AN AUTOMATED APPROACH TO INPUT DATA MANAGEMENT IN DISCRETE EVENT SIMULATION PROJECTS: A PROOF-OF-CONCEPT DEMONSTRATOR

Anders Skoogh[1], Jean-Patrick André[1], Catarina Dudas[2], Jacob Svensson[3], Matias Urenda Moris[3], Björn Johansson[1]

[1]Chalmers University of Technology, Dept. of Product and Production Development, Gothenburg, Sweden
[2]Volvo Technology Corporation, Gothenburg, Sweden
[3]University of Skövde, Centre for Intelligent Automation, Skövde, Sweden

jandre@ifma.fr (Jean-Patrick André)

Abstract

Despite the fact that Discrete Event Simulation (DES) is claimed to be one of the most potent tools for analysis and optimization of production systems, industries worldwide have not been able to fully utilize its potential. One reason is argued to be that DES projects are not time efficient enough due to extensive time consumption during the input data phases. In some companies, input data is totally missing, but even in projects where data is available it usually takes a considerable amount of time to analyze and prepare it for use in a simulation model. This paper presents one approach to the problem by implementing a software that automates several steps in the input data process such as extracting data from a database, sorting out the information needed and fitting the data to statistical distributions. The approach and the software have been developed based on a case study at Volvo Trucks in Gothenburg, Sweden. The work presented in this paper is part of a more comprehensive project called FACTS. The project scope is to develop methods and IT-tools for conceptual plant development.

Keywords: Discrete Event Simulation, Input data, Data management, Distribution fitting, Conceptual Factory Development.

Presenting Author's biography

Jean-Patrick André was born in Grenoble, France, 1984. He is studying at the IFMA (French Institute for Advanced Mechanics), France, at the Advanced Manufacturing Systems Department with the aim of obtaining a diploma of engineering in 2008. He is currently working as a guest research student in the field of DES at the Department of Product and Production Development, Chalmers University of Technology, Sweden. His email is <Jean-Patrick.Andre@ifma.fr>.

1 Introduction

Discrete Event Simulation (DES) continuously gains popularity as a tool for optimizing efficiency in production. The ability to include dynamic aspects of production systems in the analyses contributes to the popularity that ranks DES among the top three tools for decision support in production design projects [1,2]. At present, the majority of DES-projects are performed to improve already existing manufacturing systems [1]. However, to be able to fully take advantage of simulation, organizations should frontload the analyses in early concept phases of production design projects [3,4]. The major reason is that decisions taken in these project phases are important, since their impact on overall economic aspects of the project is substantial.

Regardless of project purpose, the possibility to find high-quality input data is one of the most important factors in order to succeed. Unfortunately, this process has many times proved to be difficult and time-consuming due to insufficient data availability or problems with extracting available data from a wide range of sources [5,6]. In fact, the input data phase usually occupies around one third of the total DES project time [5,7,8]. This is a major setback for simulation, especially in early concept phases of production design projects where rapid decision-making is often required. To increase the applicability for DES in running production, as well as in early project phases, a new approach needs to be taken and adopted by industry [9]. Rapidity in simulation projects is one key factor, which in many situations can be achieved by more efficient input data management.

Fortunately, some research has been performed to cope with the difficulties in input data management. As discussed in the following section, much of the previous work has focused on solutions including data that has been automatically collected and stored in Enterprise Resource Planning (ERP) systems or Corporate Business Systems (CBS). These approaches are definitely potent for the future or even today, for large companies with extensive experience of DES. However, many organizations do not have these possibilities due to absence of relevant production data in their ERP or CBS systems [10,11]. It is more common, at least in Swedish industry, to find production data in other sources such as in-house built logging systems (automatic, semi-automatic or manual).

This paper aims to present an automated approach to preparing data from such a logging system, for use in a DES model. A software is developed to take care of the time-consuming activities of extracting data from an automatic data collection system and to perform a statistical analysis, in order to suggest a statistical representation for use in the simulation model.

2 Input Data Management in DES projects

The input data process in DES projects consists of several activities, which all play important roles to transform shop-floor behavior into relevant input data for simulation models. Fig. 1 shows an outline of input data management activities, normally performed in simulation project and also in the case study presented in this paper. The model is combined, based on the authors' experience and earlier research findings [5,6,12].

Firstly, raw data is collected using automatic collection (PLC-logging etc.) and/or manual gathering (clocking, interviews, etc.). The data is then stored, usually in some kind of computer application e.g. personal spreadsheet or ERP-system, before it is extracted in order to be analyzed and prepared for use in simulation models. Furthermore, the prepared data is put into the simulation model, either manually or automatically. At the final residence, the data can be read by the model and stored as project documentation or for re-use in future simulation studies.

