FROM TRACEFILE ANALYSIS TO UNDERSTANDING THE MESSAGE OF SIMULATION RESULTS

Gaby Neumann[1], Juri Tolujew[2]

[1]Technical University of Applied Sciences Wildau, Faculty of Engineering
Bahnhofstr., 15745 Wildau, Germany
[2]Fraunhofer Institute for Factory Operation and Automation (IFF) Magdeburg
Sandtorstr. 23, 39106 Magdeburg, Germany

gaby.neumann@tfh-wildau.de (Gaby Neumann)

Abstract

Recent approaches for automatic trace file analysis mainly focus on formalizing simulation outcome in the context of a certain application area. They have in common the very much reduced role they give to the persons who build or use the simulation model. Instead they assume any result derived from simulation can directly and automatically be extracted from the trace file without any additional intervention by the simulating person. Against this background the paper reviews related work for trace file analysis with regard to its motivation, approaches and state-of-the-art. This is put into relation to simulation user needs in the logistics application area in order to discus to what extent trace file analysis helps in deriving findings, which role the user plays in receiving those results, which kind of support is missing here and how it could be provided. Conclusions emphasize that it is necessary to see behind pure simulation data in order to understand the real message of simulation results. This interpretation step requires both knowledge and understanding of the domain and mathematical/statistics skills. Automatic trace file analysis supports preparation of interpretation steps but cannot fully replace the user who brings in objectives, motivation and focus of the simulation project as well as domain-specific experiences and competences to understand the message of simulation results.

Keywords: simulation output interpretation, trace file analysis, simulation knowledge, discrete event simulation, logistics.

Presenting Author's biography

Gaby Neumann is Professor in Engineering Logistics. Since 1991 she has been part-time consultant in logistics simulation, too. Her current activities and research interests are mainly linked to problem solving and knowledge management in logistics simulation and planning, but also cover technology-based logistics learning. She has widely published in these fields. Her e-mail address is gaby.neumann@tfh-wildau.de.

1 Introduction and motivation

One of today's challenges consists in seeing simulation in the context of human-centered processes. In recent literature this is being addressed by, for example, providing simulation modelers (or users) with methods and tools for automatic trace file analysis in order to better cope with large amounts of simulation output data. Those approaches mainly focus on formalizing simulation outcome in the context of a certain application area.

In [1], for example, authors state that in tracing a simulation model a modeler finds himself in the situation where it is unclear what properties to ask to be checked by a model checker or what hypothesis to test. They assume that cyclic behavior of model components is always good behavior whereas all exceptions or disturbances in this behavior indicate errors. Therefore, the aim is to provide support by automatically identifying and removing repetition from a simulation trace in order to pay particular attention on the non-returning, progressing part of a trace. This is to be achieved by automatic trace file reduction as it is assumed that modelers do not have enough background knowledge or experience to figure out interesting parts of the trace themselves.

Authors in [2] assume that simulation usually aims to specify whether or not the concept of a material flow system meets formal requirements, but not how well it does it. This is said to be caused in limited methodological support and therefore strongly depend on the modeling/planning expert's experience and expertise. This is to be overcome by eliminating the user as weakest point through automatic analysis. For this an analysis tool is proposed that helps in identifying the concept's or system's weak points, specifying their primary reasons and pointing out system immanent potential for performance increase.

Both approaches have in common the very much reduced role they give to the key actor(s) in any simulation project: the person who builds the simulation model and the person who uses the simulation model to run experiments. Instead they assume any result derived from simulation can directly and automatically be extracted from the trace file through statistical analysis, clustering or reasoning without any additional explanation by the simulating person. If this would be the case then any simulation model and any plan of experiments can be seen as objective representation of a particular part of reality and its problem situation. Any model building or experimentation activity no matter what background or intention one has would lead to the same model and to the same collection of simulation output. A particular simulation output always would lead to the same conclusions, i.e. simulation results, no matter what is being analyzed by whom and how.

If this would be the case, why do simulation projects still require human resources of certain expertise to be involved in? It is because simulation projects are not only sequences of formalizing steps that can be fully represented by more or less complex logical algorithms, but also require intuitive problem solving, combining analyzing steps and the need for creative thinking. Whereas the first can already be formalized or will be in future, the latter always remains linked to the person carrying out or contributing to or requesting simulation projects. Approaches to increase the degree of formalization in simulation, no matter if they focus on automatic model generation or automatic trace file analysis and simulation result delivery, will always be limited by the impossibility of fully formalizing the objectives and goals of a simulation. As already concluded in [3] the simulation user will therefore continue to be the key factor in any simulation project.

