Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

AN APPROACH FOR INCREASING THE LEVEL OF ACCURACY IN SUPPLY CHAIN SIMULATION BY USING PATTERNS ON INPUT DATA

Markus Rabe
Anne Antonia Scheidler

Technical University Dortmund
Leonhard-Euler-Straße 5
44227 Dortmund, GERMANY

ABSTRACT

Setting up simulation scenarios in the field of Supply Chains (SCs) is a big challenge because complex input data must be specified and careful input data management as well as precise model design are necessary. SC simulation needs a large amount of input data – especially in times of big data, in which the data is often approximated by statistical distributions from real world observations. This paper deals with the question how the model itself and its input can be effectively complemented. This takes into account the commonly known fact, that the accuracy of a model output depends on the model input. Therefore an approach for using techniques of Knowledge Discovery in Databases is introduced to derive logical relations from the data. We discuss how Knowledge Discovery would be applied, as a preprocessing step for simulation scenario setups, in order to provide benefits for the level of accuracy in simulation models.

1 INTRODUCTION

When simulating complex real world Supply Chains (SCs), the modeling of variability is one of the most important steps. Many aspects of SCs, e.g. transactions, times for customer orders or order processing times are variable.

Table 1 Classification of data (Robinson 2004).

Category A	Available
Category B	Not available but collectable
Category C	Not available and not collectable

As shown in Table 1 the underlying data representing the variability can be divided into three categories based on availability and collectability (Robinson 2004). Robinson stated that Category A data are available e.g., because they have been collected, Category B data need to be collected e.g., arrival times, and Category C can only be estimated or treated as experimental parameter. Variability is often approximated by probability distributions like Erlang and Gamma (Banks 1998), depending on the object behavior being modeled. Despite the possibility to select the correct distributions by analyzing historical data and fitting a suitable distribution, it is also possible to analyze the current processes of the system to be modeled. Standard input modeling approaches in simulation consist of four steps: Testing input data assumptions, selecting possible distributions, estimating the distribution's parameters and analyzing the model adequacy (Merkuryeva and Vecherinska 2010). If we take a look at the standard procedure – fitting a distribution to existing raw data – we must carefully consider if we are integrating the existing information of the raw data in an adequate way. Akhavian and Behzaadan (2013) point out that unrealistic input data is one of the most important reasons why simulation frameworks fail to provide high quality output for complex tasks. Furthermore, they identify that from the simulation perspective most computer models fail to provide reliable output if the model input is not an accurate representation of the real world. The sufficient representations of the real world is comprised of two aspects, the data and the model, where normally the rules for execution logic are defined for the conceptual model. To demonstrate an approach for a higher level of accuracy taking additional logical relations into consideration, we first examine a specific simulation approach and its input data management.

2 DISCRETE EVENT SIMULATION, INPUT DATA AND MODEL SPECIES

The combination of Discrete Event Simulation (DES) and SCs is a well-established field of research. Particular difficulties in this field exist with regard to the data base of DES projects. The difficulties are characterized by different terminologies, co-existing models and specified adaptions. Therefore, a short overview is given about the state of the art of DES and SCs and the terms “input”, “data” and “accuracy” are introduced.

2.1 Supply Chains and Discrete Event Simulation

There are many approaches for simulating Supply Chains (Tarokh and Golkar 2006). The advantages and disadvantages have been extensively studied by different researchers (Terzi and Cavalieri 2004, Lee at al. 2002, Tako and Robinson 2012). Discrete Event Simulation is used widely within production and logistics. Various authors show the suitability of DES to simulate SCs, e.g. Hellström and Johnsson (2002) or Preusser et al. (2005). A literature overview for DES and SC can be found in Rabe and Deininger (2012). A specific characteristic of SC data is the amount of transactions, where a transaction is defined as exchange of material or immaterial objects between the actors. Transactions in the context of Supply Chains can have various distinctive attributes as for instance time stamps or details on quantities. In DES the success of a simulation project relies heavily on the input data quality (Skoogh and Johansson 2009). But, in industry the existing raw data quality often doesn’t meet the requirements for simulation input (März et al 2010). If raw data are available, the simulation leads to problems like outlier problems, noise and missing data. Problem definition and raw data gathering are closely linked because the selection of data differs in relation to the problem definitions and associated questions. The definition of problems, data gathering and data analysis are a mostly uncovered area in simulation (Lehtonen and Seppala 1997). Before introducing the approach for increasing the accuracy of input data by adding logical relations in DES projects, clarification of input data is needed. This denotes in particular the exact meaning of input data in the related context, the understanding of the term “level of accuracy” and the study of standard input modeling of DES focusing on the conceptual model. Due to the necessity of offering a greater practical advantage, standard procedures are adapted. For this reason, the following discussions are based on selected models and procedures.

