TOWARDS ASSISTED INPUT AND OUTPUT DATA ANALYSIS IN MANUFACTURING SIMULATION: THE EDASIM APPROACH

Tjorben Bogon, Ingo J. Timm

University of Trier
Behringstraße 21
D-54296 Trier, GERMANY

Andreas D. Lattner, Dimitrios Paraskevopoulos

Goethe University Frankfurt
P.O. Box 11 19 32
D-60054 Frankfurt/Main, GERMANY

Ulrich Jessen, Markus Schmitz, Sigrid Wenzel

University of Kassel
Kurt-Wolters-Straße 3
D-34125 Kassel, GERMANY

Sven Spieckermann

SimPlan AG
Edmund-Seng-Str. 3-5
D-63477 Maintal, GERMANY

ABSTRACT

Discrete-event simulation has been established as an important methodology in various domains. In particular in the automotive industry, simulation is used to plan, control, and monitor processes including the flow of material and information. Procedure models help to perform simulation studies in a structured way and tools for data preparation or statistical analysis provide assistance in some phases of simulation studies. However, there is no comprehensive data assistance following all phases of such procedure models. In this article, a new approach combining assistance functionalities for input and output data analysis is presented. The developed tool – EDASim – focuses on supporting the user in selection, validation, and preparation of input data as well as to assist the analysis of output data. The proposed methods have been implemented and initial evaluations of the concepts have led to promising feedback from practitioners.

1   INTRODUCTION

In recent years, discrete-event simulation has been established as an important methodology in various domains (Wenzel, Boyaci, and Jessen 2010). In particular in the automotive industry, simulation is used to plan, to control, and to monitor processes including the flow of material and information (Mayer and Spieckermann 2010). The high complexity of systems under examination and their corresponding simulation models lead to an increasing demand for assistance at different phases of simulation studies.

Simulation studies are often performed following procedure models which provide guidelines for the professional performance of simulation studies (Rabe, Spieckermann, and Wenzel 2008a). For some phases of such models, there exist supporting tools, for instance, to prepare data or to perform statistical analyses. However, there is no comprehensive support following all phases of such procedure models. In this paper, results of a collaboration between the University of Kassel, Goethe University Frankfurt, University of Trier, SimPlan AG, INCONTROL Simulation Solutions, UST Umweltsensortechnik GmbH, Continental Automotive, and VDA German Association of the Automotive Industry (industry workgroup "Ablaufsimulation – Process Simulation") are presented. In two collaborative research projects, methodologies have been developed in order to provide assistance functionalities for specific phases of a simulation study (cf. Figure 1).

The initial project AssistSim (Development and Evaluation of Assistance Functionalities for the Simulation of Logistic Processes; HessenAgentur Project No. 185/09-15) addresses support during the design and execution phase of simulation studies. In AssistSim, the experimental design of simulation studies is supported by a structured capturing of experiments' objectives of investigation as well as corresponding variables and measurements. Integrated checklists facilitate the identification of needed variables. A mapping of the identified variables and measurements to actual input and output data of specific simulation models enables the second main feature of AssistSim – the simulation execution assistance. Simulation models are executed automatically with desired numbers of replications, setting input values as specified in the experimental design phase and capturing all results in the AssistSim database. More details can be found in (Lattner et al. 2011). One result of the evaluation of AssistSim was a significant demand for assistance in the handling, analysis, and interpretation of input and output data. Therefore, the project EDASim (Development of Data Assistance for Simulation Studies in Production and Logistics; HessenAgentur Project No. 260/11-06) focuses on the assistance of the selection, assessment, and preparation of input data as well as the analysis of output data.

This paper presents a new approach to an assistance of data analysis in the context of simulation studies ranging from input data acquisition and preparation to output data analysis. Furthermore an architecture and prototype are presented.
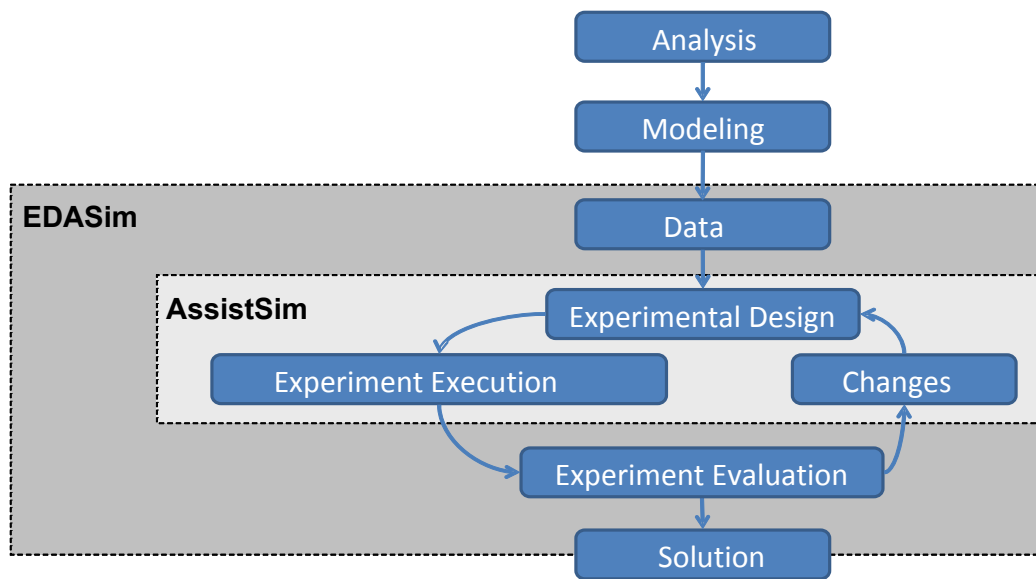

Figure 1: Focus of AssistSim and EDASim within a simulation study (restricted to operative tasks)

The paper is organized as follows: in section 2, a brief overview of related work is provided with focus on the support of data analysis regarding simulation input data as well as output data. The current state of the art analysis from a simulation expert perspective is presented in section 3. A new approach to support input and output data analysis is introduced in Section 4. The resulting software architecture and prototype of EDASim are introduced in Section 5. The article closes with a discussion of the results and future work.

