VISUAL ANALYTICS OF MANUFACTURING SIMULATION DATA

Niclas Feldkamp
Sören Bergmann
Steffen Strassburger

Department for Industrial Information Systems
Ilmenau University of Technology
P.O. Box 100 565
98684 Ilmenau, GERMANY

ABSTRACT

Discrete event simulation is an accepted technology for investigating the dynamic behavior of complex manufacturing systems. Visualizations created within simulation studies often focus on the animation of the dynamic processes of a single simulation run, supplemented with graphs of certain performance indicators obtained from replications of a simulation run or a few manually conducted simulation experiments. This paper suggests a much broader visually aided analysis of simulation input and output data and their relations than it is commonly applied today. Inspired from the idea of visual analytics, we suggest the application of data farming approaches for obtaining datasets of a much broader spectrum of combinations of input and output data. These datasets are then processed by data mining methods and visually analyzed by the simulation experts. This process can uncover causal relationships in the model behavior that were previously not known, leading to a better understanding of the systems behavior.

1   INTRODUCTION

Analyzing discrete event manufacturing simulations is usually performed by looking at a few distinct output parameters (e.g., throughput, resource utilization) according to the simulation project scope. This is typically driven by guiding questions such as "which scheduling strategy performs best" or "what size does this buffer need". As a side effect of this approach, one does not have to model aspects of the system that are not influential to answering the proposed questions.

Simulation experimentation is usually conducted manually, in the best case assisted by some kind of experiment manager. Simulation based optimization on the other hand tries to find an optimal solution for some set of output values of the simulation by varying selected input parameters and automatically executing repeated simulation runs. Input parameters in that scenario are varied according to the applied optimization algorithm.

In both cases, the simulation analyst usually takes an educated guess based on his experience which input parameters might be influential on the project scope and therefore time and effort is invested in experimenting with these focus parameters in a fixed system configuration environment. Kleijnen et al. refer to this as the trial-and-error approach to finding a good solution and argue that simulation analysts should spend more time in analyzing than building the model (Kleijnen et al. 2005).

In this paper, we suggest a much broader visually aided analysis of simulation input and output data and their relations. By adopting techniques from the database sector, a visually aided approach for simulation data analysis is outlined. We call this approach "visual analytics of manufacturing simulation data". The outlined approach combines elements from big data technologies, data mining, knowledge discovery in databases, and interactive visual analysis.

While a previous paper outlined the data mining and knowledge discovery aspects of the approach (Feldkamp, Bergmann, and Strassburger 2015), we here focus on the visualization side.

The remainder of this paper is structured as followed: In section 2 we introduce visual analytics and related work. Section 3 discusses the general process for visual analytics of manufacturing simulation data. Section 4 introduces a case study and discusses different visualization techniques which can be beneficial for interactive visual exploration of simulation input and output data. Section 5 concludes the paper and discusses future work.

2   VISUAL ANALYTICS

Visualization in general is an important tool when an interpretation of data is needed. As such, visualization is commonly applied in almost any simulation study in some way. Typical visualization techniques applied in the context of discrete event simulations include the animation of the dynamic processes of a single simulation run, often focusing on the movement of entities through the system under investigation. Other visualization techniques are applied as part of traditional simulation output analysis (Law 2014) and include graphs of certain performance indicators obtained from replications of a simulation run or a few manually conducted simulation experiments.

Visual analytics goes beyond visualization techniques that are commonly applied in simulation studies today. Visual analytics aims at integrating the human into a visual data exploration process (Keim 2002). Integrating the human with its perceptual abilities to large data sets is key for the success in this approach. The basic idea of visual analytics is to present the data in some visual form, allowing the human to get insight into the data, draw conclusions, and interact with the data to confirm or disregard those conclusions. Suitable visualizations can utilize the human ability to recognize patterns and coherence. This is also referred to as visual reasoning, which is the science of synthesizing information from massive datasets in order to provide understandable assessments that can be used effectively for communication action and decision making (Thomas and Cook 2005).

Visual analytics can be defined as "an iterative process that involves information gathering, data preprocessing, knowledge representation, interaction and decision making" (Keim et al. 2008). It combines the strengths of machines, e.g., for processing huge amounts of data, with those of humans, e.g., for pattern recognition and drawing conclusions. As such, visual analytics combines methods from knowledge discovery in databases (KDD), statistics and mathematics as driving forces behind automatic data analysis with human capabilities to perceive, relate, and conclude (Fayyad, Piatetsky-Shapiro, and Smyth 1996).