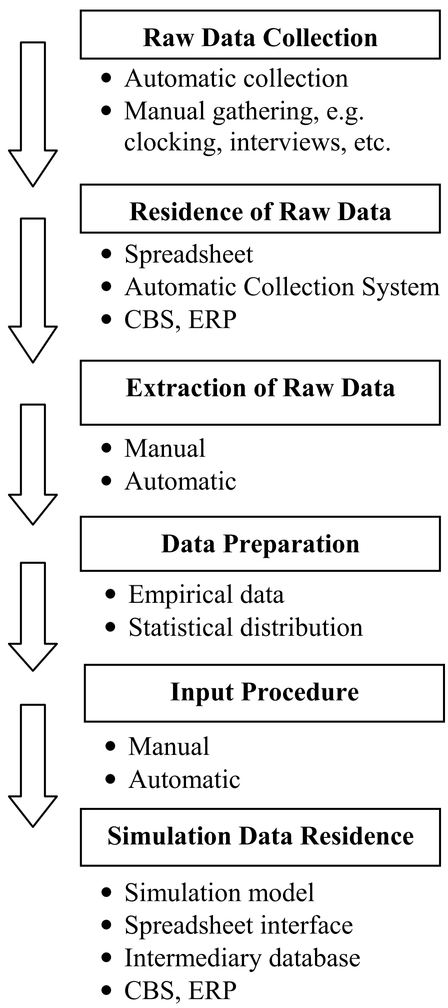

Fig. 1: Input data management activities.

Today, four major approaches are used among practitioners in manufacturing industry, to perform the activities presented in Fig. 1 [12]:

1. Tailor-made solution

- Data primarily derived from the project team
- Data manually input to model by model builder
- Data resides in the simulation tool

2. Spreadsheet solution

- Data primarily derived from project team
- Data manually input to computer application, usually MS Excel spreadsheet.
- Data automatically read by model via computer application
- Data resides in the computer application

3. Off-line database solution

- Data primarily derived from Corporate Business System
- Data automatically read by model from an intermediary simulation database, e.g. MS Access.
- Data automatically input to database from Corporate Business System
- Data resides in the intermediary simulation database

4. On-line database solution

- Data primarily derived from Corporate Business System
- Data automatically read by model
- Data automatically input to model from Corporate Business System
- Data resides in the Corporate Business System

The tailor-made solution was formerly a popular methodology, but due to the extensive amount of manual work, it has obvious drawbacks such as heavy time-consumption and difficulties to change or modify data [12]. Hence, methodologies requiring less human involvement are becoming more and more interesting for industry. The same study as referred to above, found that the spreadsheet solution is becoming increasingly used and is probably the most common solution among manufacturing companies today. However, despite this progress, input data management is still a heavily time-consuming phase in DES projects and the major reasons are the efforts related to manual gathering, extraction from available data sources and preparation for use in the simulation models [5,6].

To address the problems related to collection of raw data, the demand for various types of automatic collection systems is increasing in industry, and they will probably be a natural data source in future production systems [9,13]. Additional benefits with automatic data collection systems are objectiveness and accuracy of collected data [9]. Referring to the increased data availability, researchers argue that more and more information required for simulation will be automatically collected and obtainable via ERP systems. And since the number of ERP systems implemented in companies worldwide increases [11], they should be valuable data sources for DES projects in the future. Hence, a significant amount of research work has been performed to show the potential and prove the feasibility of connecting CBS and ERP systems to DES models, as in approaches 3 and 4 presented earlier [12,14].

The authors of this paper believe in the concept but argue that the possibilities to implement it in reality are limited at present. Organizations trying to automatically connect their ERP systems to DES models need to have all relevant data for simulation in their ERP systems, and this is only reality in a few major companies at the moment. Especially data describing the dynamics and stochastics of the production system is often missing in the ERP-systems [3]. Another important issue to keep in mind when talking about a totally automated input data procedure (e.g. approach 4), is that practitioners usually find it convenient and secure to be able to see, complete, change and experiment with the data that will be input to their simulation model. In a totally automated input data procedure, these options would be problematic to replace.

To address the two problems described above, a slightly modified approach was evaluated in the case study presented in this paper (Fig. 2). To relate to earlier research it can be seen as a combination between approaches 2 and 3, and should be a suitable solution at least until all simulation data can be integrated in the same application, e.g. ERP-system.

The authors created a “middleware” that is able to extract and compile raw data from several different data sources. The results are presented in an intermediary computer application (MS Excel spreadsheet in this case study), where the user can approve, add, change and experiment with the data before it is input to the simulation model. Above the ability to extract raw data, the middleware also prepares the raw data for use in the simulation model, for example by fitting it to statistical distributions.

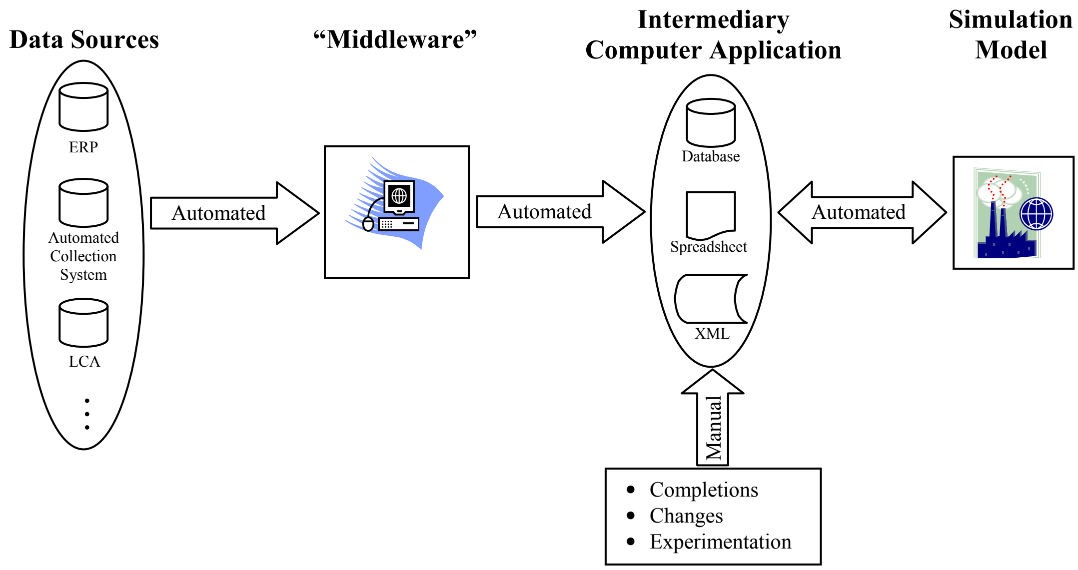

Fig. 2: The approach presented in this paper is to use a middleware solution as a bridge in between the data sources and the intermediary computer application.