2   RELATED WORK – SUPPORT FOR DATA ANALYSIS

The results of an online survey, which was sent to 181 simulation experts and users within the VDA (German Association of the Automotive Industry) in the year 2010, point out that the effort of data acquisition in simulations studies is estimated to be between 5 and 50 percent of the overall project effort; the calculated median is 20 percent (Huber and Wenzel 2011). Additionally, the effort for verification and validation of data and models uses up to 21 percent of the overall project effort. Kuhnt and Wenzel (2010) give the following reasons for the significant efforts to acquire required information: missing data, limited access to data sources and low quality of collected data. Furthermore, they point out, that there is a lack of standardized or best-practice checklists and procedure models, which are applicable for information acquisition in simulation studies. The relevance of input data and the impact of input data quality for the quality of simulation results are discussed, for example, in Rabe and Hellingrath (2001), Skoogh and Johansson (2008), VDI 3633, Part 1 (2010), Wenzel and Bernhard (2008), and Kuhnt and Wenzel (2010). In this context, Wenzel et al. (2008) point out, that non-representative data or uncertain data cause non predictable effects to simulation results and therefore can lead to doubtful decisions concerning the system. Related work investigates the impact of quality of information for a simulation study on the quality of results (Wenzel, Abel, and Willmann 2011).

However, it has to be kept in mind that the research on use and impact of information and data in simulation studies stands in the broader context of research on information and data quality in general (see, for example, Hildebrandt 2004, Eppler and Witting 2000, Eppler 2006, Naumann 2007).

First research addressing the automated analysis of experimental data has already been performed in the 1950s (cf. Valdes-Perez 1999). In the following decades, the field of automated scientific discovery has evolved which has a strong overlap to the field of data mining (Williamson 2010). In several articles, approaches to knowledge discovery in various domains have been considered, for instance, for the identification of physical laws (Schmidt and Lipson 2009) and for the automation of biochemical experiments (King et al. 2009).

Some of the aforementioned publications address the analysis of output data of simulation experiments. The system "Explora" for knowledge discovery provides support to identify interrelations in results of simulation studies for planning political decisions. These results include expected consequences if certain laws were adapted, e.g., what types of households would benefit most from new regulations with respect to children's allowance (Klösgen 1996). Better, Glover, and Laguna (2007) present an approach to simulation optimization where data mining methods are used during optimization in order to separate good quality from bad quality solutions in the parameter space. A similar approach is introduced by Burl et al. (2006) where active learning with support vector machines is used to identify parameter configurations for future simulation runs which are expected to lead to a high information gain. Robinson (2005) presents a tool based on Microsoft Excel which can be coupled to the simulation software SIMUL8 for the analysis of simulation output data. Recent works address assistance functionality in order to estimate the needed number of replications or to estimate if significant differences between two variants are expected in simulation studies (Hoad, Robinson, and Davis 2009, Lattner, Bogon, and Timm 2011).

In the work presented in this paper, operational support for structured capturing and assessment of input data is provided. In this regard, the aforementioned studies addressing input data quality constitute the basis for the requirement analysis of corresponding software components of the EDASim approach. For the analysis of output data, besides established analysis methods, the application of sequential pattern mining algorithms is proposed in this approach. Rules extracted from such frequent sequential patterns are similar to the ones presented in the work of Klösgen (1996), but can also reflect a temporal ordering of events. In order to apply these data mining algorithms to manufacturing simulation data, quantitative measurements with timestamp information need to be transformed to a qualitative representation. Furthermore, the flexible plug-in framework of EDASim enabling to easily extend the system by adding new components is a distinguishing feature of our approach in comparison to the aforementioned works.

3   PRESENT SITUATION ANALYSIS FROM A SIMULATION EXPERT PERSPECTIVE

Assisting simulation experts is not only an interesting approach from a research point of view, but is also of high relevance to application and use of simulation in an industry context. The vision of being able to guide and assist a simulation expert in a holistic manner throughout all steps of a simulation study would be a substantial progress compared with today's way of working. The high stake the VDA with engineers from the car manufacturers Audi, BMW, Daimler, Opel, and VW and the automotive suppliers Bertrandt, Continental, EDAG, Schäffler, and ZF have in both of the consecutive research activities is in itself an indicator for the relevance of the research activities for practitioners. However, the main motivation for the participating companies is different: even though procedure models for simulation studies are common sense and are widely used throughout the industry, executing the procedures, i.e., carrying out a simulation study is still a process mainly depending on the skills and experiences of the simulation experts involved. Also, capturing meta-information (e.g., the collection, preparation, storage) and documentation of simulation input and output data are rather performed individually and in many cases manually. Statistical procedures are – if applied at all – performed either in spreadsheet software, or within the simulation software itself resulting in an often poor documentation of test results, confidence intervals etc. All in all, the manual processing of the procedure steps, data handling, and experimentation leads to problems in terms of loss of efficiency or loss of knowledge about the simulation study. At the same time simulation models in the automotive industry sometimes live for several years, while the simulation expert responsible for the model changes (possibly several times) during this period (Mayer and Spieckermann 2010). In these and similar situations it is not only important to follow a simulation procedure model, but it is relevant to know, e.g., the background of modeling decisions taken one, two, or five years ago, the specific user requirements which led to the decisions, the origin of the data used in the model, the way the results were created etc.