A proper visual analysis should show the important, big picture first, and uncover details on demand. The visual analytics process is furthermore conducted through a frequent shift between visual and computational analysis of data (Keim et al. 2008), which makes it a semi-automatic process.

Among others, main benefits of visual analytics are increasing cognitive resources by expanding human working memory, reducing time exposure by representing large amounts of data in a small space and promoting recognition of patterns as well as exploring relationships that otherwise remain hidden or at least were more difficult to find (Thomas and Cook 2005).

The term "visual analytics" is commonly associated with the terms "Big Data" and "Data Mining" (Chen et al. 2009; Wong et al. 2012). Data Mining is an important substep in the KDD process involving the algorithmic tools for statistical processing, e.g., clustering or regression (Han, Kamber, and Pei 2012). Hence, a prerequisite for its applicability is the presence of large amounts of data. As we aim at adopting visual analytics for a better analysis of discrete event simulations, we propose the usage of a "data farming" approach (as suggested in (Horne and Meyer 2005; Sanchez 2007)) for designing large scale simulation experiments for obtaining detailed data sets which document the behavior of our system under investigation. In this context, farming refers to data creation and describes how one should cultivate simulation experiments to maximize the data yield (Sanchez 2014). As these datasets can grow large very fast, they become too big to manually review each observation individually, making visual analytics a reasonable tool for their analysis.

3   ADOPTING VISUAL ANALYTICS FOR MANUFACTURING SIMULATION DATA

Simulation experiments can be regarded as a black box that transforms input data into output data. In a manufacturing context, the input data are adjustable system parameters like inter arrival times, buffer sizes or scheduling strategies. Result data on the other hand is composed of the system's performance indicators like throughput times or machine utilization.

Figure 1 shows our general approach for knowledge discovery in manufacturing simulations through visual analytics. First we have to identify adjustable system input parameters. The next step is estimating bottom and top factor level limits and defining experiments accordingly. Moreover, measurable output parameters have to be defined. This includes possible data types and measurement scales. As a side note, interpretation of results certainly implies a decent understanding of the underlying model. Although we abstracted the model into a black box in our concept, knowledge of the model and creation of knowledge through result interpretation are undeniably connected in a specific application. Furthermore, we assume that the model is correct and verified, otherwise garbage-in garbage-out principle would be in effect.

After simulation experiments have been conducted, output data can be processed through data mining methods and represented with suitable visualizations. An initial investigation should explore shape and distribution of output parameters and linkage between those. Yet before looking at related input parameters, possible interesting patterns might even be discovered by only looking at output data.
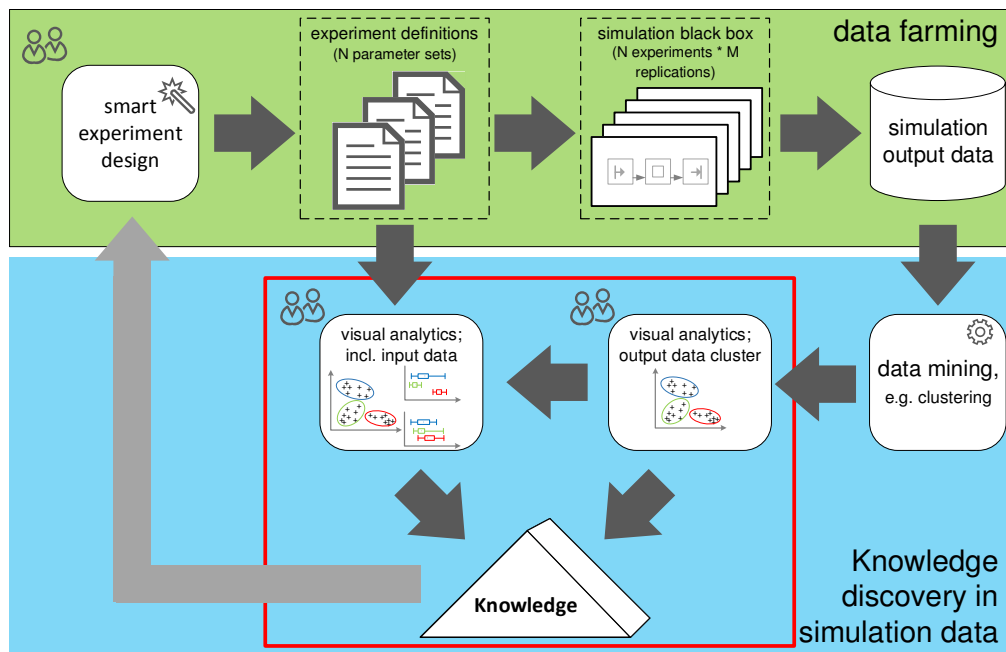

Figure 1: Visual Analytics (red border) as key technique of a knowledge discovery process for discrete event manufacturing simulations.

