Simulation in Produktion
und Logistik 2017
Sigrid Wenzel & Tim Peter (Hrsg.)
kassel university press, Kassel 2017

Functional Interaction of Simulation and Data Analytics – Potentials and Existing Use-Cases

Kopplungsmöglichkeiten von Simulation und Data Analytics - Potenziale, Lösungsansätze und Anwendungsfälle

Christoph Laroque, Westsächsische Hochschule Zwickau, Zwickau (Germany), christoph.laroque@fh-zwickau.de

Anders Skoogh, Maheshwaran Gopalakrishnan,
Chalmers University of Technology, Gothenburg (Sweden),
anders.skoogh@chalmers.se, mahgop@chalmers.se

Abstract: Discrete event simulation (DES) is very suitable to model the reality in a manufacturing system with high fidelity. Such models are easy to parameterize and they are able to consider several influences including stochastic behavior. However, simulation models are challenged, when it comes to operational decision support in manufacturing as well as logistics. Here, methods and algorithms from the area of (big) data analytics see growing importance in research as well as practical application. Much less likely is the winning combination of both in a common approach. This paper aims to describe these opportunities, both in general following the steps of a simulation study as well as on the basis of current research publications that have been researched by the authors.

1 Introduction

"*Big data analytics is where advanced analytic techniques operate on big data sets. Hence, big data analytics is really about two things—big data and analytics—plus how the two have teamed up to create one of the most profound trends in business intelligence (BI) today*" (Russom 2011).

Over the last decades, a large number of automatic data analysis methods as well as visual analytics methods have been developed and their application areas are still growing (Bange et al. 2015). However, the complex nature of many problems makes it necessary to include human intelligence at an early stage in the data analysis process. Big Data Analytics methods allow decision makers to combine their human flexibility, creativity, and background knowledge with the enormous storage and processing capacities of today's computers to gain insights into complex problems. Using advanced visual interfaces, humans may also directly interact with the data analysis capabilities of today's computer, allowing them to make well-informed decisions in complex situations (Thomas and Cook 2005).

Material flow simulation of production and logistics processes is already a well-established field of application for simulation technologies. Applications of simulation models as well as the generated data sets have grown enormously over the past years; since simulation models also get more and more detailed and by that complex, the amount of data to be analyzed is also growing. The analysis of the generated simulation results out of a large set of experiments today mostly builds on expert knowledge and human interaction during the analysis phase of a simulation study. Moreover, multiple techniques for the validation and verification of the corresponding simulation data as well as the simulation models is today mainly a manual process by the simulation expert. Big Data Analytics has the potential to enhance these processes for faster analysis during multiple steps of a simulation study and, in some cases, can even be used for a first result forecasting based on simplified mathematical models. This also may lead to faster decision making processes.

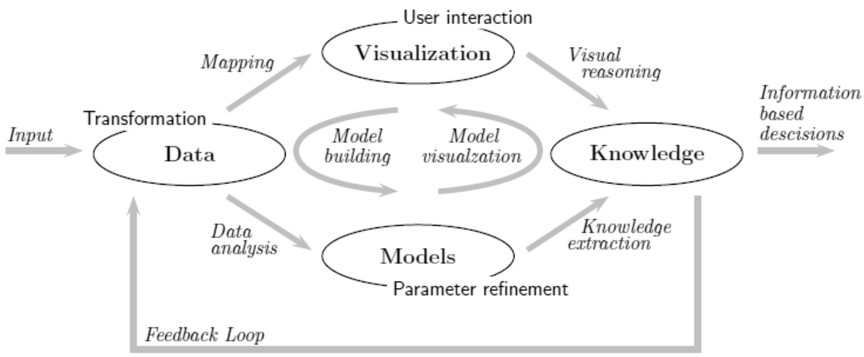

Figure 1: *Schematic "Data-to-Knowledge" process*

However, this combination of simulation and (big) data analytics methods has not yet been carried out systematically, though there are some existing use-cases in industrial practice. The capability of combined approaches coupling simulation and advanced analytical approaches is addressed only for a few specific applications. Consequently, no general framework exists to define this symbiosis and describe potential benefits.

In this paper, we propose a generalized concept to facilitate the utilization of (big) data analytics technologies in simulation studies. For each step in a sound simulation study, possible methodologies and connections are carried out and, later, existing use-cases from the field of production and logistics simulation are derived. Based on these findings and requirements, a first solution for a combined framework for simulation studies, supported by big data analytics is described.