2.2 Level of Accuracy

In this paper the terms “level of detail” and “level of accuracy” are used referring to Acken (1997). We define the level of accuracy for input data and model as: “Describes how well an entity reflects reality, completes a task or solves a problem by including or excluding relevant elements”. The amount of input data strongly depends on the level of detail chosen for a simulation setup. The level of detail is related to the level of accuracy but explicitly under the condition that accuracy and detail are not the same (Perera and Liyanga 1999). It is important to determine a scale for accuracy because “often, the goal of simulation input modeling is to provide a model that is reasonable, given the goals of the simulation; in contrast to standard statistical applications, often we are not really interested in determining whether the model is perfect” (Banks 1998).

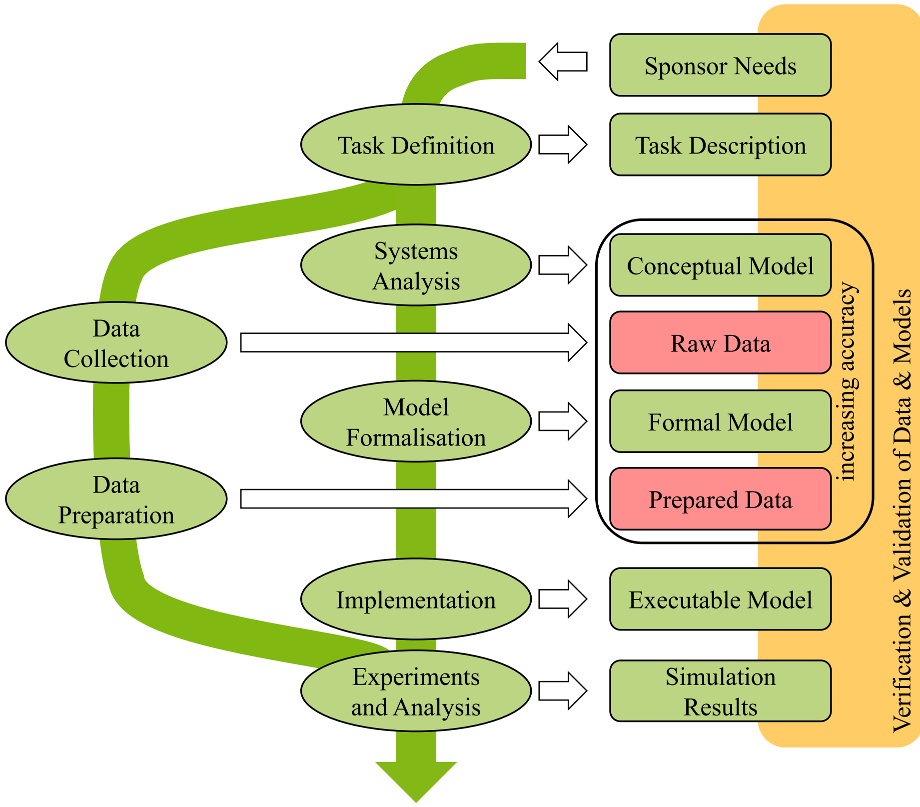

Figure 1: Procedure Model of Rabe, Spieckermann and Wenzel (2008).

Simulation can be divided into data input, model, simulation experiments and output. Level of detail and level of accuracy can be both considered for each step of the process with the aim to fulfill particular simulation requirements. For the research purposes only, the different input data stages and the model phase are considered (Figure 1). The simulation procedure model (Rabe, Spieckermann and Wenzel 2008) starts with the target description. On the right hand side the results of the corresponding task are represented. In the middle it displays the individual simulation tasks in chronological order. On the left hand side the divested data tasks are represented. The motivation of splitting is the independent handling of the data tasks (in terms of time, content and relevant persons) in relation to simulation modeling. In the present research paper only the level of accuracy for a high level of input data and model logic is considered. Both terms "input data" and "model logic" are discussed next.