Afterwards, visualizations should establish links between corresponding input and output parameter sets. This eventually yields interesting relationships between corresponding input/output parameter values. Findings should be interpreted and transformed into knowledge by verifying them through the design and conducting of additional experiments in such a way that an iterative process emerges.

The prerequisite for applying visual analytics techniques is the transformation of data obtained through the simulation experiments into a visually representable form. As we are in total control of data creation, we here do not have to deal with faulty or incomplete data, that data mining techniques typically have to eliminate through selection and preprocessing steps. Furthermore, the amount of generated output data solely depends on the experimental setup and on the performance measures of interest and can be

adjusted through intelligent design of simulation experiments.

The transformation of the raw output data includes the application of data mining methods such as clustering (Feldkamp, Bergmann, and Strassburger 2015). We suggest to apply a clustering of entire simulation runs based on selected output parameters. As a result, experiments within the same cluster are similar regarding the selected system performance measures.

Figure 2 left side shows a small notional demonstration. Note that this example includes a very small number of simulation runs, respectively cluster points. In a large scale experiment design, the number of cluster points is too big to identify single points, but rather a trend and tendency of point distribution is the focus of visual representation.
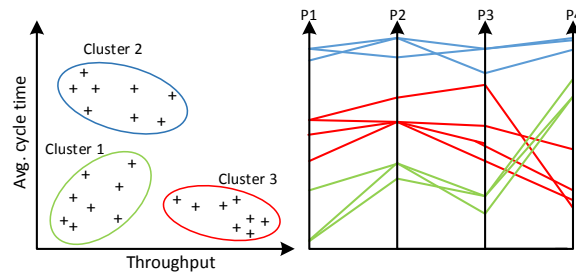

Figure 2: Left side: Two-dimensional representation of simulation output clustering. Each point represents a single experiment and each cluster represents a group of experiments that are similar regarding throughput and average cycle time. Right side: Parallel coordinates visualization with four parameters.

If the clustering algorithm needs to process more than two variables, a multidimensional data representation is necessary. The parallel coordinates method as shown in Figure 2 (right side) has proven to be a quite intuitive method for interactive visual inspection and will be further discussed in section 4.

Visualizing the clustering results will yield a first visual impression and a quick overview of the performance measures from a large numbers of experiments at once. We also get a first impression on how those experiments are distributed regarding the observed performance measures.

The next step consists in the investigation of the clusters and the experiments they are composed of. Each experiment (i.e., simulation run) represents a dataset composed of a set of output parameters with corresponding input parameters. If we look at a distinct cluster, we can also investigate which input parameters are linked to the experiments in this clusters. Following the iterative process of visual analytics, we may apply further statistical methods to gain additional insights on how input parameter values are distributed among the different clusters.

Taking a look at cluster 2 in Figure 2 as an example, the system's performance of corresponding experiments is rather poor (amplified by low throughput and high cycle times), whereas system performance of cluster 3 experiments are quite good, with rather high throughputs and low cycle times. From here we can investigate which input settings led to the corresponding systems performance measures that define this cluster.

If we are able to outline a significance of distinct input parameters values that lead to a specific cluster allocation that is yet even unexpected, we actually created new knowledge. At best this knowledge is utilized to support factory planning decisions. For a better understanding, we clarify the visual analytics process with a simple use case example in the following chapter.

4   USE CASE SCENARIO

4.1   Design of Experiments

For a demonstration of our approach, we use an enhanced single server model. Seven different types of products enter a manufacturing system through a single source. Each product type has distinct processing

and setup times. Additionally, each job is assigned a unique due date when entering the system. Jobs are sorted before entering the processing station. The station requires a proper setup according to the product type. Setup time of the station varies depending on the product type. Figure 3 shows a screenshot of simulation model.
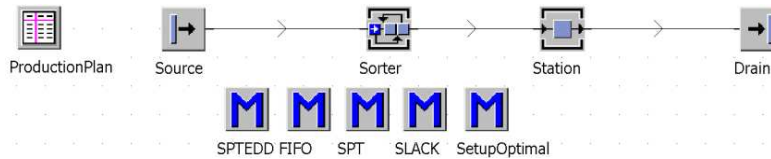

Figure 3: Screenshot of the Plant Simulation Model.

Adjustable input parameters of the model include the inter arrival time of jobs, the sorter's maximum capacity, the sorting strategy, and the mixture of product types. Output parameters include throughput, mean cycle times, setup times, due date delays, and station utilization.

Table 1: Experimental design.