The ERP may be one source of input, and can be complemented by breakdown information from an automated logging system and environmental data from a Life Cycle Assessment (LCA) database, just to give an example. The type of intermediary computer application can be a spreadsheet, database or an XML solution depending on organizational prerequisites.

3 Conceptual Plant Development – FACTS

The work presented in this paper is part of a more comprehensive project with a broader scope. The project name is "Factory Analyses in ConcepTual phases using Simulation" (FACTS) [15]. The aim of FACTS is to develop and implement methods and IT-tools for conceptual plant development. If DES is introduced in the early stages of production system design, before crucial decisions already have been made, it gives the project team the opportunity to see the impacts of product mix changes or new production setups. Having this insight offers the possibility to make better decisions and evaluate better system design configurations.

It has been argued though, that lack of data and long model development time makes DES inappropriate as a technique to use in the early phases of a system design project [16]. The FACTS initiative addresses both these statements, proposing an innovative framework based on aggregation techniques, automated input data handling/manipulation and the combination of optimization and output data analysis. The project realization is therefore organized into five main areas, related to the project objective but with separate focus (see Fig. 3).

- Abstraction/aggregation methods
- Input data handling
- Optimization
- Integration (integrates results from the earlier three to developed a support tool and work procedures)
- Information dissemination

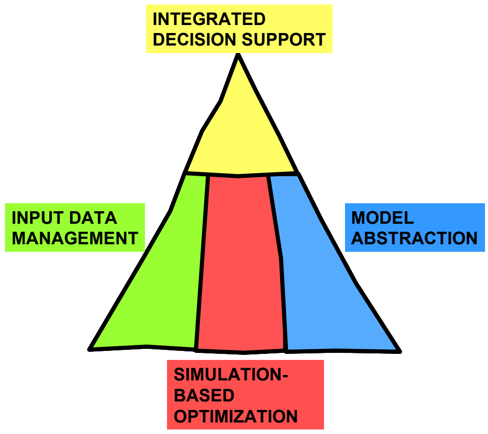

Fig. 3: Schematic description of the project's work packages. The project delivers methods and tools to support concept analyses (the top in the triangle). This support is based on results and data from the three work packages/areas: (1) Simulation-based optimization, which needs (2) abstract models, and is fed with (3) input data. [15]

The first three areas are the main parts of the work procedure. The project gathers empirical data from case studies, which are carried out and evaluated successively during the project. A toolset based on the integration of the sub-projects results, is already partly developed and presented in [17] under the name of FACTS analyzer.

The project is funded by VINNOVA, Sweden (within the MERA framework) and the industrial partner companies, including Volvo Car Corporation, Volvo Powertrain, Volvo Technology and Volvo Trucks. Academical partners are University of Skövde, Chalmers University of Technology (Gothenburg) and IVF Industrial Research and Development Corp (Mölndal).

4 Case Study Description

In 2006 Volvo Truck Corporation produced more than 100 000 trucks and is therefore the second largest producer of heavy trucks worldwide, creating reliable transport solutions for customers all over the world. Roughly one fifth of these are produced at the Tuve plant in Gothenburg, Sweden. Staying competitive requires efficient virtual manufacturing tools.

Virtual manufacturing tools are currently used in late phases at Volvo Trucks, mainly to develop existing production facilities. Volvo Trucks is in urgent need of a decision tool in earlier phases of the development process. The decision-making in early phases demands information quickly and this implies a more efficient input data handling.

4.1 Purpose and Objectives

The objective is to initiate the use of IT-tools and methods for decision-making earlier in the process, forcing the organization to state what kind of input data and information is needed. The result of this procedure will imply that the production development can state their demands on product development. In consequence, this will be an opportunity for the departments of product and production development to collaborate and communicate earlier, avoiding time-consuming and expensive misunderstandings.

In order to develop an efficient production line, analyzing and optimizing it by simulation before the line exists or has started production is necessary. Fig. 4 shows how this is achieved by focusing on simulation early, resulting in less testing at the end of the project. This approach is often referred to as frontloading.

With a consistent tool and an efficient input data handling, Volvo Trucks will use this method for testing possible to-be scenarios. Since different variants are assembled on the same line, the process times at each station may differ. This entails a great interest in testing different scenarios for the takt time, mainly on how it should be determined. Takt time is here used to describe a pre-determined time that controls the pacing rate of parts going from station to station in a synchronized manner.

One common opinion for not using DES in early phases is that appropriate input data is unavailable, which was pointed out in section 3. The input data might be insufficient, and as a consequence simplification based on experience has to be done. The input data handling is approximately 30 % of a simulation study [5], and therefore one of the most time-consuming parts. Making this process more efficient would gain a considerable amount of time and will therefore be well applicable in early phases when the time horizon is constricted.