2.3 Input Data and Input Data Modeling for DES

The process of information acquisition is divided into different steps. The approach of this paper is oriented towards the model of information acquisition of Bernhard and Wenzel (2005). The model starts by identifying information. During the following step the collection of this information is prepared by selecting information sources and acquisition methods. The next step is collecting information and also recording the data. These results are stored within an information system (e.g. database, data warehouse). If we take a look at the process of input data management for DES presented by Skoogh and Johansson (2008), we can map the storing of data in an information system to the stage before the preparation of statistical or empirical representation is conducted. The research approach introduced in this paper utilizes this state of data as its input. There is no concrete distinction related to the various kinds of data input.

The main idea of simulation input modeling is to create a valid input model that reproduces the input process of the real world system. Simulation input modeling for unpredictable variability can be divided into different categories, e.g. bootstrapping or distribution (Law 2007). Subdivision can be made between raw data, preprocessed data for standard simulation input modeling or another kind of processed data as a result preprocessing adaption. The improvement of the process of data collection is not methodically related to data processing which is considered in this paper and therefore has to be regarded independently. We stated that our approach is related to the period after data collection, but more experiments are needed to make a decision on which of the different data types should serve as input data. As a consequence, the term “level of accuracy” only refers to the stages after the collection of data. Hence, the method being described in this paper has the objective to extend an already existing solution. The starting point of the method is the information system’s current state. In general the system’s current state analysis along with the task definition serves as the foundation for the documentation of the conceptual model. Hereinafter, we analyze the impact of higher accuracy on the conceptual model.

2.4 Developing Model Species – The Conceptual Model

We have clarified the terms data and level of accuracy. In the next step, an explanation is needed which model species should be adapted for gaining higher accuracy in DES. In the current research phase no concrete distinction is made between the different model species (Figure 1). As the conceptual model is the foundation of the execution logic, it is consequentially the first model to adapt. We stated that additional knowledge is related to the accuracy. The accuracy itself has two connection objectives, namely the input data and simulation modeling. The conceptual model which contains the logical relations for simulation execution needs to be affected because execution rules may change. This is an adaption of the system’s logical behavior and needs to be reflected in the conceptual modeling phase. The reason given for this statement is that a conceptual model should always reflect the knowledge about the system (Wand et al. 1995). If other models are affected in the same extend further clarification is necessary, because model theory of relationship between conceptual and computation models must be taken into account (Juristo and Moreno 2000). We have identified the connection of input and model for increasing accuracy. One way to increase accuracy is to take more knowledge into account. But what can be the enhanced knowledge and how can we identify it?

3 KNOWLEDGE DISCOVERY IN DATABASES: WHAT IS A PATTERN?

In the past, some techniques of Knowledge Discovery in Databases (KDD) were successfully applied to the field of logistics (Rahman, Desa and Wibowo 2011). Recent projects combine KDD and simulation in a successful manner by focusing on certain data aspects (Bogon et al. 2012). Therefore, knowledge discovery in SCs by using KDD techniques is much encouraged. When talking about finding knowledge in data, we refer to KDD, meaning the overall process of discovering useful knowledge from data as it is defined by Fayyad, Piatetsky-Shapiro and Smyth (1996).

The process starts with the selection of data, data preprocessing, for instance cleansing, followed by the transformation of data. Transformation is necessary because data mining algorithms require specialized input, e.g. a target attribute needs to be labeled or nominal values of attributes need to be mapped to numeric values. The process uses data mining algorithms as an essential step and ends with the interpretation of the results (Figure 2).

The step in the KDD process called data mining is the extraction of implicit, previously unknown and potentially useful information from data (Witten, Eibe and Hall 2011). This information, extracted from data, can be represented in various forms and can be distinguished according to the limit of their validity. Information derived from data can be any kind of structure, logical relation between attributes or in general a description. This description, extracted from the data, is called pattern (Hand 2002). Patterns can be specified by many different categories in relation to the focus of the KDD run. As an example, a pattern can be distinguished between local and global coverage or the different ways in which a pattern can be expressed.