Input Parameters	Margins	Levels
Inter arrival time	60s-240s	18
Sorter capacity	10-1000	10
Sorter strategy	5 strategies	5
Product mixture (Seven product types)	0-100% per product	47 Experiments (NOLH-Sampling)
Random number stream	1-10	10

Table 1 shows the design of experiments. For inter arrival times we chose fixed values from 60 seconds up to 240 seconds, which are also the minimum and maximum processing times per job. Sorter capacity was varied from 100 to 1000 slots. 5 different sorting strategies were investigated: First in first out (FIFO), shortest processing time (SPT), minimum slack time (SLACK), a weighted combination of SPT and earliest due date (SPTEDD), and sorting according to current station setup state.

Regarding the product mixture it was not possible to perform experiments with a full factorial design from 0%-100% per product type, as this would result in at least 7^{101} product mixes. Therefore, a sampling method was needed to reduce the number of experiments while maintaining large coverage of the parameter space.

We created our experiments based on the nearly orthogonal latin hypercube (NOLH), which is a sampling method that offers a realistic distribution of parameter variability. Orthogonal means that these designs try to minimize correlation between input variables and factor-level combinations are evenly sampled (Hernandez, Lucas, and Carlyle 2012; Ye 1998). With the sampling design we ended up with 47 manifestations regarding the product mixture and all input variables combined resulted in a total number of 491.160 experiments. Note that we used the sampling method for the product mixture and full factorial design for the other variables. This is possible because our test scenario is simple. When models become more complex and the number of input variables increases, sampling over all input variables is necessary.

We portioned the simulation runs onto multiple machines and conducted them with Tecnomatix Plant Simulation. For storing input/output datasets we used a MongoDB noSQL database (MongoDB Inc. 2010). We choose the noSQL approach, because the schemaless and highly scalable database implementation allows to quickly adapt to dataset modifications while neglecting revisions of data schema.

4.2   Clustering Simulation Output Data

A prerequisite for visual analytics is the clustering of the datasets under investigation. In conjunction with simulation data analysis, we propose to group individual simulation runs into clusters of similar output performance values.

The first question to answer is therefore which output parameters should be taken into account for the clustering algorithm. Because we can observe a large number of distinct output parameters, datasets are likely multi-dimensional. In multi-dimensional data, similar datasets are composed of subsets that often correlate among each other. Furthermore, correlation among subsets indicates valuable relationship information between attributes and eventually hidden casualization (Böhm et al. 2004).

Taking this into account, we can identify possible parameter combinations for clustering by looking at spatial shapes of subsets and searching for interesting patterns. A valuable form of visualization for this purpose are scatterplot matrix visualizations. They draw a two-dimensional scatterplot for each variable pair of the multi-dimensional data set. This allows us to visually search for any noticeable, interesting structures among subsets. For example, a rather structure less point cloud with a high proportion of outliers would result in a decrease of the clusters entropy, so pairwise visual investigation of subsets eventually leads to additional insights (Wilkinson, Anand, and Grossman 2006) and therefore meaningful clusters.

Figure 4 shows selected output parameters of our case study as a scatterplot matrix. From this figure, we can identify correlated structures between throughput, station utilization, and station setup proportion. What we see are strongly correlated structures among those parameters, which can be expected in a single server model.
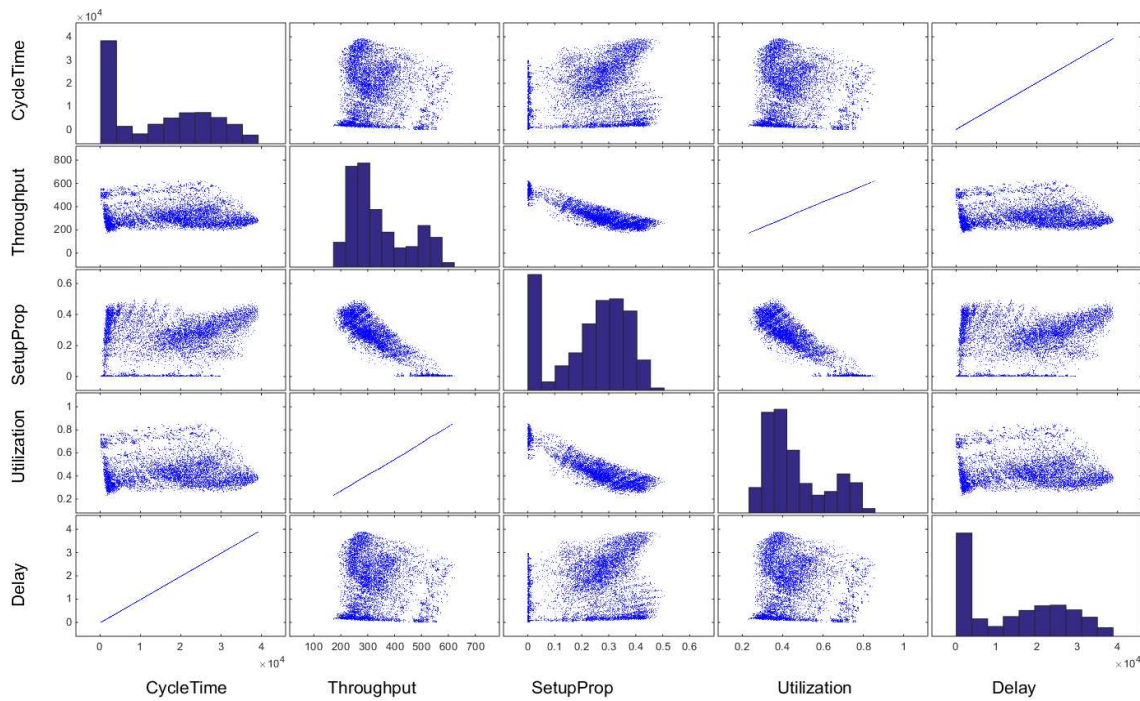