The paper will therefore start with a brief motivation and the clarification of both methodological approaches. Then, the combinatorial integration of these methods will be carried out by following the steps of a simulation study. In a further section, existing use-cases will be classified in the framework. A summary will conclude the paper.

2 Relevant Terms and Definitions

Modern business computing, especially in the area of operations research, offers a wide variety of methods for solving complex problems planning, scheduling and control of production and logistic processes. Typically, new processes have to be designed or existing once have to be improved. For this, the processes are projected into models and then optimized by the use of simulation and/or optimization technologies in order to improve decision variables and resulting key performance indicators under a given set of restrictions. In simulation, this improvement is usually achieved by the iterative evaluation of multiple scenarios and their subsequent simulation results (Law and Kelton 2000). In the case of optimization, the optimal configuration is achieved by mathematical optimization algorithms or (meta-) heuristic approaches (Rardin 1997). Due to the high computational demand of both iterative evaluation and mathematical optimization, specific procedures as a combination of both simulation and optimization were derived. These procedures combine the advantages of both methods: an optimization algorithm can be used to automatically generate a specific model configuration, which then can be evaluated with simulation runs (Fu 2002). Those and other methods are typically described as Operations Research (OR).

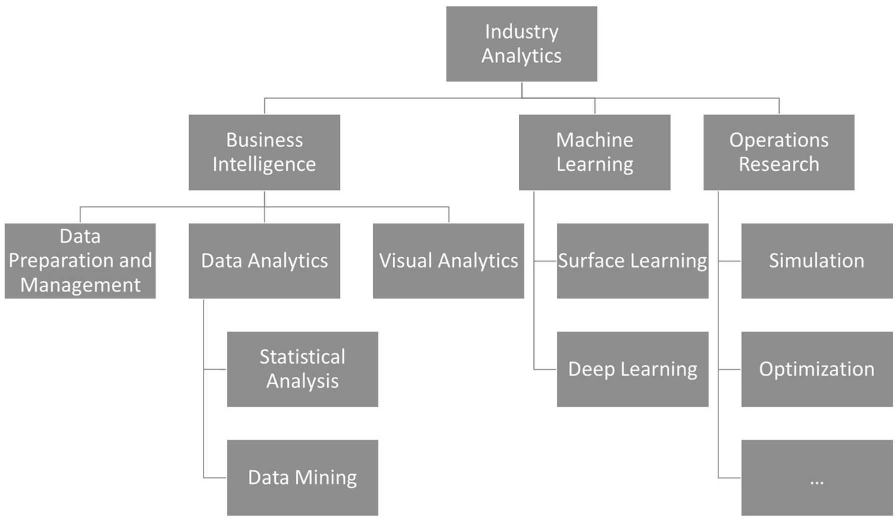

Figure 2: *Conceptual structure of Keywords and Methods in Business and Industry Analytics*

In the data-driven analysis of existing manufacturing or logistical systems, further mathematical methods are applied regularly in order to derive information that may support decision making processes, out of the data, that is already available in specific information systems, e.g. Enterprise Ressource Planning (ERP), Customer Relationship Management (CRM), Manufacturing Execution Systems (MES) or Product Lifecycle Management (PLM). They are summarized as data analytics (for some insights, see Morabito 2015 and Runkler 2016). Together with some more fundamental methods on data management and handling and interactive visualizations, this topic is covered as Business Intelligence (BI, e.g. Kemper et al 2010). The focus of these approaches is to deliver insights on what and why some effects happened. They are widely characterized as ‘descriptive’ and ‘diagnostic’ analytics. Based on these approaches, domain experts today work more and more on solutions that allow some significant forecasting based on this knowledge, either by some rather simplified mathematical models (predictive analytics, which answers what might happen) or by more sophisticated approaches that are summarized as prescriptive analytics. Here, deeper understanding of the systems behaviour is needed and more complex modelling and analysis from the OR-area are considered.

Besides that, a number of approaches can be found during the last years, where algorithms from the area of machine learning (Michalski et al. 1983) and artificial intelligence (ibid.) are used to automate these analytical steps, e.g. the pattern recognition in data or to derive new patterns and correlations in the growing amount of data, that is available in these systems. Figure 2 tries to give an overview about the links between these terms.