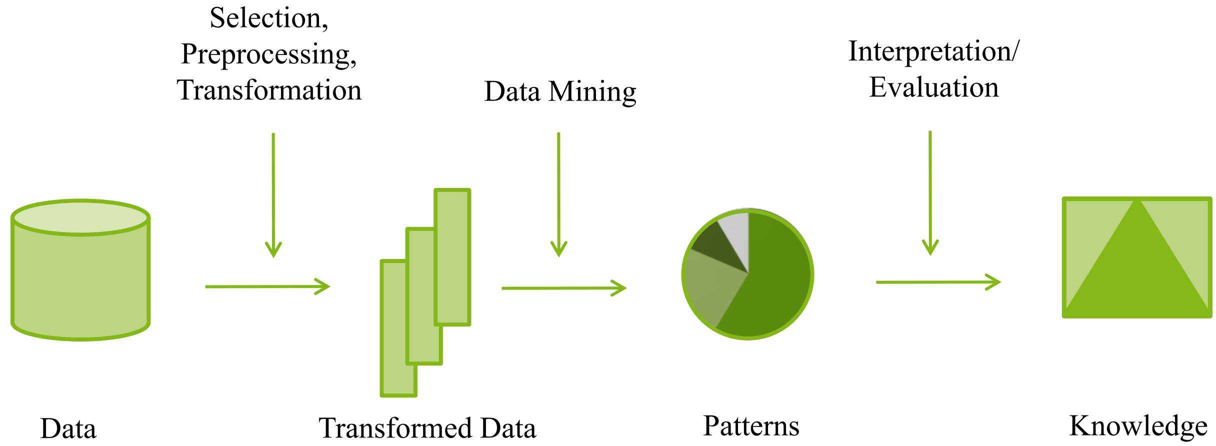

Figure 2: Knowledge Discovery in Databases – Process View (reflecting the work of Fayyad, Piatetsky-Shapiro and Smyth 1996).

Typical ways of expressions are rules, trees or functional relations (Klösgen and Zytgow 1996). A pattern itself is generic, which means it contains placeholders. An instance is created by replacing the placeholders with specific values or by quantifying the variables. A generic pattern can have more than one instance, however the difference between a generic pattern, no matter to which category it belongs, and an instance is important because a particular simulation activity is connected to an instance. A special kind of pattern is the association rule, which describes a relation between two or more features and is well suited for transactions (Agrawel et al. 1993). If we take a look at the SC industry we will find a lot of transactions in the related information systems. For example the transport of a good from supplier to vendor and its possible intermediate stops are often part of the SCs data structure. This transaction consists of different features, such as time stamps or quantities (Moody und Kortink 2000). An example of such a rule can be: *If Supplier "Big Producer" delivers good "Premium" with a quantity of more than 500 and supplier "Commodity producer" delivers good "Top-Quality" with a quantity of more than 1000, the consignments of supplier "Small Producer" are delayed.* In this rule the object's supplier, quantity, consignment and delay are features. Within simulation, these features or their excerpts and aggregations represent the simulation input. But, it is not possible to simply replace the standard input modeling by association rules. The principal reasons for this are the need of a complete input data set for simulation scenarios. A rule just contains a subset of parameters, depending on thresholds and probabilities and might not cover the simulation requirements. Thus, a rule does not automatically lead to a probability distribution.

It should be noted that a rule should not be confused with correlation but they are closely linked, e.g. correlation analyses can be used to find out how useful a rule might be (for more information on rules and correlation see Brin, Motwani and Silverstein 1997). In any way, the knowledge generated through KDD processes seems promising for increasing simulation accuracy. This knowledge may describe a logical relation of parameters which is not covered by the standard input modeling of parameters. We have pointed out that we cannot substitute one for the other, therefore a possible combination should be further examined.

4 COMBINATION – SIMULATION AND KNOWLEDGE DISCOVERY

Combining DES and KDD to support a higher accuracy of simulation input seems to be a promising opportunity. We have identified two possible connections for increasing accuracy, the model and the input data. As a consequence, the acquired knowledge of the KDD process can be connected to both, the data input and the simulation model.