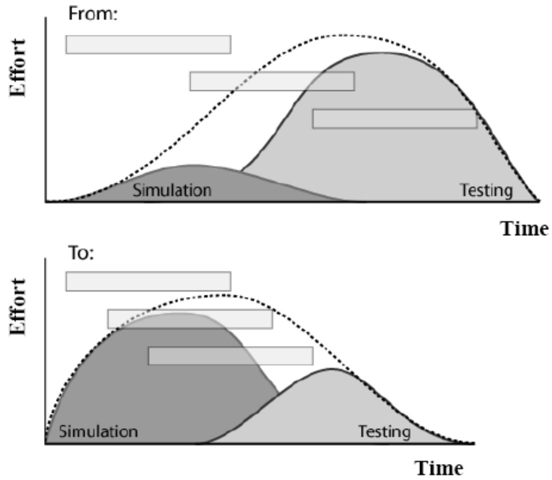

Fig. 4: Increased use of simulation in early phases will reduce the testing in late phases [3].

4.2 The Production System

The ambition is to handle all pre-assembly on sub-lines in order to eliminate disturbance on the main line. Pre-assembly is completed at several sub-lines preceding the main line at the Tuve Plant. The engine sub-line is in focus for this study.

Volvo Powertrain is the supplier of engines from Skövde, Sweden, and transmissions from Köping, Sweden. The engine and the transmission are assembled together with the generator and the compressor into a complete engine.

The engine line consists of two parallel lines, Engine line 1 and Engine line 2, with nine stations on each line, Fig. 5. The engine is put on an AGV (Automated Guided Vehicle) at station 1 from a roller conveyer track. Work is performed at the next eight stations and after station 9 there are three buffer places for each line. Thereafter, the finished pre-assemblies are transported to the main line by forklifts.

4.3 The Simulation Model

The simulation model of the engine lines at the Tuve plant concentrates on the nine stations (x2), the AGV system being their material handling system, and the two takt systems controlling the pace of the two lines [17].

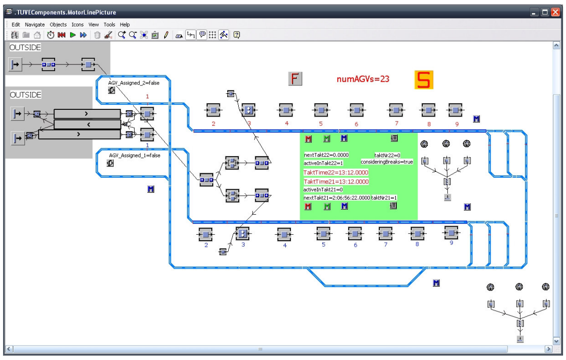

Fig. 5: The model of the two engine lines and the object controlling the takt logic in the middle.

If one or several stations on the engine lines exceed their corresponding takt time, the exceeding time on each station is logged as a failure. The way in which the failure data is logged on the engine lines and the fact that both lines are controlled by a takt system with a corresponding takt time enables a simplified modeling of the two lines. In this simplified model, concern about different engine variants can be completely neglected. The processing time on each station is simply set to the takt time, and aside from some transport times in the model, the real input needed in the model is the failure data. The drawback of this way of modeling is that the variant sequence is lost. However, the variant sequence is not of interest in this case. The simulation model is created using UGS Plant Simulation.

4.4 Process Information System - PI

Failure data in the Tuve plant is automatically collected in a system called the Process Information System (PI). In this case, some of the more vital data logged in the PI-database is what station that caused the failure, the time when the failure occurred, and the duration of the failure. However, this type of raw data is of no use in a simulation software before it has been processed and adapted to the software. The failure data needs to be processed into two statistical distributions, one modeling the time to repair (TTR) and one modeling the time between failures (TBF).

4.5 The Manual Input Data Process

Both engine lines are manual lines; the failure data was therefore not expected to contain many long failures. However, this was not the case. The failure data showed a lot of long failures. The long failures turned out to be influenced by breaks, nights, weekends, and in some cases holidays. Because of this the failure data was truncated with an upper bound of one hour, chosen after careful studies of the production system. The failure data processing did not stop there. It provided useful observations of the failure duration, i.e. the TTR, but observations of the TBF still needed to be calculated. The TBF observations were calculated with the help of an Excel macro. The observations of TTR and TBF were then, in turn and per station, transferred into a statistical distribution fitting software that produced the corresponding distribution parameters. Finally these distribution parameters were compiled in a table that could be interpreted by the simulation model.

The amount of manual work needed to process and adapt the failure data into such a parameter table was, as one easily can imagine, rather large. This is the case in almost all failure modeling, since modeled systems commonly have several objects that require failure modeling.

5 Software Description

The aim of this research is to implement a software that automates several steps in the input data process, from the extraction of empirical data to the delivery of fitted distributions that can be used directly as input to the simulation model.

Today, the software extracts TTR data from an off-line copy of the PI-database, in order to prove the concept. However, it can be extended to include more parameters, which also will happen within the frame of the FACTS project. After extraction, a sorting procedure is applied to group the data according to which line number, station number and data type the information belongs to. A statistical analysis including a summary of empirical parameters, estimation of input parameters for statistical distributions and a distribution fitting is then performed. The final results of the software are the most appropriate statistical distributions for each of the stations at the two engine lines. The input suggestions are automatically presented in the MS Excel spreadsheet interface used in the case study, and the user is able to approve the results and use them in the simulation model. The input of model data from the spreadsheet was previously automated in the case study.