3 Integration from a Simulation Perspective

One of the main ideas of this contribution is to stress the potential benefit that methods from the area of data analytics may deliver in order to support the execution of a simulation study. Therefore, the typical steps of a sound simulation study according to (Law and Kelton 2000) are described shortly, followed by potential methods that could be carried out in order to support the single process steps.

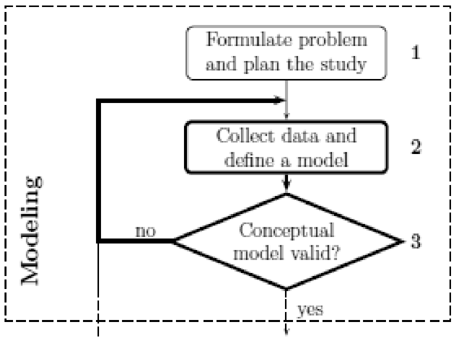

Figure 3: *application of descriptive and diagnostic analytics for preparation and validation of the data for a simulation study*

The usual and meaningful kick-off for a simulation study is the formulation of the goal of the research and a validation, that simulation is the appropriate method. Of course, in some cases, one might already see here some possibilities to gain some necessary insights with some simpler mathematical approaches. Since simulation studies aim at more complex systems, this consideration is skipped here, although there are lot of existing examples, where the benefits of a simulation study could be derived much easier with some fundamental data analytics or queuing theory applications. The next valuable step is the conceptual modeling of the system and, in parallel, the gathering, validation and verification of the necessary data. If some empirical data already exists, then visual analytics and a couple of the statistical methods are typically carried out, that are also used within descriptive analytics, e.g. identification of the distribution, min-max-mean-analysis, etc. For validation, outlier detection and pattern recognition can be used manually or automatically as data mining applications. Moreover, out of the amount of available data, some first correlation analysis can be derived to gain insights on significant dependencies in the data. Available data sources can be filtered via cause-and-effect-analysis for those parameters that probably will be significant on the systems performance and should be considered therefore during the later experimentation phase.

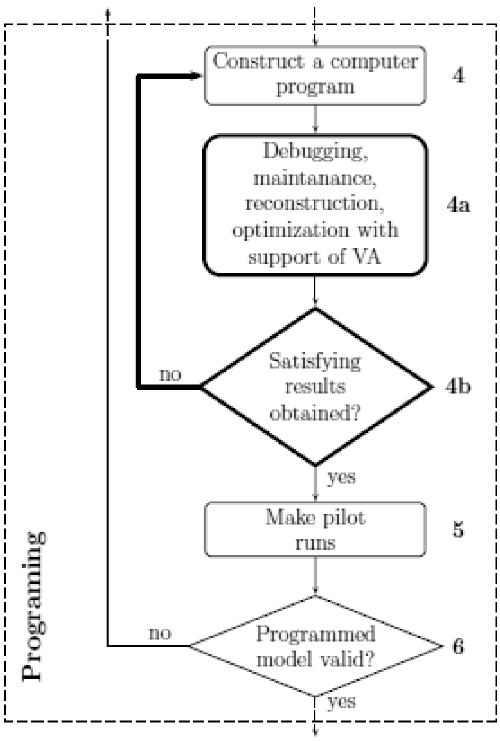

Figure 4: *Validaton of the formal simulation model by descriptive and visual analytics*

After the successful completion of these steps, the formal representation of the conceptual model is applied in a simulation software. Here, more visual approaches are today regularly applied for the validation and verification of the simulation model and some first pilot tests are run. Here some first data as a result of the simulation model derived, which again can be analysed in order to validate the gathered results. The presumably identified correlations can be checked (again) and some further insights on the models behaviour is derived. If, for example, known correlations cannot be validated via the simulation model, this also might indicate some programming mistakes during the adoption of the conceptual to the formal, computation model within the simulation software. The major advantage of some systematic application of the data analytics approaches, though, can be derived during the following phase of the simulation experimentation.