In order to better understand the kind of support that is needed and how it can be provided the following sections of the paper discus the role the user plays in simulation (Section 2), specify what is expected from simulation using logistics as exemplary application area (Section 3), and propose a methodology for supporting the user in specifying and achieving output data needed (Section 4). Section 5 summarizes research findings and presents conclusions derived from them.

2 Which role does the user play in simulation?

In general, simulation projects in the field of logistics – as in other fields too – are organized in the form of a service involving both, simulation experts and logistics experts with individual knowledge to be of use at certain stages of the project: Simulation experts are primarily responsible for model building and implementation steps, whereas logistics experts mainly provide application-specific knowledge for problem description, identification of input data and evaluation of results [4]. In order to better understand the role of the user in simulation it is worth to take a closer look at simulation knowledge sources and stakeholders for identifying which knowledge comes from where and in which form.

In general, input information for a *simulation project* usually come with the tender specification or are to be identified and generated in the problem definition and data collection phases of the simulation (see Fig. 1). Here, the user decide (and bring in) what is to be taken into consideration for model building and which information is required for the investigation.

The *model-building process* should be seen as another important phase of collecting, evaluating and structuring information. As discussed in [5] a simulation model is more than just a tool necessary to achieve certain objectives of experimentation and cognition. In the course of a simulation project the simulation model is developed, modified, used, evaluated and extended within an ongoing process. Therefore, it is also a kind of dynamic repository containing knowledge about parameters, causal relations and decision rules gathered through purposeful experiments. Even further, knowledge is "created" systematically through simulation based on the systematic design of experiments (including a meaningful definition of parameters and strategies) and the intelligent interpretation of results.

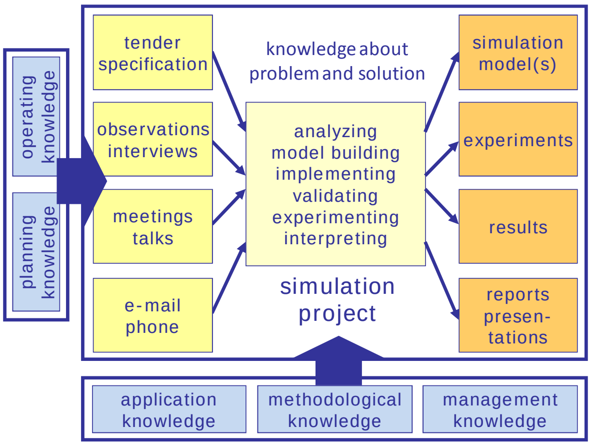

Fig. 1 Sources and evolution in simulation knowledge

Simulation experiments, for example, to support logistics planning and operation might be oriented towards modifications in either functionality or structure or parameters of a model and its components or even in a combination of those variations leading to more complex fields of experiments. Experimentation efforts are directly related to the type of variation required. The latter depends on the specific design of the simulation model resulting from the underlying modeling concept of the simulation tools and the design of the conceptual model by its developer. To correctly interpret simulation output it is necessary to understand what the objectives, parameters and procedures of a certain series of experiments were and to relate this to the results and findings. Consequently, the objectives of a simulation and the questions to be answered by experiments should already be taken into consideration when designing the conceptual model. Specific opportunities and features offered by the selected simulation tool then influence transformation of the conceptual model into the computer model when it comes to model implementation.

All steps again and again require input and background information based upon the knowledge and experience of the user, i.e. the simulation expert and the domain expert. In terms of simulation target definition it is particularly necessary to understand what the domain expert expects from simulation. As this is typically specific to the application area, we continue discussions using logistics as example.

3 What is expected from the use of simulation in logistics?

In the course of a logistics simulation project both partners, logistics expert (simulation customer) and simulation expert (simulation service provider), use to face the ever challenging task to interpret numerous and diverse data in a way being correct with respect to the underlying subject of the simulation study and directly meeting its context. These data are usually produced and more or less clearly presented by the simulation tool in the form of trace files, condensed statistics and performance measures derived from them, graphical representations or animation. Problems mainly consist in

1) clearly specifying questions the simulation customer needs to get answered,

2) purposefully choosing measures and selecting data enabling the simulation service provider to reply to the customer's questions, or

3) processing and interpreting data and measures according to the application area and simulation problem.