The described software functionality is visualized in Fig. 6, and more detailed descriptions of the sequential steps are presented in the following sub-sections. The software is developed in C++.

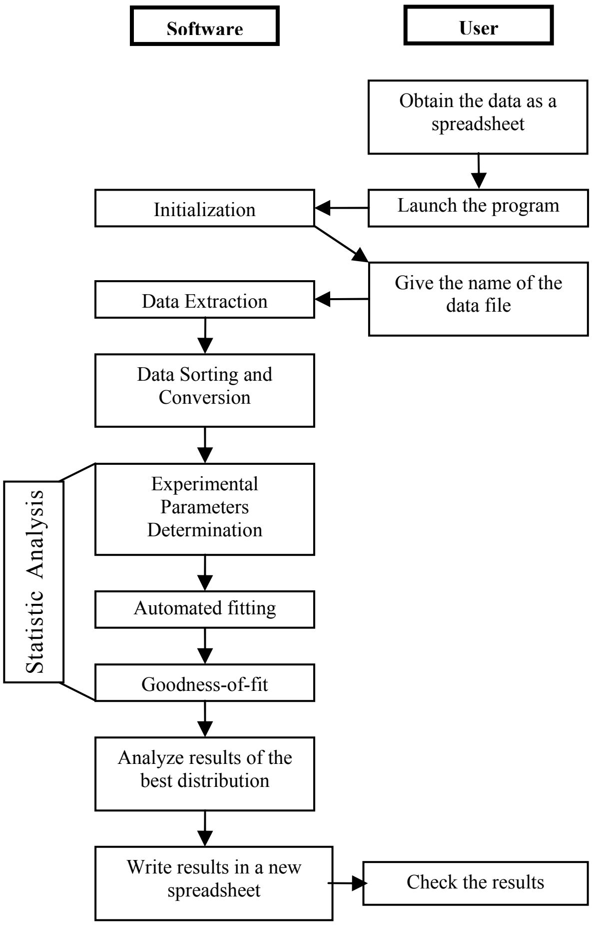

Fig. 6: Software functionality and user involvement.

5.1 User Interface and Involvement

One essential improvement with this solution is the significant reduction of user involvement. Traditionally in DES projects, the user has been involved in several handling steps required to finally obtain an input data representation for the simulation model. These steps typically consist of manual data gathering or extraction from data sources, followed by efforts to estimate parameters to statistical distributions. The actual choice of representation is often performed by graphical comparisons or complicated mathematical calculations. Even when using a distribution fitting software, some manual steps usually have to be carried out.

In the approach presented in this case study, the user involvement is reduced to the following activities:

- Create a copy of the on-line database (e.g. MS Excel spreadsheet, .txt-file or .csv-file)
- Run the program
- Give the name of the file including all raw data to the software (Fig. 7)
- Check the results

This has led to the possibility to reduce the time required to prepare input data to the simulation model.

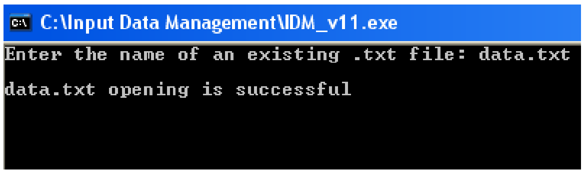

Fig. 7: User states the name of the file containing the raw data

The final intention for the software is to connect it directly to the PI-database for an even more automated procedure. However, at this stage a copy of the real-world database is created in order to ensure that the IT-support for running production is not interfered. An already existing web-interface is used to state what raw data to collect from the PI-database. Subsequently, the user just copies the entire content from the web-interface into the chosen off-line format (in this case a MS Excel spreadsheet, later saved as a .txt-file).

1	2	3	4	5	6
Input Data Analysis					
	Mean Time To Repair (MTTR)				
Engine Line 21					
	Distribution	Location	Parameter 1	Parameter 2	Parameter 3
Station 1	Exponential	0	442.09		
Station 2	Lognormal	0	377.42	2072.26	
Station 3	Exponential	0	540.20		
Station 4	Exponential	0	88.82		
Station 5	Lognormal	0	230.98	1095.13	
Station 6	Exponential	0	112.25		
Station 7	Lognormal	0	300.46	1908.91	
Station 8	Exponential	0	193.72		
Station 9	Lognormal	0	492.23	2557.50	

Fig. 8: Example of final results

The whole process, transforming seven months of raw data that corresponds to 18000 lines in the database into a suitable representation for the simulation model, required less than two minutes of time. The manual approach for the exact same work and for the same parameter took six hours. In fact, during these two minutes, most of the time consumption is due to the user handling. One example of the final results presented in the simulation model input data interface can be seen in Fig. 8.

5.2 Data Extraction

Since the shop-floor data in this case study was presented in an off-line copy of the PI-database, the extraction of data is consequently performed from the .txt-file representation (the user can also select other file formats such as .csv or .xls).

The implemented solution for the extraction activity is divided into several steps. The first one is to read data from the database copy (.txt-file) and sort it according to which line number, station number and data category the observation belongs to. Based on the sorting procedure, all the data is grouped and automatically put into several other .txt-files, only containing observations of the same origin with regard to the three sorting criteria stated above. From the second level of text files, the data is converted into the analysis unit of interest, which is seconds in the Volvo Trucks case study. The converted data is then stored in a chained list within the software and arranged starting with the lowest value and ending with the highest. Before the data is put into the software the user is also allowed to implement specific constraints, for example to only include observations within the interval ranging from 1 to 3600 seconds. The solution using a chained list reduces the complexity of the algorithm and increases the speed of the process. It also provides a flexible space to store data from database readings of different dates and numbers of breakdowns.