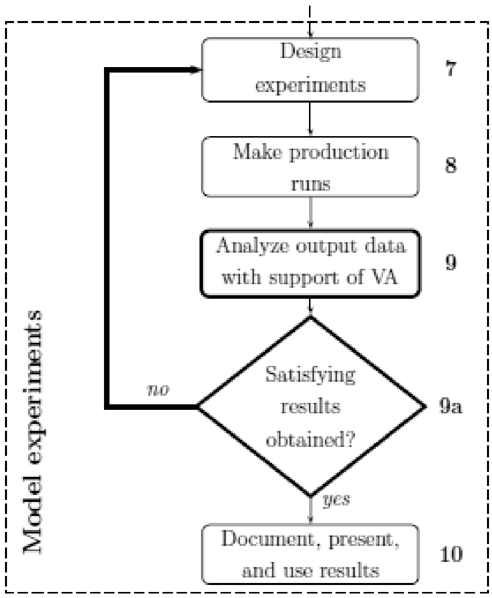

Figure 5: *Applying Data Analytics in the output analysis of a simulation study*

During the execution of the simulation scenarios, typically with multiple replications due to the system's stochastic behaviour, huge amounts of data are being created, that are to be analysed systematically in order to gain some useful insights according to the overall goal of the simulation study. Since during the last years, approaches like simulation optimization, simheuristics as well as distributed computation of the simulation experiments in cloud-based environments are of growing importance, the amount of data to be analysed is also growing rapidly. Here, especially automatic techniques from data analytics come into play. Data Mining as well as Machine Learning algorithms are both able to cluster the gained results and to identify patterns that cause changes in the performance indicators. This information might be presented to the simulation expert in order to enhance the process of system understanding. Moreover, powerful visualization tools are able to interactively visualize those huge amounts of data and thereby allow a manual interpretation of the performance metrics of the simulation model (visual analytics).

4 Use Cases

The authors did considerable research during the creation of this contribution in order to identify existing use-cases. Some of the use-cases are described as examples in this section. (However, there's of course more out there and further work existing. The authors are willing to share their entire research results on-demand.) Most of the latest research is carried out within the construction of modern manufacturing systems that are highly integrated with their corresponding IT-system, so called cyber-physical systems. Industry 4.0 as the application of the Internet-of-Things (IoT) in this area play a major role in these modern applications for a decision support in the more and more dynamic, flexible and complex manufacturing processes.

4.1 Simulation and Data Analytics in Production and Manufacturing

In 'Determining the optimal level of autonomy in cyber-physical production systems' (Gronau et al. 2017) the authors state that traditional production systems are enhanced by cyber-physical systems (CPSs) and Internet of Things. As kind of next generation systems, those cyber-physical production systems (CPPSs) are able to raise the level of autonomy of its production components. To find the optimal degree of autonomy in a given context, their approach is formulated using a simulation concept. Based on requirements and assumptions, a cyber-physical market is modeled and qualitative hypotheses are formulated, which will be verified with the help of the CPPS of a hybrid simulation environment. Data Analytics are used to extract influence factors which explain the optimal degree of autonomy.

(Brodsky et al. 2015) propose an architectural design and software framework for fast development of descriptive, diagnostic, predictive, and prescriptive analytics solutions for dynamic production processes. The proposed architecture and framework shall support the storage of modular, extensible, and reusable Knowledge Base (KB) of process performance models. The approach requires a variety of underlying analysis tools, including data manipulation, optimization, statistical learning, estimation, and simulation.

(Jackson et al. 2016) focus on the aspect of reconfiguration of an aerospace production system. Here, advanced manufacturing technologies broadly used in automotive industry have limited application for typical UK aerospace manufacturing, as they require production volume and repetition of operations to deliver value. This paper discusses a framework of key technologies ranging from digital manufacturing concepts to a flexible fixture that enables reconfiguration in aerospace manufacturing systems. Initially, the overall architecture of the framework is presented illustrating the key components such as a cloud based data storage mechanism, an intelligent multi-product assembly station, kitting boxes embedded with sensors, a manufacturing network management portal and a decision support tool that combines data analytics and discrete event simulation. Afterwards, the main functionalities and technologies of the components are described and finally an industrial application scenario for the proposed framework is presented.