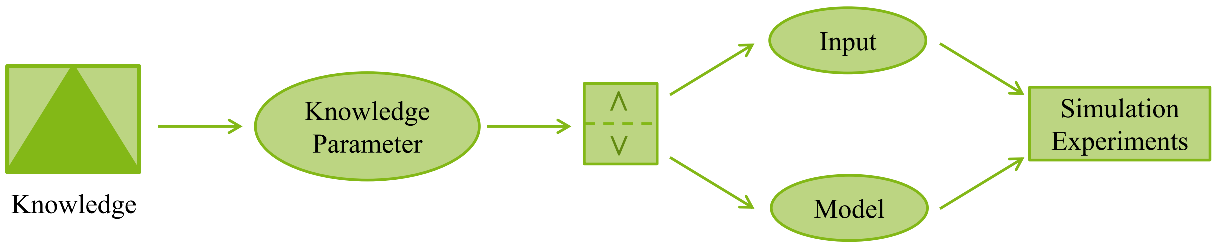

Figure 3: From KDD to simulation experiments.

For this connection, one necessary preprocessing step has been identified (Figure 3), which is connected to a term called "interestingness". This term evaluates how well a particular rule meets a specified KDD goal. For pattern evaluation often objective measures such as interest factors or entropy are used (Tan, Kumar and Srivastava 2004). In KDD processes, interpretation and evaluation of patterns are important for taking the most interesting (Silberschatz and Tuzhilin 1995) rules into account. Furthermore, the DES focus must be taken into account. Because of this reason, it must be clarified what could be an appropriate measurement for interestingness. The combined interestingness feature that covers simulation focus may differ from single KDD processes. Finally, a rule which fits an interestingness measure may need special post processing, e.g. mapping of feature names to make useful additions for input data or simulation modeling. In Figure 3 this area is represented by the cloudy term "Knowledge Parameters", where cloudy implies the previously unknown transformation step from knowledge representation to proper input format for subsequent steps

The appropriate method for these parameters with the related DES areas depends on various influencing factors. As an example, the data mining algorithm itself determines the resulting knowledge and the knowledge can be represented in different ways. Another important factor of influence is the connection object, which will affect the requirements for extension itself. For sure, if the connecting object is the simulation model the requirements will differ from the requirements needed if the connecting object is located at the input data level.

5 FIRST EXPERIMENTS

The data sets supplied at hand consist of aggregated multi-level SC data of a company manufacturing motorcycles. In total the existing database extract is composed of six tables and more than 120 attributes in total. The data contain SC transactions in relation to time, suppliers and products. In the beginning of the examination it immediately became clear that two out of six tables did not have any logical relation to the four remaining tables and after further review it became obvious that these two tables were not going to contribute to the intended experiments. Consequently these tables were dropped. The remaining four tables were comprised of 107 attributes. Several of these attributes were redundant because of the same reason: They had the exact same value in each row. So the total number of attributes was reduced to 67 after dropping the redundant attributes. Looking at maximum and minimum values it became apparent that there were illogical values in some columns. In one column for example we were able to find 52 rows with a timestamp of the year 2201. These values were removed from the database tables as part of the outlier detection. Nine of the 67 attributes were timestamps. However, the exact date and time was not necessary for the examinations so we were able to save a lot of memory space by mapping the timestamps to integer values: The original date was mapped to an integer value representing the difference in days from January 1st 2000 to the day of the timestamp. In further examinations only the difference between certain dates in days was significant so this seems to be an appropriate approach.