From a technical point of view common procedures to data analysis in simulation experiments are based on the brute force method: before simulation it is important to prepare the input data, such that meta-information is linked to the data. Collected data sets are validated to ensure valid simulation runs and to get feasible output data. Furthermore, the simulation expert has to pre-process and to select input data for simulation by applying various statistical methods to clean-up and prepare the data set for simulation. Additionally, the data have to be connected to the simulation model, which requires detailed domain knowledge. After performing the experiments, the simulation expert has to run various statistical tests and statistical models to summarize and to prepare the experiment evaluation. Numerous test results are created and have to be screened. All these steps must be documented to get a report and to summarize all relevant results. Only users with a certain degree of expertise are able to generate high quality reports of simulation studies and even for them, a support in this process is desired. Especially the interpretation of output data includes major challenge as interesting results can be missed in the resulting "crowd of information".

4   EDASIM: A NEW APPROACH TO INPUT AND OUTPUT DATA ANALYSIS

EDASim addresses an integrated input and output data analysis in the context of simulation studies. Therefore, it contains assistance functionality for input data (acquisition, collection and preparation, meta-information annotation, and data validation), statistical analysis and inference, data visualization and reporting tools. In the following two subsections more details on the EDASim approach on input data as well as on output data analysis and management are discussed.

4.1   Information Acquisition and the Impact of Input Data Quality

At the beginning of the information acquisition for a simulation study it is necessary to clarify which input data are generally necessary to achieve the goals of the study, and which quality these data need to have in order to be credible. Models and results in simulation are then referred to as credible, "if enough indicators, indications, arguments or perhaps even evidence are available that show that the simulation

model fulfills the purpose intended, as far as possible" (translated from German original: Berchtold et al. 2002, p. 135). According to Wenzel and Bernhard (2008) the fulfillment of enterprise-specific, project-based or individually defined acceptance criteria is a measure of the credibility that the user assigns to the model (see also Robinson 2004, Rabe, Spieckermann, and Wenzel 2008b, Balci et al. 2000). Transferring this statement to credibility of information and data it seems to be relevant to ask when is credibility of data needed and how it can be achieved and identified.

According to the differentiation of the terms "Verification and Validation (V&V)" discussed in the literature (cp. Balci 2003, Rabe, Spieckermann, and Wenzel 2009), verification ("Are the data correct?") deals rather with syntactical questions ("Is the data format correct? Are the data accurate and consistent?") and validation ("Are that the right data?") with questions in terms of correctness of data relating to system pre-settings and to questions about the suitability for the appropriate task. Analogous to the V&V in the simulation modeling process Kuhnt and Wenzel (2010) as well as Wenzel and Bernhard (2008) specify in their works in the Collaborative Research Centre "Modeling of Large Logistics Networks" (SFB 559), supported by the Deutsche Forschungsgemeinschaft, that the entire process of information acquisition (identification of the necessary information, information collection, if necessary data recording, statistical data analysis and data usability test) has to be combined with V&V of the respective results of the steps. Finally, the process of information acquisition has to be finalized with the question whether the identified data satisfy the qualitative information needs of the later application.

In this context Bernhard, Dragan, and Wenzel (2007) define information quality as the set of quality attributes of information or of an information process, which refers to its ability to satisfy given requirements. To assess the quality of information, they define quality criteria for the content, the meaning, the origin and the use of information:

1. Content-related aspects of the quality of information refer to the purely technical information quality (recording errors and gaps, corruptions) and therefore to the correctness of the information, independent of the meaning of information, satisfaction of information needs (completeness) and the degree of the independence of the information content of the persons involved (objectivity).

2. Aspects of the meaning of information refer to the quality of information regarding the interpretability and hence the contextual quality. These aspects comprise the timeliness of the information (currency), the degree of clearness and clarity of the information content (comprehensibility) and the degree of coverage of the necessary information needs (congruence).

3. Aspects of the origin are used to evaluate the source of information and thus the expected quality of the information contained therein. The availability and accessibility of information in principle, but also the reconstruction of the process of gathering information (traceability) and the reliability of the information from the source are evaluated. If multiple information sources are used, in addition, the degree of similarity of structure and semantics of information (homogeneity) can be important.

4. With respect to the utilization of information, the correspondence between the collected and prepared data and the requirements from the application are checked. The importance of information for the application (relevance) and the suitability in terms of the granularity of the data and the use of the data in the method respectively in the simulation tool used have to be ensured.

The above theoretical preliminary work and industry-related research to obtain databases for simulation studies (cp., for example, Bockel, Csanady, and Wenzel 2008) are incorporated into the research project EDASim to provide assisting functionality that leads to an increase in the quality of input data. In order to address item 4 of the list above, EDASim provides checklists to systematically collect and prioritize information needs for a simulation right from the start. This helps, e.g., to document the rele-

vance of the information for a specific objective within the scope of a project itself but also for re-use for future projects.