In 'Integrating data analytics and simulation methods to support manufacturing decision making' (Kibira et al. 2016) the authors denote, that in CPS, the manufacturing system itself is installed with smart devices such as sensors that monitor system performance and collect data to manage uncertainties in their operations. However, multiple parameters and variables affect system performance, making it impossible for a human to make informed decisions without systematic methodologies and tools. Further, the large volume and variety of streaming data collected is beyond simulation analysis alone. Simulation models are run with well-prepared data. The authors agree, that novel approaches, combining different methods, are needed to use this data for making guided decisions. Their contribution proposes a methodology whereby parameters that most affect system performance are extracted from the data using data analytics methods. These parameters are used to develop scenarios for simulation inputs; system optimizations are performed on simulation data outputs. A case study of a machine shop demonstrates the proposed methodology. This paper also reviews candidate standards for data collection, simulation, and systems interfaces.

Many industries are applying various methods for optimizing energy use across the manufacturing life cycle. These methods are either physics-based or data-driven. Manufacturing systems generate a vast amount of data from operations and in simulations. Advances in data collection systems and data analytics (DA) tools have enabled the development of predictive analytics for energy prediction. Many of these prediction methods do not account for the uncertainty quantification-UQ (both in data and model). The work of Ronay and Bhinge (2015) addresses the issue of uncertainty in predictive analytics. It focuses on metal cutting processes and presents a Neural Networks (NNs) model to predict the required energy consumption during the manufacturing of a part on a milling machine. Moreover, it is shown that with advanced data collection and processing techniques, one can construct a model to predict the energy consumption of a machine tool for machining a part with multiple operations and process parameters.

4.2 Simulation and Data Analytics in Logistics

In ‘Application of big data technology in support of food manufacturers' commodity demand forecasting’ Nita (2015) states, that forecasting commodity demand for food manufacturers is very difficult to achieve because it is easily affected by variable factors such as the weather and the success or otherwise of advertising campaigns. The ambiguity of demand forecasting results in burdensome supply and demand adjustments and causes an increase in the logistics and production costs, stock-out, excessive stock and/or disposal losses. NEC applies one of its big data technologies, the heterogeneous mixture learning technology, to perform commodity demand forecasting and to automate and optimize the supply chain management systematization. In addition, NEC also applies machine learning to the simulation of sales measures in order to maximize their effects and to increase the sales volume.

Xu et al. (2015) expand their simulation optimization approach with big data analytics. In the first part of their contribution, they classify simulation optimization techniques afterwards, they review applications of simulation optimization in various contexts, with a more detailed discussions on logistics and manufacturing systems. Specifically, the authors then discuss how simulation optimization can benefit from various IT-technologies like cloud computing and high-performance computing and the integration of simulation optimization with big data analytics.

The focus of the contribution of Schuh et al. (2014) is the Short-term cyber-physical production management. However, a major trigger for the planning and safeguarding of the production facility, a high adherence to delivery dates is a logistic target. In consideration of the fact that material and information flows in production plants are getting more intersected and networked than ever before and customer demand tailored products in short throughput times, keeping an overview as well as responding properly becomes a huge challenge for the production manager. In their paper, the authors describe a new approach of cyber-physical short-term assistance of the production manager with the goal to support the production controller by providing prioritized short-term actions through a combination of new sensor technologies, big data processing and simulation. The paper outlines the roadmap to short-term cyber-physical production management. With the help of visualization analytics, the application displays the effects of a performed action.

5 Summary and Outlook

As the general approach, as well as the given use-cases, show that in order to deploy a decision-support system for modern manufacturing and logistics systems, it is meaningful to integrate the classical material flow simulation approach with modern Data Analytics approaches and algorithms. Especially for applications, that aim at some model insight close to real-time, intended for the use in operational production planning and/ or (re-)scheduling,

