cases} \quad (l = 1, \dots, M) \quad (5)$$

$$y_{d,l} = \sum_j \delta_{j,d,l} \quad (j = 1, \dots, N, \quad d = -D, \dots, D-1, \quad l = 1, \dots, M) \quad (6)$$

where

$$\begin{aligned} \delta_{j,d,l} \\ = \begin{cases} 1 & \text{if } j \neq i \text{ and } t_i^{in} + d \cdot \Delta D \leq t_j^{in} < t_i^{in} + (d+1) \cdot \Delta D \text{ and } pt_j = l \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (7)$$

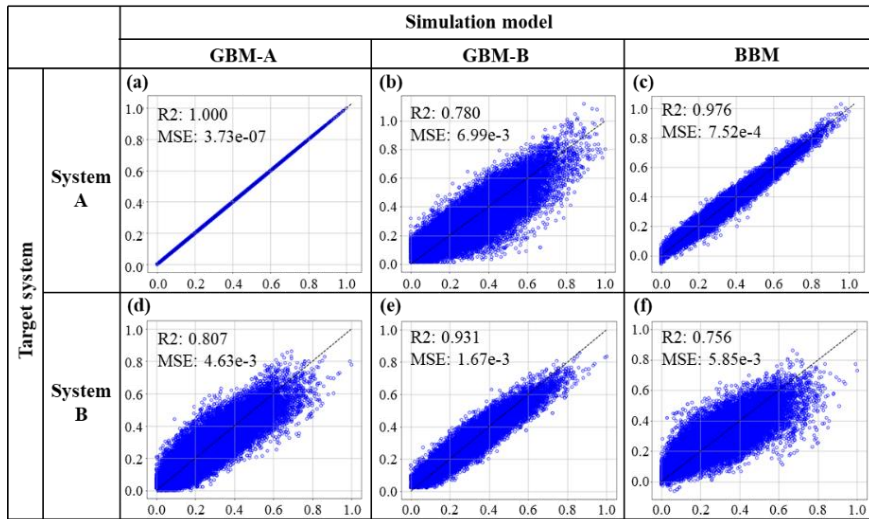
Here,  $M, N$  is the number of product types and jobs respectively.  $y_{d,l}$  is the number of arrival jobs of product type  $l$  in a certain period determined by the arrival date-time of job  $i$  and parameters  $D$  and  $\Delta D$ .  $D$  indicates how long the number of arrival jobs before and after job  $i$  is considered, and  $\Delta D$  is the unit length of the above period.



#### 4.5 Experimental result

The experimental results are shown in Fig.3. The horizontal and vertical axes of each plot in Fig. 3 are the normalized lead time in the target system and GBM/BBM respectively, and each point represents each job. (a), (b) and (c) in Fig. 3 are the results of the target system A, and (d), (e), and (f) are the results of the target system B. (a) and (e) are the results when the activity configuration in GBM is the same as that of the target system, and (b) and (d) are the results when the activity configuration is different.

From the results shown in Fig. 3, the most accurate model for system A and system B is GBM-A and GBM-B respectively. It means that GBM could be superior to BBM if the activity configuration of the target system is known. Additionally, the accuracy of (b) and (d) is lower than that of (a) and (e) respectively. (d) is a result of GBM-A in which some activities are lacked compared to system B. Therefore, there is not the solution of parameters that completely matches GBM-A with system B. (b) is a result of GBM-B in which some activities are excessive compared to system A. Although there is the solution that completely matches GBM-B with system A, the parameter search by PSO fell into a local optimal. This indicates that if a model is too complicated, the parameter calibration may become difficult. BBM shows high accuracy for system A (Fig. 3(c)) and more accurate than GBM-B (Fig. 3(b)). It means that BBM may be superior if the background knowledge about activity configuration of the target system is insufficient. On the other hand, BBM shows less accuracy than GBM-A and GBM-B for system B (Fig. 3(f)). This result indicates that it is difficult for ANN to express a system that includes activities strongly affected by the processing order of jobs such as sequence-dependent setup.



**Fig. 3.** Experimental results: Comparison of prediction accuracy for job lead time. (Horizontal axis: job lead time in target system, Vertical axis: job lead time in simulation model)

## 5 Conclusion

In this research, we proposed a novel modeling approach for material flow simulation called as data-driven and multi-scale simulation modeling. We classified the existing modeling methods from the perspective of activity, process and system scale and White, Black and Gray Box modeling. Then, we introduced a modeling approach that optimizes model configuration by combining those modeling methods. It has been confirmed that the proposed method is expected to contribute to automate the design phase of model configuration and execute simulation more accurate. In addition, we verified the effects of differences in available data on the superiority of modeling methods through computational experiments.

As future works, the comparison of WBM/GBM/BBM for more complicated system will be conducted. Additionally, we should consider the parameter calibration method for GBM and the design of feature variables and model structure for BBM suitable to express production system. Furthermore, we'll develop methods which automatically generate multiple configurations of simulation model to realize the proposed concept.

## References

1. Monostori, L.: Cyber-physical production systems: Roots, expectations and R&D challenges. *Procedia CIRP*, Vol.17, pp. 9-13 (2014).
2. Rosen, R., Wichert, G., Lo, G., Betternhausen, K. D.: About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine*, Vol. 48, Issue 3, pp.567-572 (2015).
3. Kirchhof, P.: Automatically generation flow shop simulation models from SAP data. *Proceedings of the 2016 Winter Simulation Conference*, pp.3588-3589 (2016).
4. Barlas, P. and Heavey C.: Automation of input data to discrete event simulation for manufacturing: A review, *International Journal of Modeling. Simulation and Scientific Computing*, Vol. 7, No. , 1630001 (2016).
5. Karnok, D., Monostori, L.: Determination of routings and process time information from event logs. *Preprints of the 18th IFAC World Congress*, pp.14055-14060 (2011).
6. Popovics, G., Monostori, L.: ISA standard simulation model generation supported by data stored in low level controllers. *Procedia CIRP* 12, pp.432-437 (2013).
7. Nagahara, S., Sprock, T. A., Helu, M. M.: Toward data-driven production simulation modeling: dispatching rule identification by machine learning techniques. *Procedia CIRP* 81, pp. 222-227 (2019).
8. Nagahara, S., Serita, S., Shiho, Y., Zheng, S., Wang, H., Chida, T., Gupta, C.: Toward data-driven modeling of material flow simulation: automatic parameter calibration of multiple agents from sparse production log. *16th IEEE International Conference on Automation Science and Engineering*, pp.1096-1101 (2020).
9. Lingitz, L., Gallina, V., Ansari, F., Gyulai, D., Pfeiffer A., Sihn, W., Monostori, L.: Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. *Procedia CIRP* 72, pp.1051-1056 (2018).
10. Choi, H., Ohmori, S., Yoshimoto, K.: Improvement of particle swarm optimization: Proposal of R-best model and parameter adjustment with consideration to searching phase and state. *21st International Conference on Production Research*, pp. 2079-2083 (2011).