Figure 4: Scatterplot matrix of selected output parameters. Visual inspection can identify parameter sets of apparently correlated structures (e.g., throughput – setup proportion, utilization – setup proportion, etc.).

For demonstration of the concept we therefore use these parameters for clustering our simulation runs. As a result, we receive a good separation of clusters that allows us to build a group hierarchy from good to poor performance runs (Figure 5). The three dimensional data points are represented in a one

dimensional parallel coordinates plot as described in the prior chapter. For a better visual representation, each column has been standardized to a 0 mean and a standard deviation of 1, so this visualization is used to evaluate the comparison between experiments, not to look at the actual parameter values. Regardless of cluster allocation, this visual representation gives a quick overview (regarding the three selected performance measures) of almost half a million simulation experiments.

A counter example for a bad clustering would be the usage of throughput, setup proportion, and delay for the clustering algorithm, because the underlying variables have a very high proportion of outliers and are less structured.
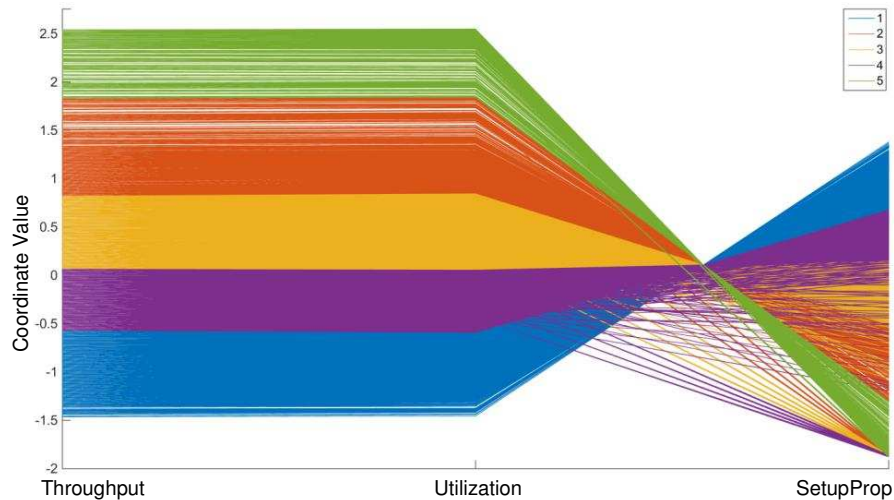

Figure 5: Parallel coordinates visualization of the clustering based on three parameters. Each vertical axis represents one output parameter.

Looking at Figure 5, the experiments that compose cluster 1 (blue color) have the lowest performance, with a low throughput, high proportion of machine setups but still low overall station utilization. This performance is worse because e.g. experiments in cluster 4 (purple color) perform much better among throughput and station utilization to the proportion of station setup times. Experiments in cluster 5 (green) and cluster 2 (orange) perform best, even though cluster 2 has a few outlier experiments that have very low setup proportion but do not fulfil the expected linear relation to throughput and station utilization. Further investigation should – among other things - find out why those experiments have lower throughput than cluster 5 experiments despite low setup time proportion.

4.3 Investigation of Relationships to Input Data Parameters

After we grouped the experiments in clusters regarding system performance (of selected output parameters), the next step in the proposed visual analytics process is finding relations to experiment input parameters and investigating which input parameters values the clusters are composed of.