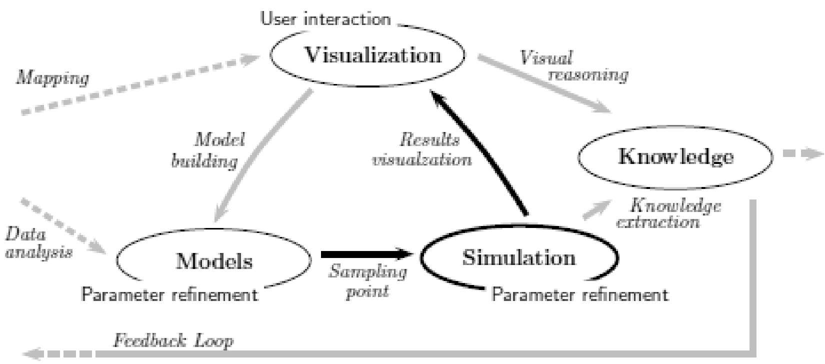

Figure 6: *Simulation as an enhancement for the 'Data-to-Knowledge' process*

From the data analytics perspective, simulation can be regarded as an additional source of knowledge and as an important multiplier of those systems, where real-world data are not (yet) existing or specific scenarios are to be 'tested' in a most realistic way. For the simulation expert, these advantages of our method is basic knowledge; for some data scientists, this might be new. Nevertheless, Data Analytics as a method pool in production and logistics is of growing importance and lot of current research deals especially with predictive approaches that should carry out some insights on the systems behavior. Their next step, prescriptive analytics, will strengthen the application of simulation-based approaches in decision-support systems on an operational level at the factory floor. Simulation tools as well as the applying experts better be ready then to handle complex models with huge amounts of data in more or less automated approaches with massive algorithmic support.

Acknowledgement

Some of the work presented in this contribution, especially in the interplay of simulation and visualization was carried out with the support of Betram Ledwa, student at the Technical University of Dresden in July 2014.

References

Bange, C.; Grosser, T.; Janoschek, N.: Big Data Use Cases – Getting real on data monetarization. BARC Research Study, 2015.

Brodsky, A.; Shao, G.; Krishnamoorthy, M.; Narayanan, A.; Menasce, D.; Ak, R.: Analysis and optimization in smart manufacturing based on a reusable knowledge base for process performance models. IEEE International Conference on Big Data, DOI:10.1109/BigData.2015.7363902, 2015.

Fu, M. C.: Optimization for simulation: Theory vs. practice. INFORMS Journal on Computing, 2002.

Gronau N.; Grum, M.; Bender, B.: Determining the optimal level of autonomy in cyber-physical production systems. IEEE International Conference on Industrial Informatics (INDIN), DOI: 10.1109/INDIN.2016.7819367, 2017.

Jackson, K.; Efthymiou, K.; Borton, J.: Digital Manufacturing and Flexible Assembly Technologies for Reconfigurable Aerospace Production Systems. DOI: 10.1016/j.procir.2016.07.054, 2016.

Kemper, H.-G.; Baars, H.; Mehanna, W.: Business Intelligence - Grundlagen und praktische Anwendungen – Eine Einführung in die IT-basierte Managementunterstützung, DOI: 10.1007/978-3-8348-9727-5 Springer Vieweg 2010.

Kibira, D.; Hatim, Q.; Kumara, S.; Shao, G.: Integrating data analytics and simulation methods to support manufacturing decision making. Proceedings of the Winter Simulation Conference 2016, DOI: 10.1109/WSC.2015.7408324, 2016.

Law, A.; Kelton, D.: Simulation Modeling and Analysis, McGraw Hill 2000.

Michalski, R.-S.; Carbonell J.G.; Mitchell, T.M.: Machine Learning – An Artificial Intelligence Approach. DOI: 10.1007/978-3-662-12405-5, Springer, 1983.

Morabito V.: Big Data and Analytics - Strategic and Organizational Impacts, DOI: 10.1007/978-3-319-10665-6, Springer 2015.

Nita, S.: Application of big data technology in support of food manufacturers' commodity demand forecasting. NEC Technical Journal, 2015.

Rardin, R. L.: Optimization in Operations Research, Prentice Hall, 1997.

Ronay, A.K.; Bhinge, R.: Data analytics and uncertainty quantification for energy prediction in manufacturing. Proceedings of the 2015 IEEE International Conference on Big Data, DOI: 10.1109/BigData.2015.7364081, 2015.

Runkler, T.-A.: Data Analytics - Models and Algorithms for Intelligent Data Analysis, DOI: 10.1007/978-3-658-14075-5, Springer 2016.

Russom, P.: Big Data Analytics. TDWI BEST PRACTICES REPORT, 2011.

Schuh, G.; Potente, T.; Thomas, C.; Hempel, T.: Short-term cyber-physical production management. DOI: 10.1016/j.procir.2014.10.024, 2014.

Thomas, J.; Cook, K.: Illuminating the Path: Research and Development Agenda for Visual Analytics. IEEE-Press 2005.

Xu, J.; Huang, E.; Chen, C.-H.; Lee, L.H.: Simulation optimization: A review and exploration in the new era of cloud computing and big data. Journal of Operational Research, DOI: 10.1142/S0217595915500190, 2015.