To overcome these problems and give support in defining simulation goals and understanding simulation results, methods and tools are required that are easy to use and able to mediate between knowledge and understanding of the simulation customer (the logistics expert planning or operating that process and system to be simulated) and the simulation expert (the expert from the point of view of data and their representation inside computers). Within this context, it is worth thinking in more detail about what a simulation customer (the logistics expert) might look for when analyzing the outcome of simulation experiments [6]:

- *Typical events*. The logistics expert specifically looks for moments at which a defined situation occurs. This kind of query can be related, for example, to the point in time at which the first or last or a specific object enters or leaves the system as a whole or an element in particular. Other enquiry might be oriented towards identifying the moment when a particular state or combination of states is reached or conditions change as defined.

- *Typical phases*. The logistics expert is especially interested in periods characterized by a particular situation. In this case s/he asks for the duration of the warm-up period, for the period of time the system, an element or object is in a particular state, or how long a change of state takes.

- *Statements*. The logistics expert looks for the global characteristics of processes, system dynamics or object flows such as process type (e.g. steady-state, seasonal changes, terminating/

non-terminating), performance parameters of resources (e.g. throughput, utilization, availability), parameters of object flows (e.g. mix of sorts, inter-arrival times, processing times). This information is usually based on statistics resulting from trace file analysis and replies to either a specific or more general enquiry by the user.

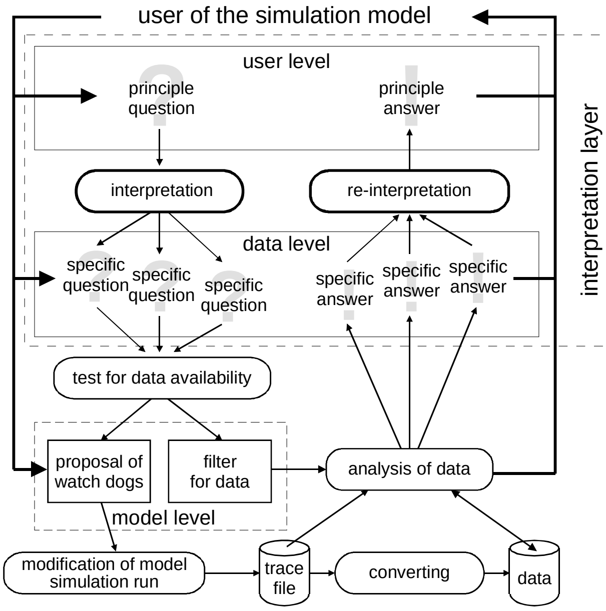

Fig. 2 User-data interaction for simulation output analysis

When the potential interests of a simulation customer are compared, one significant difference emerges: whereas the first two aspects need specific questions formulated by the logistics expert directly at data level, the last aspect is characterized by usually fuzzy questions of principle from the more global user's point of view. Before these questions of principle can be answered, they have to be transferred to the data level by explaining them in detail and putting them in terms of concrete data (see Fig. 2). As result of this process of interpretation a set of specific questions is defined with each of them providing a specific part of the overall answer in which the user is interested. Questions at data level correspond to results that can be delivered directly by the simulation even if minor modifications to the simulation model should be required [7]. This is the kind of study also current approaches for trace file analysis support [1], [2]. To derive an answer in principle to a question of principle the respective set of specific answers needs to be processed further. These steps of additional analysis and condensing can be understood as a process of re-interpretation to transfer results from data to user level.