Furthermore, EDASim functions make sure that the information sources within an enterprise can be administrated with the help of meta-information and be linked to the information required for the simulation study, so that the coverage of required information with the actual available information can be assured (congruence). At the same time, the evaluation of information regarding objectivity and actuality is supported (item 1 and 2) and traceability of information acquisition (item 3) is maintained.

With additional functionalities providing statistical methods, the user receives assistance to ensure the completeness, correctness and accuracy of its input data (item 1). These functionalities provide methodological expertise, which is needed to identify possible errors in data. In this manner, an important step in support of the input data verification can be achieved.

4.2 Output Data and Output Data Analysis

Besides the aforementioned task to identify needed sources for simulation studies and to assess the input data quality, another challenge is to analyze simulation output data and to support the expert to draw conclusions from the results. Complex models can consist of numerous parameters as well as a number of measurements. Multiple simulation runs including replications of particular variants can originate huge amounts of output data. In order to avoid "getting lost" in these data, support in the form of statistical analysis as well as the identification of patterns in the data are desired in this context.

The approach for output data analysis, as implemented in EDASim, is intended to be generic in the sense that results from arbitrary simulation models can be examined, e.g., the buffer states of simulated factories or the output of a production line. The simulation models under consideration typically are stochastic and hence will use random variables. Besides simulation output data that are measured once (e.g., a factory's output at the end of a simulation run), there can be time-dependent measurements (like the capacity level of a buffer over time).

With respect to the analysis of the output data generated by user-defined simulation studies, EDASim provides functionalities that include the computation of certain statistics (mean value, standard deviation, confidence intervals, etc.), significance tests, outlier identifications, distribution fitting, identification of interrelations (supervised and unsupervised learning), and an automated categorization of numeric values. The corresponding parameters which are relevant for the particular functionality are stored in a common database. The user has the possibility to change these values, e.g., the error rate of the confidence interval or of the significance test, both temporary and permanent for specific projects.

Although other tools are available providing statistical functionalities, it is useful to provide an integrated solution in order to avoid system discontinuities, i.e., that the user has to deal with multiple programs for the analysis. Concerning the assessment of data accuracy, a functionality has been designed which computes the confidence interval of data. This functionality provides in addition to the interval, basic statistical values like arithmetical mean or standard deviation and a chart of the single values as well as of the confidence interval. For the performance of significance tests from two samples, a t-test has been integrated (paired/unpaired sample, one-/two-tailed test). An interesting aspect regarding output results of simulations is to estimate the existing distribution of measurements. For this purpose, the Chi-squared test is used (e.g., Greenwood and Nikulin 1996). For more details about the distribution fitting approach see (Lattner et al. 2012).

In order to identify regularities in simulation output data, supervised as well as unsupervised machine learning algorithms are integrated into our approach. In the first case, a supervised learning task is set up to train a classifier using results of performed simulation runs. The use of symbolic machine learning approaches like decision tree or decision rule learning is proposed as they lead to comprehensible descriptions which can be interpreted by human decision makers (e.g., Quinlan 1993, Cohen 1995). Such a trained classifier can be used, for instance, to separate good quality solutions (regarding a measurement like factory output) or to find descriptions for certain situations (e.g., simulation runs where a specific buffer capacity is exceeded or a machine's workload is very low). These classifiers can be used to provide

a general description about regularities of the different classes or can later be used to predict certain outcomes of settings even without performing simulation runs. The predicted outcome can be of interest if it has to be decided whether certain settings should be taken into account for simulation. A filtering based on such a classification can thus help to focus computational time on those settings which are expected to be more interesting.

For the unsupervised identification of regularities, an approach to sequential pattern mining is integrated. These pattern mining approaches take as input a sequence of "events" and identify frequent patterns which then can be transformed into (temporal) association rules like "If breakdown(machine7) then empty(buffer2) and low_workload(machine9)". In order to generate the needed input, events and their corresponding timestamps have to be extracted from simulation runs. In our work, an implementation of RuleGrowth (Fournier-Viger et al. 2011) of the Sequential Pattern Mining Framework provided by Fournier-Viger (2012) is used in order to identify frequent sequential patterns.

5   EDASIM ARCHITECTURE AND PROTOTYPE

In the previous section, the fundamental concepts of EDASim have been introduced. These conceptualizations have led to the development of an assistance system for users of simulation software. In this section the architecture of the prototype and the prototype itself are introduced.

EDASim is based on a flexible architecture (see Figure 2), which is characterized by the EDASim database, the plug-in container, and the GUI. As a programming language, C# has been chosen for implementation of the core architecture. The underlying and integrating component for the assistance system is the EDASim database. Here, the assistance process is logged; project specific details and information about the system configuration are stored. Related to the database the core system's behavior is controlled. The conceptual design of EDASim provides a modular, extensible structure where different functionalities can be integrated as so called plug-ins. Plug-ins provide certain functionalities and can be loaded dynamically to the EDASim main system using a specific interface. As a result of this component-oriented design, the programming language of the plug-in is not restricted by the EDASim system. In our project, plug-ins have been developed in C# as well as Java. The plug-in interface specifies methods for GUI layout as well as required methods for communication with other plug-ins and the EDASim database. Additionally, plug-in developers have to implement mandatory functions in the plug-in used by the EDASim plug-in container for recognizing the new plug-in. As a result of the plug-in mechanism and specification, the system can easily be extended as desired. The assistance functionalities and developed plug-ins of EDASim range from data handling, statistics, and visualization and are introduced in the following in more detail.