For example, a certain input parameter value range that appears frequently in a distinct cluster indicates that this parameter is likely influential to the cluster allocation. For that purpose, again, different options for visually aided investigation are possible.

4.3.1 Relationships to Job Inter Arrival Time

As a start, we may want to investigate the influence of the input parameter "inter arrival time" on the cluster allocation. For reasons of simplicity, we want to experiment further with the parallel coordinates visualizations introduced above. We now include the desired input parameter "inter arrival time" in the plot (Figure 6).
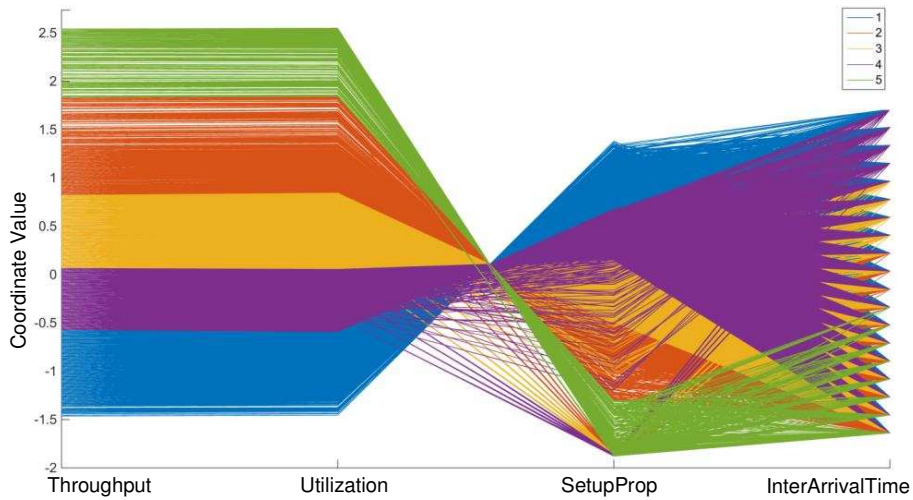
Figure 6: Visualization of the clusters now including input parameter inter arrival time.
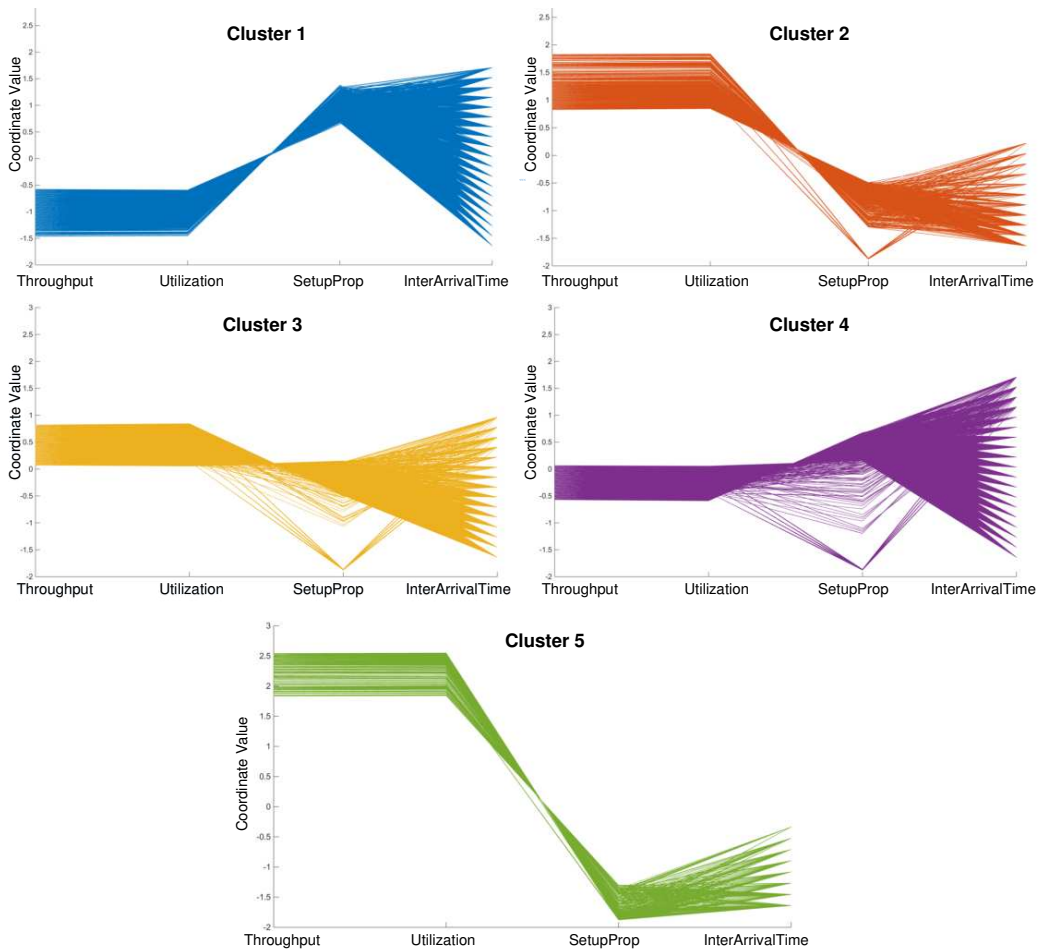