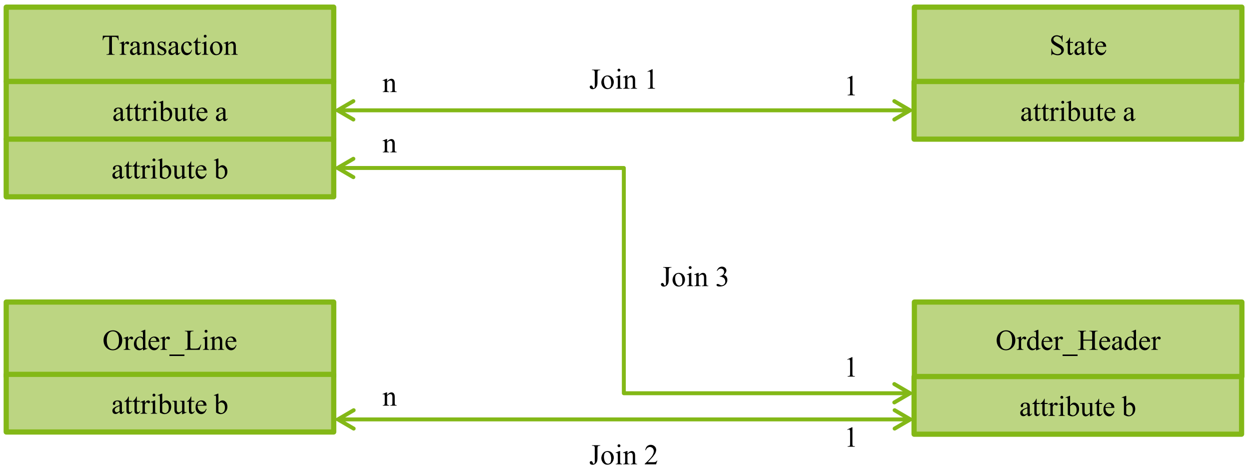

Figure 4: Data model.

The transaction table is the largest table and consists of nearly 1,600,000 rows. Join 1 maps an attribute of this table to the same attribute of the state table (Figure 4). The database table State only contains 180 rows. This leads to the fact that the resulting table consists of the same amount of nearly 1,600,000 rows after this inner join. Join number two connects the tables Order_line (about 300,000 rows) and Order_header (9,000 rows) by an inner join using Order_header's primary key. The resulting table contains over 303,000 rows. These resulting tables of the first joins are joined again using attribute b. The n:1 relationship between Order_line and Order_header leads to an n:n relationship during this inner join. Therefore, the resulting table is expanded to include about 201,000,000 rows. After Join 3 the resulting table contains the previous number of rows and 64 attributes.

The required amount of memory is almost 100 GB. The resulting table can definitely be considered as big data because *"*Big Data is a loosely defined term used to describe data sets so large and complex that they become awkward to work with using standard statistical software*"* (Snijders, Matzat and Reips 2012).

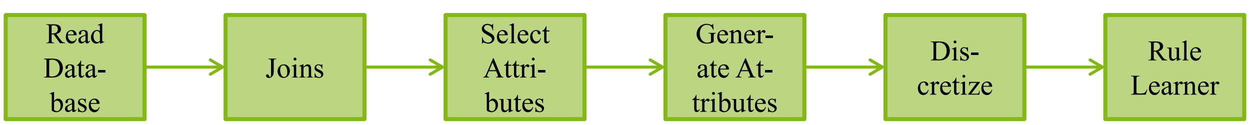

Figure 5: Experiment setup for the first phase.

Figure 5 shows the generic setup for experiment studies in the first phase of the upcoming scientific work. After reading and joining the tables, attributes are selected: By focusing on the relevant attributes for research, it is possible to reduce computation time and memory usage. The next process step is used to create aggregated attributes which represent new features, e.g. the distance between two dates. This step and the next step "Discretize" are used for preparing and transforming the data into an appropriate format. The appropriate format is needed for the main step of the experiment, the data mining algorithm itself. In the experiment setup introduced here, a rule learner is used.

6 CONCLUSION

In this paper, we discuss why simulation needs to take all available knowledge into the account. For this purpose, a new approach for increasing accuracy of DES input has been introduced. This approach is based on data mining which is merged with standard simulation input modeling with the aim to extend the data input or the conceptual model. We motivated why neither distribution fitting nor data mining as isolated step in simulation setup is generally advantageous. We demonstrated the two possible connections for merging KDD and simulation techniques with the aim to generate more realistic input for DES. With that in mind we gave an insight into the first experiment setups in the concluding section. Our next steps will be the extension of the experiment phase, in particular the testing of different hypothesis and data mining methods apart from rule learning. A comprehensive experiment circle will deal with the questions what is the best level in simulation processes for integrating the learned rules - input data or the simulation model itself. A further object of this research will be process automation; in particular for data preprocessing. But, this must be a subsequent step because automation differs in relation to the used data mining input and therefore the target function cannot be defined in this early stage.

REFERENCES

Acken, J. M. 1997. “Logic Level Modeling.” In *Encyclopedia of Computer Science and Technology*, edited by J. G. Williams and A. Kent, 281–306. Boca Ration, Florida: CRC Press.