All steps of interpretation and re-interpretation aim to link the user's (logistics expert's) point of view to that of the simulation expert. They not only require an appropriate procedure, but, even more importantly, an interpretative model representing the application area in which simulation takes place. This model needs to be based on knowledge and rules expressed in the user's individual expertise, but also in generalized knowledge of the (logistics) organization regarding design constraints or system behavior and the experience of the simulation expert derived from prior simulations. As this knowledge might not only be of explicit nature, i.e. existing independent of a person and suitable to be articulated, codified, stored, and accessed by other persons, but also comprises implicit or tacit knowledge carried by a person in his or her mind often not being aware of it simulation users as individuals or team need to remain involved in the steps of interpretation and re-interpretation at least. Whereas explicit knowledge might be transferred into rules and algorithms, tacit knowledge cannot be separated from its owner and therefore requires direct involvement of the knowledge holder in the interpretation process. More specifically this means support is required for translating any principle question into corresponding specific (data-related) questions as well as for deriving principle answers from a number of specific (data-related) answers. Although a set of (standard) translation rules might be known, formalized and put into the rule base already, always further questions remain that are unknown to the rule base yet. Here, the logistics expert needs support in (i) correctly formulating the right question and (ii) getting the full picture from the puzzle of available data and their analysis.

One approach for enabling this could be based on viewpoint descriptions. Viewpoint descriptions were introduced into model validation as a new kind of communication and interaction between the human observer of simulation results and the computer as the simulation model using authority that was called oracle-based model modification [3]. Here, the principle idea is that the user presents his or her observations (in the animation) as a viewpoint description to the computer that initiates a reasoning process. This results in definition and realization of necessary changes to the simulation model in an ongoing user-computer dialogue. The main advantage of this concept lies in the reduced requirements for rule-base definition. Those aspects that easily can be formalized (e.g. typical quantitative observations or unambiguous logical dependencies) are translated into questions to the user (What is it s/he is interested in?) or various forms of result presentation (as figures or diagrams), whereas those that are non-imaginable yet or individual to the user or simply hard to formalize do not have to be included to provide meaningful support to the user. There is no need to completely specify all possible situations, views and problems in advance, because the person who deals with simulation output brings in additional knowledge, experience and creativity for coping with non-standard challenges. Even further, this way the rule-base continuously grows as it "learns" from all applications and especially from those that were not involved yet.

On the other side the user benefits from prior experience and knowledge represented in the computer by receiving hints on what to look at based upon questions other users had asked or which were of interest in earlier investigations.

This approach helps in designing the interpretation layer for mediating between simulation customer and simulation model or output no matter how many data have been gathered and how big the trace file grew. Nevertheless, effectiveness and efficiency of this interpretation process depends on the availability of the right data at the right level of detail. This quite often does not only depend on the simulation model and tool used for its implementation, but also on the opportunity to aggregate data in always new ways.

4 How specific results can be achieved from DES experiments?

As discussed in the previous section simulation results derived from running experiments by use of a particular simulation model are as good as they finally respond to the questions the simulation user is interested in. The challenge consists in knowing about questions a user in a specific project might have. Generally, a certain amount of (standard) questions can be pre-defined in correspondence with the application area and another set of questions might be defined by the user when starting into simulation modeling and experimentation. This might even lead to a specific focus in trace file generation and recording of simulation output data by purposefully introducing a cohort of observers to the model that directly correspond to the type and amount of data required for responding to questions already addressed by the user [7].

However, it is not that exceptional that new questions arise in the cause of the simulation project when seeing results from previous experiments. In those situations it might either be necessary to re-run simulation with a modified observation concept or to aggregate or derive results from already existing simulation output in a different way. Concerning the first, there are two options for interpreting simulation output: online and offline. Online interpretation might focus on:

- visualizing changes in the position of moving objects;
- visualizing states (e.g. stock development);
- identifying or recognizing pre-defined situations.

Offline interpretation typically is used for:

- calculating freely definable characteristics;
- identifying or recognizing pre-defined situations;
- preparing and showing special-focused animations.

Although being specific to a certain simulation project, those analysis steps are possible to be pre-specified and also in the focus of approaches as presented in [1] and [2]. But beyond this, specific questions relevant in a certain simulation project might eventually even require to summarize (primary) objects as simulated into new (secondary) classes not simulated yet. In a transportation model with a number of trucks moving different types and different volumes of goods, for example, it suddenly might be of interest to know something about all those trucks arriving Tuesdays only. The simulation model itself knows trucks as one class of objects, but does not contain “Tuesday trucks” as a specific sub-class to this. This new class needs to be formed out of the situation and might then be added to the rule-base for trace file analysis, but cannot be pre-defined as simply not specified before. Consequently, any tool to support trace file analysis must allow and even support those interactions with the trace file which again goes far beyond formal statistical analysis.