Figure 7: Separate parallel coordinates visualizations of the individual clusters. Interesting aspects previously hidden due to overlays can now be uncovered (e.g., outliers concerning setup proportion in clusters 2-4, value range overlaps concerning inter arrival time in almost all clusters).

With some imagination, we could already conclude from the good (green) cluster in Figure 6 that low values for the input parameter inter arrival time seem to have a positive influence on the selected output performance indicators. However, as parallel coordinate plots can become unhandy rather quickly due to multiple overlays of the depicted simulation parameters, we have the option to interactively experiment with the data sets by fading in and out desired or undesired datasets. That way we can eliminate any undesired overlays in the visualization.

Inspecting the clusters now individually as depicted in Figure 7 we see overall low inter arrival times in cluster 5 (green, good performance). However, as also simulation runs from other clusters have overlapping inter arrival time value ranges, we can conclude that inter arrival time cannot be the only input variable that is influential to cluster allocation and good system performance.

In cluster 5, we can further see an apparent correlation between inter arrival time and the output parameter "setup proportion", leading to the desire to further investigate on reasons for that. As a consequence, we may now want to look at the influence both the product mixture and the sorting strategy in these runs.

4.3.2 Relationships to Product Mixture

Several options for investigating the product mixture are possible, including histograms of class distributions and box plots of certain ratio parameters (Feldkamp, Bergmann, and Strassburger 2015). Due to space constraints, we here stay with parallel coordinates plots and visualize overall throughput alongside individual throughputs of each product type (Figure 8).

We can see that the highest cumulative throughputs (parameter to the farthest left on x-axis) are correlated with very unbalanced proportions of throughput regarding the product type. This can be concluded from the spikes in the parameters showing throughput per product type. The simulation runs in the good (green) cluster have a tendency to prefer product mixes with a single product type leading to the spikes in throughput of this product type (green spikes in product type throughput). Please note that the individual spikes in throughput of a certain product type result from different simulation runs.

Also from Figure 8 we see that the flatter the lines become, the more balanced is the throughput over the product types, leading to a decreased overall throughput. Moreover, the plot also reflects the hierarchy of clusters from good to poor system performance. Analysis of this could be confirmed by individual cluster plots, which are omitted here due to space constraints.
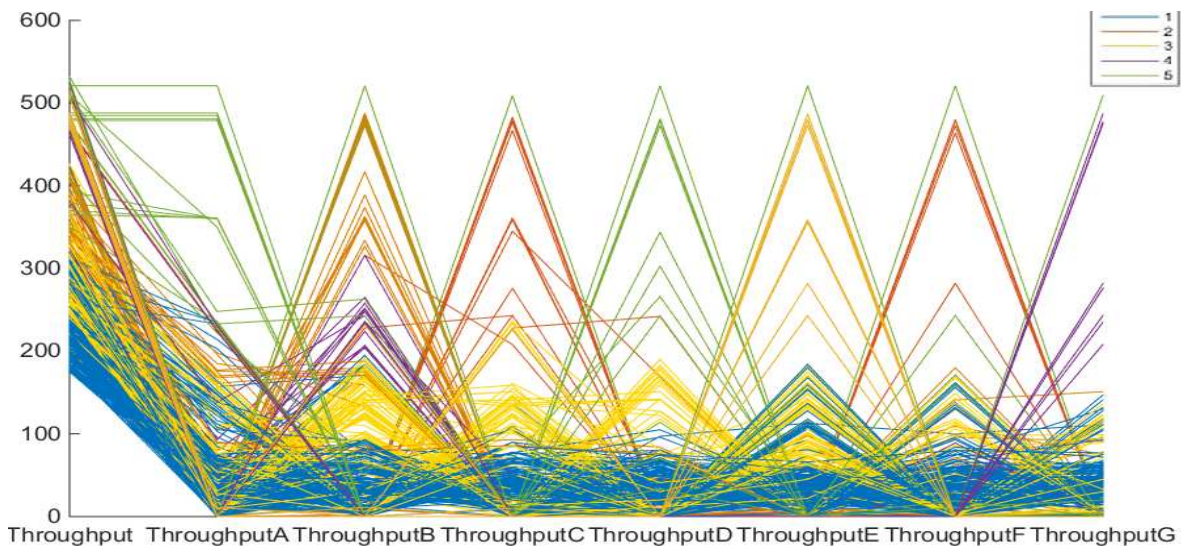