The first cluster of assistance functionalities deals with data management (plug-ins for database access, data table visualization, and selection of data sources). Functionalities for meta-information annotation (plug-in to annotate information about data sources) are included. The next cluster of functionality is dedicated to statistics. The assisted analysis of input data as well as output data is based upon the application of statistical methods. Among others, in the domain of logistics it could be important to provide algorithms for confidence intervals, statistical tests, categorization, pattern mining, and outlier identification. All these algorithms are integrated and supported by basic statistics plug-ins and offer standard configurations as a starting point. In conjunction with the previous functionalities it is important to provide a visual access to analysis and simulation results. Therefore, data visualization plug-ins are integrated in EDASim, namely Gantt charts and multi-level pie charts. One valuable extension would be the assistance for an automatic generated documentation of the simulation runs with the identification of the important results to generate an overview about the performed study supporting beginners as well as professionals.
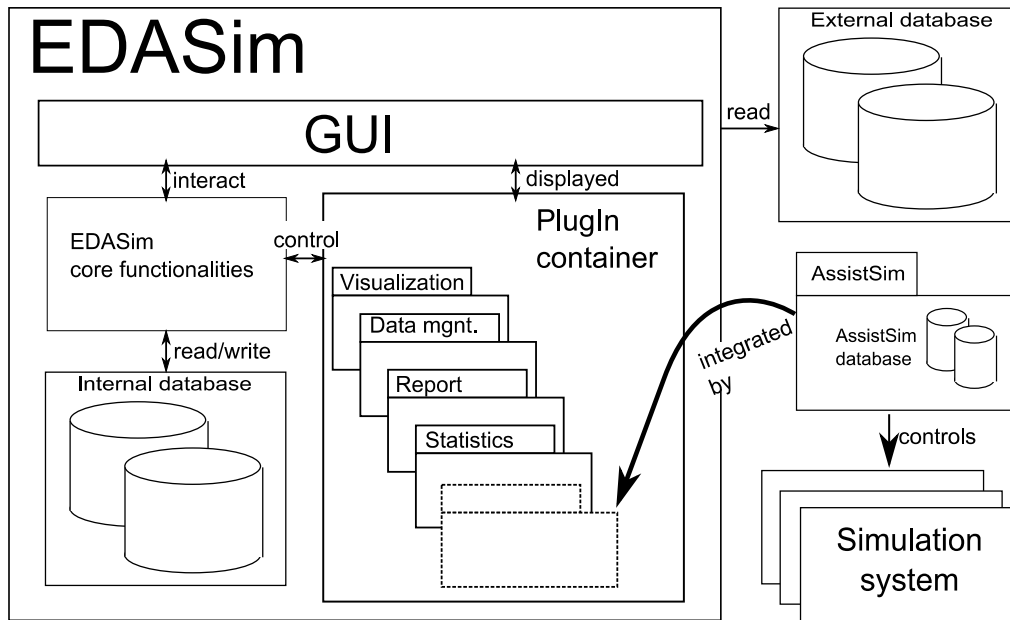
Figure 2: Sketch of the EDASim architecture

The principle of the developed assistance system is that different projects can easily be managed and worked with. If a new project is created, specific requirements will arise and this potentially will lead to new functionality needs. Therefore, the major design principle of EDASim is to enable maximum flexibility by means of a sophisticated plug-in mechanism as introduced above. This technique allows for a project-specific configuration of plug-ins. The user can apply a predefined set of configurations (so-called standard setups) or can set up plug-in parameters individually. While users with less experience can retain default parameters, experts can change configurations as requested. Plug-in interfaces are documented well such that experienced users may implement new plug-ins. Furthermore, the plug-in mechanism allows for various programming languages to be used (e.g., C# or Java). Following the logic of plug-ins, the AssistSim system which has been developed in a preceding research activity can be integrated as a plug-in to EDASim as well. Doing so, the automated control of performing simulation runs is utilized in EDASim.

To support the specific form of dynamic functionality in EDASim, a flexible GUI concept is necessary. Every plug-in can be added to the project out of the list of all plug-ins by a drag and drop mechanism. Every plug-in offers its functionality on an encapsulated extra frame in the program and lets the users adjust their specific settings. All plug-ins loaded to the current project can interact with each other using a communication interface provided by the main application. The data of different types of loaded databases can be used by different plug-ins in the project. Furthermore, if the user wants to work with a specific set of data from the database, he can specify a query as needed and store it under a user-defined identifier. This data view (query) can be used in other plug-ins as input.
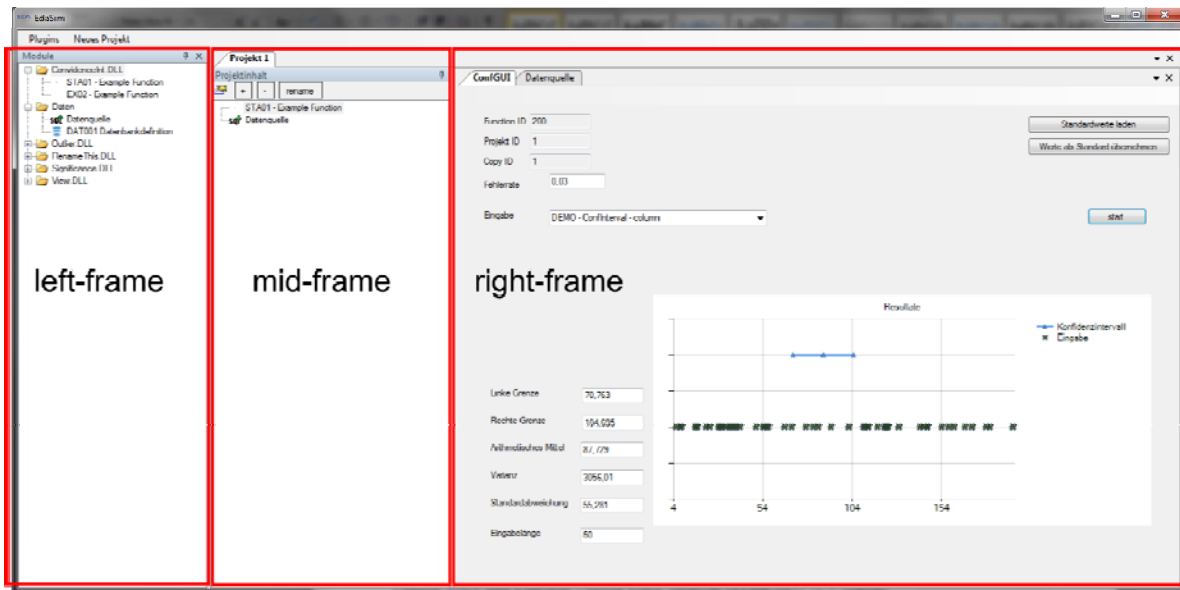
Figure 3: GUI of the EDASim system