5 Conclusions

To understand the message of simulation results formal trace file analysis is one important step. The other one is the non-formal, more creative step of directly answering all questions that are of interest to the user (in our case the logistics expert). Precondition is to know (and understand) what the questions of the user are, but also the ability of the user to ask questions relevant to a particular problem. For the latter, the framework for trace file analysis and interpretation provides even further support: Typical questions no matter if they are of generic or specific nature help the user in identifying the problem or the questions to be asked or the aspects to be investigated. As discussed, this can be supported by the approaches for viewpoint description and defining observers or specifying analysis focus. Additionally, a pattern combining typical symptoms (i.e. visible situations or measurable characteristics) with the underlying problems causing those symptoms would be of huge benefit as this might also guide the user in truly understanding what happens in a specific material handling or logistics system.

Current approaches to trace file analysis mainly focus on deriving (standard) parameters and (typical) characteristics by use of statistical methods, clustering or reasoning. With this they provide results at data level (see Fig. 2) allowing basic interpretation based upon (externalized) domain-specific knowledge. This step works automatically for those aspects that can be formalized and shows limited results only for those aspects that require intuitive, creative thinking by the user.

Against this background the paper concludes that it is necessary to see behind the simulation results by interpreting simulation output in order to understand the real message of simulation results. This interpretation requires knowledge and understanding of the domain / application area as well as mathematical and statistics skills. Trace file analysis supports preparation of interpretation steps but cannot fully replace the user who brings in objectives, motivation and focus of the simulation project as well as domain-specific experiences and competences to understand the message of simulation results. A sophisticated framework especially supports in reducing routine work like statistics calculations through incorporated powerful analysis tools and stimulating creative thinking by proposing, asking, suggesting in a really interactive communication between the simulation user and the computer.

6 References

[1] P. Kemper and C. Tepper. Automated trace analysis of discrete-event system models. *IEEE Transactions on Software Engineering*, vol. 35, no. 2, March/April 2009, pp. 195-208.

[2] D. Wustmann, V. Vasyutynskyy, T. Schmidt. Ansätze zur automatischen Analyse und Diagnose von komplexen Materialflusssystemen. In W. Scheid (editor), *Proc. of the 5th Expert colloquium of WGTL - Wissenschaftliche Gesellschaft für Technische Logistik*, Ilmenau (Germany), 2009, pp. 1-19.

[3] C. Helms and T. Strothotte. Oracles and Viewpoint Descriptions for Object Flow Investigation. *Proc. of the 1st EUROSIM Congress on Modelling and Simulation*, Capri (Italy), 1992, pp. 47-53.

[4] G. Neumann and D. Ziems. Logistics Simulation: Methodology for Problem Solving and Knowledge Acquisition. In: N. Callaos; M. Bica; M. Sanchez (editors), *Proc. of the 6th Multiconference on Systemics, Cybernetics and Informatics*, vol. XII, Orlando (Florida/USA): International Institute of Informatics and Systemics, 2002, pp. 357-362.

[5] G. Neumann. The role of knowledge troughout the simulation lifecycle: what does a simulation model know? In B. Zupančič, R. Karba, S. Plažič, editors, *Proc. of the 6th EUROSIM Congress on Modelling and Simulation*, vol. 2; Ljubljana (Slovenia), 2007, paper TH-1-P12-5.

[6] G. Neumann. Simulation and Logistics. In B. Page and W. Kreutzer, editors, *The Java Simulation Handbook – Simulating Discrete Event Systems in UML and Java*, Shaker, Aachen, 2005, pp. 435-468.

[7] J. Tolujew. Werkzeuge des Simulationsexperten von morgen. *Proc. of Simulation and Animation '97*, SCS - Society for Computer Simulation International, San Diego, 1997, pp. 201-210. (Tools of the simulation expert of tomorrow, in German)

[8] J. Tolujew, T. Reggelin, C. Sermpetzoglou. Simulation und Interpretation von Datenströmen in logistischen Echtzeitsystemen. In: C. Engelhardt-Nowitzki; O. Nowitzki; B. Krenn (editors): *Management komplexer Materialflüsse mittels Simulation. State-of-the-Art und innovative Konzepte*. Leobener Logistik Cases, Wiesbaden: Deutscher Universitäts-Verlag 2007, S. 215-232. (Simulation and interpretation of data streams in logistics real-time systems, in German)