5.3 Parameter Estimation

There are basically two approaches available to determine the input parameters used to represent a dataset using statistical distributions. One is to estimate the parameters, based on calculations using statistics from the empiric data set, e.g. mean and variance. The other is to optimize the input parameters using a Maximum Likelihood Estimation (MLE) algorithm or an iterative algorithm that detects the best settings according to the results of a goodness-of-fit test. Today the software is programmed to use the first approach, since it gives results that are good enough compared to the alternative manual data handling procedure. However, to get more precise data representation the second approach would be preferable, which is further described in the section called “future improvements”.

Thus, to be able to estimate the input parameters, the software calculates a set of summary statistics from the raw data:

- Number of observations
- Minimum and maximum
- Mean
- Median
- Variance
- Coefficient of variation
- Skewness

First of all, it gives the user a good summary of the data-set he or she has to deal with. Secondly, some of these parameters are used to estimate the input parameters for the various statistical distributions used in the following goodness-of-fit tests. For example, the gamma cumulative distribution function needs two parameters (α and β), which are determined according to Eq. 1 and 2 [18]:

$$\alpha = \frac{Mean^2}{Variance} \quad (1)$$

$$\beta = \frac{Variance}{Mean} \quad (2)$$

5.4 Distribution Fitting

Based on experiences from previous data analyses from the engine lines at the Volvo Tuve plant, some distribution types were commonly applicable to represent the station behaviors. Hence, four of them were implemented in the software as a first step; the normal, lognormal, exponential and gamma distributions.

To find the best representation for all empiric data sets, input parameters for all four of the distribution types were estimated as described earlier. Thereafter, the software compares all possible representations and states the best representation for all TTR values at all stations using a Kolmogorov-Smirnov goodness-of-fit test (K-S test). However, normally the K-S test is used to decide whether a hypothesized distribution is a good representation of an empiric data-set or not, given a specified level of significance [19]. In this case it is only interesting to compare the different distributions to each other in order to find the best representation. This is performed by calculating the Kolmogorov-Smirnov test statistic (Eq. 3) [19] for all possible representation and choose the distribution with the lowest value of the K-S test statistics.

$$D = \max_{1 \le i \le N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right) \quad (3)$$

N = number of empirical observations
F = the cumulative distribution function (CDF)

The CDF completely describes the probability distribution of a real-valued random variable *X*. For every real number *x*, the CDF of *X* is given by Eq. 4 [18]:

$$F_X(x) = \mathrm{P}(X \le x) \quad (4)$$

For example the Normal CDF is defined by (Eq. 5) [20]:

$$F_{NX}(x) = \frac{1}{2}\left(1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right) \quad (5)$$

where σ is the standard deviation and μ is the mean. Erf (x) is the Gauss error function.

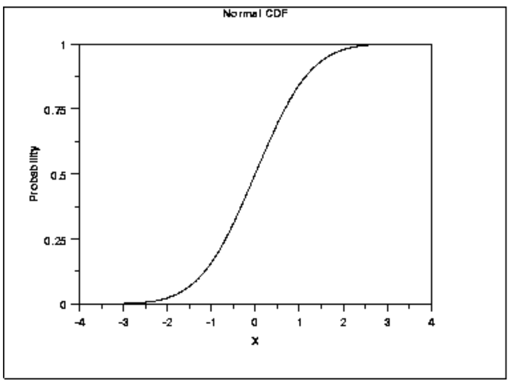

Fig. 9: Normal CDF [19]

The K-S test statistics has some advantages compared to other goodness-of-fit tests for use in this case study. For example, the Chi-Square test has a drawback in the difficulty to group the empirical data and decide interval widths for the analyses [21]. The Anderson-

Darling test would be another test to implement but since it is not applicable for all continuous distributions it would place restrictions for further development, concerning the number of distribution families available for fitting.

Despite these advantages, the K-S test has several important limitations [19]:

- It only applies to continuous distributions.
- It tends to be more sensitive near the centre of the distribution than at the tails.
- The distribution must be fully specified.

The last limitation results in a drawback for the presented approach until the parameter optimization module has been implemented, which is further discussed in section 5.6. To apply the K-S test using estimated parameters, results in a smaller probability than specified to do a Type 1 error (to reject the proposed distribution function even though it is a proper representation) [21]. However, since the actual K-S test is not fully performed in the software, the test statistics is still of help to compare distributions to each other.

5.5 Data Validation

The statistical correctness of the distribution fitting process performed by the software, was validated using output from the manual input data analysis, previously carried out in the case study. Moreover, the automatically analyzed data was used as input to the simulation model of the engine sub line. The result showed a production output 7% higher than the real-world system for the same period of time as the data was logged. However, despite the over-performance in model output, the result was satisfying. The discrepancy is explained as a direct result of the exclusion of the impact that the buffers after station 9 have on the production performance. Occasionally, these buffers cause line stops due to needs for quality adjustments. This behavior is not included in the simulation model.

5.6 Future development

This case study has shown the potential to increase rapidity in input data phases of DES projects using a middleware solution. However, the presented tool is not final, and will be further improved in order to fully meet organizations' needs of statistical reliability.