The GUI of the EDASim system is split up into three frames (cf. Figure 3). On the left frame all loaded plug-ins are listed. In the mid-frame the current project and all other loaded projects are displayed. If the user wants to integrate plug-in functionality into the current project, it can be integrated via drag and drop from the plug-in-frame or from another project adapting the plug-in's configuration of this project. In the right frame the user dialogue of the currently active plug-in is shown. Here, all the user specific and plug-in depending settings can be done to configure the behavior of the plug-in. For example, to calculate the confidence interval the input dataset has to be chosen. All GUI frames of projects and plug-ins can be used as part of the main window as seen in Figure 3 or as a separate window.

6   DISCUSSION AND FUTURE WORK

Based on the analysis in cooperation with experts from practice in simulation we have identified relevant shortcomings in the process of input and output data handling. As a next step, assistance functionalities have been conceptualized and an architecture for an assistance system has been developed. The software architecture is implemented as a modular, extensible structure. Different plug-ins for statistics, visualization and information management for data sources are available. The implementation of additional individual plug-ins is possible. The prototype shows the feasibility of our concept, evaluations on realistic simulation scenarios are done for in-depth analysis of the EDASim approach. The cooperation of a simulation software vendor and a simulation service provider within the research consortium has turned out to be very fruitful. One of the simulation companies has committed on pushing the development of the EDASim approach into the direction of a commercial software roll-out. The company also intends to move pre-existing software for visualization of experimental results to the new platform. The commercialization of AssistSim and EDASim ideas is accelerated by the commitment of the consortium member VDA to support the activities beyond the formal end of the publicly funded research.

Continental, and VDA German Association of the Automotive Industry (industry workgroup "Ablaufsimulation – Process Simulation").

REFERENCES

Balci, O., W. F. Ormsby, J. T. Carr III, and S. D. Saadi. 2000. "Planning for verification, validation, and accreditation of modeling and simulation applications." In *Proceedings of the 2000 Winter Simulation Conference*, Edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 829-839. Piscataway: IEEE.

Balci, O. 2003. "Validation, verification, and certification of modeling and simulation applications." In *Proceedings of the 2003 Winter Simulation Conference*, Edited by S. Chick, P. J. Sanchez, E. Ferrin, and D. J. Morrice, 150-158. Piscataway NJ: IEEE.

Bernhard, J., M. Dragan, S. Wenzel. 2007. "Bewertung der Informationsgüte in der Informationsgewinnung für die modellgestützte Analyse großer Netze in der Logistik." Technical Report 07006 – Sonderforschungsbereich 559 "Modellierung großer Netze in der Logistik". Dortmund, ISSN 1612-1376.

Better, M., F. Glover, and M. Laguna. 2007. "Advances in analytics: Integrating dynamic data mining with simulation optimization." *IBM Journal of Research and Development* 51(3-4): 477-487.

Berchtold, C., D. Brade, M. Hofmann, A. Köster, T. Krieger, and A. Lehmann. 2002. "Verifizierung, Validierung und Akkreditierung von Modellen und Simulationen." BMVG-Studienauftrag Nr. M/GSPO/Z0076/9976, Abschlussbericht. München: ITIS e.V..

Bockel, B., K. Csanady, and S. Wenzel. 2008. "Methodik zur systematischen Informationsgewinnung für Simulationsstudien." In *Advances in Simulation for Production and Logistics Applications*, Edited by M. Rabe, 959-604. Stuttgart: Fraunhofer-IRB Verlag.

Burl, M. C., D. DeCoste, B. L. Enke, D. Mazzoni, W. J. Merline, and L. Scharenbroich. 2006. "Automated knowledge discovery from simulators." In *Proceedings of the Sixth SIAM International Conference on Data Mining*, Edited by J. Ghosh, D. Lambert, D. Skillicorn, and J. Srivastava, 82-93. Bethesda, MD, USA: SIAM.

Cohen, W. W. 1995. "Fast effective rule induction". In *Proceedings of the 12th International Conference on Machine Learning*, Edited by A. Prieditis and S. J. Russell, 115-123. Morgan Kaufmann.