Agrawal, R., T. Imieliński and A. Swami. 1993. “Mining association rules between sets of items in large databases.” In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, edited by P. Buneman and S. Jajodia, 207–216. New York (U.S.A.): ACM Press.

Akhavian, R. and A. H. Behzadan. 2013. “Automated Knowledge Discovery and Data-driven Simulation Model Generation of Construction Operations.” In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 3030–3041. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Banks, J. 1998. *Handbook of simulation.* New York (U.S.A.): Wiley; Co-published by Engineering & Management Press.

Bogon, T., I. J. Timm, U. Jessen, M. Schmitz, S. Wenzel, A. D. Lattner, D. Paraskevopoulos and S. Spieckermann. 2012. “Towards assisted input and output data analysis in manufacturing simulation: The EDASim approach.” In *Proceedings of the 2012 Winter Simulation Conference - (WSC 2012)*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose and A.M. Uhrmacher, 1–13. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Brin, S., R. Motwani and C. Silverstein. 1997. “Beyond market baskets.” In *Proceedings ACM SIGMOD International Conference on Management of Data 1997*, edited by J. M. Peckman, S. Ram and M. Franklin, 265–276. New York (U.S.A.): ACM Press.

Fayyad, U. M., G. Piatetsky-Shapiro and P. Smyth. 1996. “From Data Mining to Knowledge Discovery in Databases.” In *American Association for Artificial Intelligence*, edited by U. M. Fayyad; G. Piatetsky-Shapiro and P. Smyth, 37–54.

Hand, D. J. 2002. “Pattern Detection and Discovery.” In *Pattern Detection and Discovery*, edited by D. J. Hand, N. M. Adams and R. J. Bolton, 1–12. Berlin, Heidelberg: Springer.

Hellström Daniel and M. Johnsson. 2002. “Using discrete event simulation in supply chain planning.” In *Proceedings of the 14th Annual Conference for Nordic Researchers in Logistics*, edited by O. Solem, S. 13-28.

Juristo, N. and A. M. Moreno. 2000. “Introductory paper: Reflections on Conceptual Modelling.” *Data & Knowledge Engineering* 33:103–117.

Klösgen, W. and J. M. Zytkow. 1996. “Knowledge Discovery in Databases Terminology.” In *Advances in Knowledge Discovery and Data Mining*, edited by U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, 573–592. Menlo Park, Cambridge, London: AAAI/MIT Press.

Law, A. M. 2007. *Simulation modeling and analysis.* 4th ed. Boston (U.S.A.): McGraw-Hill.

Lee, Y.H., M. K. Cho, S. J. Kim and Y. B. Kim. 2002. "Supply chain simulation with discrete-continuous combined modeling." *Computers & Industrial Engineering* 43:375-392.

Lehtonen, J.-M. and U. Seppala. 1997. "A methodology for data gathering and analysis in a logistics simulation project." *Integrated Manufacturing Systems* 8*:*351–358.

März, L., W. Krug, O. Rose and G. Weigert. 2010. *Simulation und Optimierung in Produktion und Logistik.* Berlin, Heidelberg: Springer.

Merkuryeva, G. and O. Vecherinska. 2010. "Randomness Modeling in Supply Chain Simulation." In *2010 International Conference on Intelligent Systems, Modelling and Simulation (ISMS),* 128–133. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Moody, D. L. and M. A.R. Kortink. 2000. "From Enterprise Models to Dimensional Models." In *Proceedings of the Second International Workshop on Design and Management of Data Warehouses, DMDW 2000*, edited by M. A. Jeusfeld, H. Shu, M. Staudt and G. Vossen, 5.1 - 5.12.

Perera, T. and K. Liyanage. 2000. "Methodology for rapid identification and collection of input data in the simulation of manufacturing systems." *Simulation Practice and Theory* 7:645–656.

Preusser, M., C. Almeder, R. F. Hartl and M. Klug. 2005. "LP Modelling and Simulation of Supply Chain Networks." *Supply Chain Management und Logistik*, edited by H.-O. Günther, D. C. Mattfeld and L. Suhl, 95–113. Berlin [u.a.]: Springer.