To increase the statistical correctness and user reliability, the software requires improved fitting functionality. Today, statistical distributions are compared to each other and the best representation is chosen. To be able to determine whether it is a significant representation or not, entire goodness-of-fit tests must be performed.

The problem described earlier that K-S tests do not work to a satisfactory level using estimated input parameters, needs to be solved before the K-S goodness-of-fit test is possible to fully implement. However, there are two ways of solving the problem. The solution that will be developed is to include a parameter optimization algorithm that calculates the optimal parameter setting instead of using the estimated ones. The alternative way is to perform other goodness-of-fit tests (e.g. Chi-Square and Anderson-Darling) providing better results for estimated parameters. This may also be done but with the motivation to give the user more options regarding the statistic analysis. Furthermore, the number of distribution families will be increased.

It could be desirable to connect the middleware directly to the PI database in order to reduce user involvement even more. The user interface will also be improved to facilitate for the user to specify the observations to extract, for example regarding dates and failure types. Finally, the concept will be enlarged to include more simulation inputs than only breakdown data.

6 Discussion

The main contribution of the approach presented in this paper is the possibility to automate the entire chain of input data handling. The demonstrator developed in the case study shows how extraction of data from an automated collection database and statistical preparation of the raw data can be performed for breakdown data, e.g. Time To Repair (TTR). However, the concept will be further developed to include more required input parameters to DES models at Volvo Trucks. The final aim is to fully utilize the increased use of automated data handling systems, not only for breakdown data. The approach provides a convenient first step for companies searching for a more efficient way to provide data for production analysis tools like DES. The applicability increases in companies, which do not have the uniform data structure needed to immediately adopt more automated concepts, such as ERP or CBS based simulation.

A future extension of the middleware solution could be to integrate the intelligence of the software into the intermediary computer application, e.g. off-line database. This would be one way of approaching the third "level" of input data management presented in section 2 [12]. Another very important issue for further development and implementation of the approach is to build the middleware functions according to a standardized data structure, like the Core Manufacturing Simulation Data (CMSD) model developed by NIST [22]. There is a need for a more unified way of representing the data to simulation models, which is owing to industry's problems with sharing data between simulation and other manufacturing applications [23]. Since many of the data sources discussed in this paper have gone far in their own development of data structures, the middleware can serve as a good tool to transform the representation into a standardized form for simulation.

One of the drawbacks associated with the presented approach is that efforts are required to write the code that extracts the data from the automatic collection system. Due to the variations among data collection systems, a general solution is not possible to build at present. Therefore, the approach is most appropriate for companies using DES on a continuous basis and consequently are able to benefit from the re-use of the middleware functionalities.

The main objective for the software was to enable fast input data analyses of the same quality as provided by manual analyses. However, despite the fulfilment of these objectives, some further improvements are needed and will be performed. To increase the level of statistical correctness, the middleware will be further developed by including more statistical distributions and a function for parameter optimization. Furthermore, the possibility to extract data of specified categories (e.g. breakdown reasons) needs to be improved and made more easily accessible via the user interface.

7 Conclusions

This paper has presented an approach to increase rapidity in input data management in DES projects. The strategy is to provide an IT-tool that automates the time-consuming activities of extracting raw data from original sources and preparing it for use in simulation models.

The tool is programmed in C++ and serves as a bridge in between the data sources and an intermediary computer application directly connected to the simulation model. The preparation of raw data includes calculation of input parameters to statistical distributions and distribution fitting using Kolmogorov-Smirnov test statistics.

The approach was tested in a case study at Volvo Truck's plant in Gothenburg, Sweden. The user involvement in the input data management phase was reduced significantly, which had a positive impact on the required time-consumption (from six hours to two minutes). However, the tool itself needs some further extensions and improvements to fully replace the traditional input data handling within the organization.

8 Acknowledgements

The funding for this research is granted by VINNOVA (Swedish Agency for Innovation Systems, integrates research and development in technology, transport and working life.) The authours would also like to thank Cecilia Berlin for providing valuable suggestions to enhance the presentation of this paper.

9 References

[1] U. Ericsson. Discrete Event Simulation, The Truth: Doctorial dissertation. Department of Production Engineering. Chalmers University of Technology, Sweden. 2001.

[2] Y. B. Moon and D. Phatak. Enhancing ERP system's functionality with discrete event simulation. *Industrial Management & Data Systems*, 105:1206-1224, 2005.

[3] J. Johnsson, Discrete Event Simulation as a Supporting Tool for Decision-Making in Virtual Enterprise, Licentiate Thesis, Department of Product and Production Development, Chalmers University of Technology, Sweden, 2004.

[4] S. Thomke and T. Fujimoto. The effect of "Front-Loading" problem-solving on product development performance. *Journal of Product Innovation Management*, 17:128-142, 2000.

[5] A. Skoogh and B. Johansson. Time-consumption Analysis of Input Data Activities in Discrete Event Simulation Projects. *Proceedings of the Swedish Production Symposium,* Gothenburg, August 28-30, 2007. (Submitted for publication).

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

[7] W. Trybula. Building simulation models without data. *IEEE International Conferance on Systems, Man, and Cybernetics*, Humans, Information and Technology, 1:209-214, San Antonio, October 2-5, 1994.

[8] S. Umeda and A. Jones. Simulation in Japan: State-of-the-art update. Technical Report NISTIR 6040, National Institute of Standards and Technology, U.S. Department of Commerce, Technology Administration, 1997.