Eppler M. J. and D. Wittig. 2000. "Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years." In *Proceedings of the International MIT Conference on Information Quality*, Edited by B. D. Klein and D. F. Rossin, 83-96. Cambridge.

Eppler M. J. 2006. *Managing information quality*, 2nd ed. Berlin Heidelberg: Springer.

Fournier-Viger P., R. Nkambou, and V. S.-M. Tseng. 2011. "RuleGrowth: Mining Sequential Rules Common to Several Sequences by Pattern-Growth." In: *Proceedings of the 2011 ACM Symposium on Applied Computing*, Edited by W. C. Chu, W. E. Wong, M. J. Palakal, and C.-C. Hung, 956-961. New York: ACM.

Fournier-Viger, P. 2012. Sequential Pattern Mining Framework (SPMF). Accessed March 3. http://www.philippe-fournier-viger.com/spmf/.

Greenwood, P. E. and M. S. Nikulin. 1996. *A guide to chi-squared testing*. New York: Wiley.

Hildebrand, K. 2004. "Datenqualität im Supply Chain Management." In *Informatik 2004 – Informatik verbindet*, Band 1, Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Edited by P. Dadam and M. Reichert, 239-243. Bonn: Gesellschaft für Informatik e.V.

Hoad, K., S. Robinson, and R. Davies. 2009. "Automated selection of the number of replications for a discrete-event simulation." *Journal of the Operational Research Society* 61(11):1632-1644.

Huber, L. and S. Wenzel. 2011. "Trends und Handlungsbedarfe der Ablaufsimulation in der Automobilindustrie." *Industrie Management* 27 (5): 27-30.

King, R. D., J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. N. Soldatova, A. Sparkes, K. E. Whelan, and A. Clare. 2009. "The automation of science". *Science* 324 (5923): 85-89.

Klösgen, W. 1996. "Explora: A multipattern and multistrategy discovery assistant." In *Advances in knowledge discovery and data mining*, Edited by U. M. Fayyad, G. Piatetsky-Shapiro, and R. Uthurusamy, 249-271. Menlo Park: AAAI Press.

Kuhnt, S. and S. Wenzel. 2010. "Information Acquisition for Modelling and Simulation of Logistics Networks." *Journal of Simulation* 4:109-115. Published online 6. Nov. 2009, doi10.1057/jos.2009.9.

Naumann, F. 2007. "Datenqualität." *Informatik Spektrum* 30 (1): 27-31.

Lattner, A. D., H. Pitsch, I. J. Timm, S. Spieckermann, and S. Wenzel. 2011. "AssistSim - Towards Automation of Simulation Studies in Logistics." Appears in *Simulation News Europe*.

Lattner, A. D., T. Bogon, and I. J. Timm. 2011. "An Approach to Significance Estimation for Simulation Studies." In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART'2011), Edited by J. Filipe and A. Fred, 177-186. SciTePress.

Lattner A. D., J. Dallmeyer, D. Paraskevopoulos, and I. J. Timm. 2012. "Approximation of Pedestrian Effects in Urban Traffic Simulation by Distribution Fitting", In: *Proceedings of the 26th European Conference on Modelling and Simulation (ECMS 2012)*, Edited by K. G. Troitzsch, M. Möhring, U. Lotzmann, 567-573. European Council for Modelling and Simulation.

Mayer, G. and S. Spieckermann. 2010. "Life-cycle of simulation models: requirements and case studies in the automotive industry." *Journal of Simulation* 4: 255-259.

Quinlan, J. R. 1993. *C4.5 - Programs for Machine Learning.* Morgan Kaufmann Publishers, Inc.

Rabe, M. and B. Hellingrath. 2001. *Handlungsanleitung Simulation in Produktion und Logistik.* San Diego: SCS.

Rabe, M., S. Spieckermann, and S. Wenzel 2008a. "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.Hill, L. Mönch, O. Rose, T. Jefferson, and J.W. Fowler: 1717-1726. San Diego: IEEE Press.

Rabe, M., S. Spieckermann, and S. Wenzel. 2008b. *Verifikation und Validierung für die Simulation in Produktion und Logistik – Vorgehensmodelle und Techniken.* Berlin: Springer.

Rabe, M., S. Spieckermann, and S. Wenzel. 2009. "Verification and Validation Activities within a New Procedure Model for V&V in Production and Logistics Simulation." In *Proceedings of the 2009 Winter Simulation Conference*, Edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 2509-2519. IEEE Omnipress.

Robinson, S. 2005. "Automated analysis of simulation output data." In *Proceedings of the 2005 Winter Simulation Conference*, Edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 763-770.

Robinson, S. 2004. *Simulation: The Practice of model development and use.* Chichester: John Wiley & Sons.

Schmidt, M. and H. Lipson. 2009. "Distilling free-form natural laws from experimental data." *Science* 324 (5923): 81-85

Skoogh, A. and B. Johansson. 2008. "A Methodology for Input Data Management in Discrete Event Simulation Project." In *Proceedings of the 2008 Winter Simulation Conference*, Edited by S. J. Mason, R. Hill, L. Mönch, O. Rose, 1727-1735. San Diego: IEEE Press.

Valdés-Pérez, R. E. 1999. "Principles of human-computer collaboration for knowledge discovery in science." *Artificial Intelligence* 107 (2): 335-346.

VDI 3633, Part 1. 2010. *Simulation von Logistik-, Materialfluss- und Produktionssystemen. Grundlagen.* Simulation of Systems in Material Handling, Logistics and Productions. Fundamentals. Berlin: Beuth.