Rabe, M. and M. Deininger. 2012. "State of Art and Research Demands for Simulation Modeling of Green Supply Chains." *International Journal of Automation Technology* 6:296–303.

Rabe, M., S. Spieckermann and S. Wenzel. 2008.: "A New Procedure Model for Verification and Validation in Production and Logistics Simulation." In: *Proceedings of the 2008 Winter Simulation Conference,* edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson and J. W. Fowler, 1717-1726. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Rahman, F. Abdul, M. Ishak Desa and A. Wibowo. 2011. "A Review of KDD-Data Mining Framework and its Application in Logistics and Transportation." In *2011 7th International Conference on Networked Computing and Advanced Information Management (NCM)*, edited by Y. Cho, S. Kawata and F. Ko, 175–180. Piscataway, New Jersey: IEEE.

Robinson, S. 2004. *Simulation..* Chichester, England, Hoboken, New Jersey: Wiley.

Silberschatz, A. and A. Tuzhilin. 1995. "On Subjective Measures of Interestingness in Knowledge Discovery." In *Proceedings of KDD-95: First International Conference on Knowledge Discovery and Data Mining*, edited by U. M. Fayyad and R. Uthurusamy, 275–281. Menlo Park (U.S.A.): AAAI Press.

Skoogh, A. and B. Johansson. 2009. "Mapping of Time-Consumption During Input Data Management Activities" *Simulation News Europe* 19:39–46.

Snijders, C., U. Matzat and U.-D. Reips. 2012. "Big Data: Big Gaps of Knowledge in the Field of Internet Science." *International Journal of Internet Science.* 7:1-5.

Tako, A. A. and R. Stewart. 2012. "The application of discrete event simulation and system dynamics in the logistics and supply chain." *Decision Support Systems* 52:802-815.

Tan, P.-N., V. Kumar and J. Srivastava. 2004. "Selecting the right objective measure for association analysis." *Information Systems* 29 : 293–313.

Tarokh, M. J. and M. Golkar. 2006. "Supply Chain Simulation Methods." In *2006 IEEE International Conference on Service Operations and Logistics, and Informatics // IEEE International Conference on Service Operations and Logistics, and Informatics,* 448–454. Piscataway, New Jersey: IEEE.

Terzi, S. and S. Cavalieri. 2004. "Simulation in the supply chain context: a survey." *Computers in Industry* 53: 3-16.

Wand, Y., D.E. Monarchi, J. Parsons, and C.C. Woo. 1995. "Theoretical foundations for conceptual modelling in information systems development." *Decision Support Systems* 15:285–304.

Wenzel, S. and J. Bernhard. "Information Acquisition for Model Based Analysis of Large Logistics

Networks." In *Proceedings 19th European Conference on Modelling and Simulation,* edited by Y. Merkuryev, R. Zobel and E. Kerckhoffs, 37–42. Nottingham, UK: European Council for Modelling and Simulation.

Witten, I. H., E. Frank and M. A. Hall. 2011. *Data mining.* 3rd ed. Burlington, Massachusetts: Morgan Kaufmann.

AUTHOR BIOGRAPHIES

MARKUS RABE is full professor for IT in production and logistics at the Technical University Dortmund. Until 2010 he had been with Fraunhofer IPK in Berlin as head of the corporate logistics and processes department, head of the central IT department and member of the institute direction circle. His research focus is on information systems for supply chains, production planning and simulation. Markus Rabe is vice chair of the "Simulation in Production and Logistics" group of the simulation society ASIM, member of the editorial board of the Journal of Simulation, member of several conference program committees, has chaired the ASIM SPL conference in 1998, 2000, 2004, and 2008, and was local chair of the WSC'2012 in Berlin. More than 150 publications and editions report from his work. His e-mail address is markus.rabe@tu-dortmund.de.

ANNE ANTONIA SCHEIDLER is a research assistant since 2012 at IT in production and logistics. Until 2012 she worked as IT consultant for different large companies. Key areas of her activity were designing business processes and conception of data imports and models. She studied computer science at the Technical University Dortmund and graduated in 2008 with a thesis on forecasting of time series using local and global data patterns. Her research focus is on data mining, input data, data in simulation and supply chain information. Her e-mail address is anne-antonia.scheidler@tu-dortmund.de.