[9] A. Ingemansson, T. Ylipää and G. S. Bolmsjö. Reducing bottle-necks in a manufacturing system with automatic data collection and discrete event simulation. *Journal of Manufacturing Technology Management*, 16:615-628, 2005.

[10]C-F. Ho, W-H. Wu and Y-M. Tai. Strategies for the adaption of ERP systems. *Industrial Management & Data Systems*, 104:234-251, 2004.

[11]S.C.L. Koh and S.M. Saad. Managing uncertainty in ERP-controlled manufacturing environments in SMEs. *International journal of production economics*, 101:109-127, 2006.

[12]N. Robertson and T. Perera. Automated data collection for simulation?. *Simulation Practice and Theory*, 9:349-364, 2002.

[13]S. Jain. Simulation in the next millennium. *Proceedings of the 1999 Winter Simulation Conference*, Simulation, a bridge to the future,

2:1478-1484, Phoenix, Arizona, December 5-8, 1999.

[14]L.G. Randell and G.S. Bolmsjö. Database driven factory simulation: a proof-of-concept demonstrator. *Proceedings of the 2001 Winter Simulation Conference*, pp. 977-983, Arlington, Virginia, December 9-12, 2001.

[15]M. Jägstam, A. Ng, B. Johansson, P. Gullander, P. Klingstam and J. Svenningstorp. FACTS-Factory Analyses in Conceptual Phase using Simulation. *Proceedings of the Swedish Production Symposium*, Gothenburg, August 28-30, 2007. (Submitted for publication).

[16]A. Patchong, T. Lemoine and G. Kern. Improving Car Body Production at PSA Peugeot Citroën. INTERFACES, Vol. 33, No. 1, January-February 2003.

[17]A. Ng, M. Urenda Morris, J. Svensson, A. Skoogh and B. Johansson. FACTS Analyser: An innovative tool for factory conceptual design using simulation. *Proceedings of the Swedish Production Symposium*, Gothenburg, August 28-30, 2007. (Submitted for publication).

[18]D.C. Montgomery and G.C. Runger. Applied Statistics and Probability for Engineers. John Wiley & Sons, Inc., New York. 1999.

[19]*NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/ (Accessed on April 23, 2007).

[20]http://mathworld.wolfram.com/DistributionFunction.html (Accessed on May 11, 2007).

[21]A. M. Law and W. D. Kelton. Simulation Modelling and Analysis (3rd ed.). McGraw-Hill, New York. 2000.

[22]C. McLean and S.L. Leong. Simulation standards: a framework for standard modular simulation. *Proceedings of the 2002 Winter Simulation Conference*, Exploring New Frontiers, pp 1613-1620, San Diego, California, December 8-11, 2002.

[23]M. Johansson, S. Leong, F. Riddick, B. Johansson, A. Skoogh and P. Klingstam. Using standardized input data structure for a flow shop simulation model – a test implementation of core manufacturing simulation data (CMSD). *Proceedings of the 2007 Winter Simulation Conference*, Washington DC, December 9-12, 2007. (Submitted for publication)

10 Authors Biographies

ANDERS SKOOGH was born in Floby, Sweden, 1980. He attended Chalmers University of Technology at Automation and Mechatronics Engineering, where he obtained his M.Sc. degree in 2005. Anders has industrial experience of Discrete Event Simulation from his former employment as logistics developer at Volvo Car Corporation. He is now working as a PhD student in the field of Discrete Event Simulation at the Department of Product and Production Development, Chalmers University of Technology, Sweden. His email is <anders.skoogh@chalmers.se>.

JEAN-PATRICK ANDRÉ was born in Grenoble, France, 1984. He is studying at the IFMA (French Institute for Advanced Mechanics), France, at the Advanced Manufacturing Systems Department with the aim of obtaining a diploma of engineering in 2008. He is currently working as a guest research student in the field of DES at the Department of Product and Production Development, Chalmers University of Technology, Sweden. His email is <Jean-Patrick.Andre@ifma.fr>.

CATARINA DUDAS was born in Gothenburg, Sweden 1978. She went to the University of Stockholm where she obtained her M.Sc degree in Numerical analysis in 2004. She is now working as a Discrete Event Simulation engineer at Volvo Technology, Gothenburg, Sweden. Her e-mail address is <catarina.dudas@volvo.com>

JACOB SVENSSON was born 1982 in Alingsås on the west coast of Sweden, in the vicinity of Gothenburg. Early spring 2007 he got his degree in the Master of Science program Industrial Engineering and Management at Linköping University, Sweden. During his Master Thesis work at Volvo Technology, Gothenburg Sweden, he was introduced to DES. Since the 1st of December in 2006 he has worked intensely with DES as a project employee at the University of Skövde, Sweden. His email is <jacob.svensson@his.se>.

MATÍAS URENDA MORIS is a Ph.D.-student and part-time lecturer at the Centre for Intelligent Automation, University of Skövde, Sweden. He received a BSc in Automation Engineering at University of Skövde and an MSc in Manufacturing Management at University of Skövde in collaboration with Loughborough University, UK. His main research area is DES for health care and manufacturing systems. His email address is <matias.urenda.moris@his.se>.

BJÖRN JOHANSSON is an assistant professor at the Department of Product and Production Development, Chalmers University of Technology. His research interest lies in the area of Discrete Event Simulation for manufacturing industries. He is interested in the specifics of modular modeling methodology for effective and swift model-building, which is accompanied by software development, user interfaces, and input data architectures. His email address is <Bjorn.Johansson@chalmers.se>.