Wenzel, S., D. Abel, C. Willmann. 2011. "Wissensarbeit in der Digitalen Fabrik - Der Zwiespalt zwischen Systematisierung und Kreativität." In *Wissensarbeit – Zwischen strengen Prozessen und kreativem Spielraum.* Schriftenreihe der Hochschulgruppe für Arbeits- und Betriebsorganisation e.V. (HAB), Edited by D Spath, 251-276. Berlin: GITO-Verlag.

Wenzel, S. and J. Bernhard. 2008. "Definition und Modellierung von Systemlasten für die Simulation logistischer Systeme." In *Beiträge zu einer Theorie der Logistik*, Edited by P. Nyhuis, 487-509. Berlin: Springer.

Wenzel, S. , P. Boyaci, and U. Jessen. 2010. "Simulation in Production and Logistics: Trends, Solutions and Applications." In *Advanced Manufacturing and Sustainable Logistics.* Proceedings of the 8th International Heinz Nixdorf Symposium, IHNS 2010, Paderborn, Edited by W. Dangelmaier, A. Blecken, R. Delius, and S. Klöpfer, 73-84.

Wenzel, S., M. Weiß, S. Collisi-Böhmer, H. Pitsch, O. Rose. 2008. *Qualitätskriterien für die Simulation in Produktion und Logistik - Planung und Durchführung von Simulationsstudien.* Berlin: Springer.

Williamson, J. 2010. "The philosophy of science and its relation to machine learning." In *Scientific Data Mining and Knowledge Discovery*, Edited by M. M. Gaber, 77-89. Berlin, Heidelberg: Springer.

AUTHOR BIOGRAPHIES

TJORBEN BOGON studied computer science at the University of Bremen (Diploma in 2007). In April 2007 he worked as software engineer in CoachAndWin GmbH on an agent-based soccer simulation game system. From November 2007 he started his PhD on distributing particle swarm optimization by agent technology at Goethe University Frankfurt. Since November 2010, he followed Ingo Timm to the chair for business informatics at the University of Trier and continues his PhD there. His email address is bogon@uni-trier.de.

ULRICH JESSEN studied computer science at the University of Dortmund (Diploma in 1990). From 1990 to 2006 he was working at Fraunhofer Institute for Material Flow and Logistics. At that time he managed several national and international research and industry projects whose aim the development and realization of simulation, planning, and assistance systems was. Since 2007 he has been working at University of Kassel as manager of research and consultancy projects as well as manager of the master degree program Industrial Production Management. His email address is jessen@uni-kassel.de.

ANDREAS D. LATTNER received the Diploma (2000) and the doctoral degree (2007) in computer science at the University of Bremen, Germany. From 2000-2007 he was working as research scientist at the Center for Computing Technologies (TZI) at the University of Bremen. He works now as a postdoctoral researcher at the chair for "Information Systems and Simulation" at the Goethe University Frankfurt. His research interests include knowledge discovery in simulation experiments, temporal pattern mining, and multi-agent systems. His email address is lattner@cs.uni-frankfurt.de.

DIMITRIOS PARASKEVOPOULOS received the diploma degree (2010) in computer science at the Goethe University of Frankfurt am Main, Germany. Since February 2011, he has been a PhD student at the chair for Information Systems and Simulation at the Goethe University Frankfurt. His research interests are Process Mining, Business Process Management and IT-Risk Management. His email address is paraskev@informatik.uni-frankfurt.de.

MARKUS SCHMITZ studied computer science at the University of Kassel (Diploma in 2009) after finishing his professional training as IT specialist in Bonn in 2002. He has been working at the University of Kassel, Department of Production Organization and Factory Planning, since 2009. His area of expertise includes software-based factory planning, modern 3D-techniques and the application of current consumer techniques in professional environments. His email address is m.schmitz@uni-kassel.de.

SVEN SPIECKERMANN, Ph.D., is Chief Executive Officer at SimPlan AG, Maintal, Germany, mainly working as a senior consultant and project manager in simulation projects for several industries. Since

1992, he has been participating in over 200 simulation projects and various joint research initiatives. Additionally, he has been giving lectures in simulation at the Technical University of Braunschweig since 1995 and at the Technical University of Darmstadt since 2008. He has published several papers on simulation, simulation-based optimization and related topics. His email contact is sven.spieckermann@simplan.de.

INGO J. TIMM held various positions in research at University of Bremen, Technical University Ilmenau, and Indiana University Purdue University – Indianapolis (IUPUI) from 1998 until 2006 before being appointed full professor for Information Systems and Simulation at Goethe-University Frankfurt/Main. Since 2010, he holds a chair for Business Information Systems at University of Trier. Ingo Timm is speaker of the German SIG on Distributed Artificial Intelligence (FG-VKI) of German Informatics Society. His special interest lies on business information systems, intelligent assistance systems, (actor-based/multiagent-based) simulation, and knowledge-based support to simulation system. His email address is ingo.timm@uni-trier.de.

SIGRID WENZEL is full professor and managing director of the Institute for Production Engineering and Logistics, University of Kassel, and head of the Department of Production Organization and Factory Planning. In addition to this, she is a board director of the Arbeitsgemeinschaft Simulation (ASIM), spokesperson for the ASIM working group Simulation in Production and Logistics, member of the advisory board of the Association of German Engineers Society of Production and Logistics (VDI-GPL), and head of the Committee Modeling and Simulation of the VDI-GPL. Her email address is s.wenzel@uni-kassel.de.