Figure 8: Parallel plot of measured cumulative throughput (parameter on the farthest left) alongside with throughputs per product type (from A to G), color-coded with corresponding cluster membership.

4.3.3 Relationships to Job Sorting Strategy

As of now, we have learned that low inter arrival times and unbalanced product mixes seem to have a positive effect on the selected performance indicators. While this may be obvious for the depicted single server model, visual analysis of more complex models may well reveal less obvious knowledge.

Staying with the case study, we may now want to investigate for the reasons of our observations, leading us to the need to investigate the applied sorting strategy. Since this variable is on a nominal scale and has no numeric ratio, we now need to find other visual representations than previously applied. An informative visualization should show the proportion of the 5 strategies per cluster. A potential visualization would be pie charts. We here suggest the usage of mosaic plots, as they are visually somewhat more expressive and more intuitive to read (Figure 9).

We see that Cluster 1 does not include experiments conducted with the setup optimal strategy. We can therefore exclude this strategy as a driving factor for bad system performance. On the contrary, looking at the high proportions of this scheduling strategy in Clusters 2, 3 and 5 we may even associate this strategy as an important factor that leads to good system performance. For this case study, this is obviously an accurate conclusion since this strategy leads to less machine blocking by setup processes.
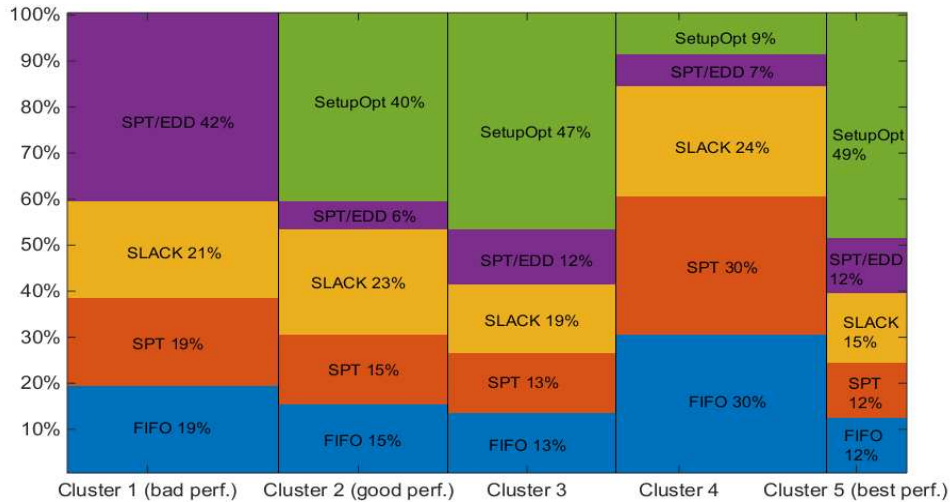

Figure 9: Mosaic plot of sorting strategy proportion per cluster. Note that the width reflects the overall amount of simulation runs in the cluster.

In summary, regarding to our simple single server model, we found out by visual analytics that short inter arrival times of jobs are key to a good system performance, at least up to a break-even point where machine blockings due to setup processes decrease the systems performance again.

In that situation a setup optimal scheduling strategy lessens this effect. Furthermore, focusing on reducing the product mixture variety is always superior to any sophisticated product mixture, regardless of the applied scheduling strategy.

In line with the proposed iterative process for visual analytics based knowledge discovery, new sets of experiments could now be designed to either confirm or dismiss these findings, in order to create further knowledge. Next steps would include the design of additional experiments that focus on exploring the limits of those input values that apparently lead to good system performance.

5   DISCUSSION AND SUMMARY

Although options for finding hidden knowledge in a simple single server model are quite limited, we demonstrated how a visual analytics based knowledge discovery process for manufacturing simulation data can be performed. We have shown how this process can be adopted for simulation output analysis

and how visual analytics combined with elements from big data analytics can be usefully applied in simulation studies. Our approach brings in an additional viewpoint to discrete event simulation analysis, which can be considered alongside traditional experimentation techniques conducted in simulation studies.

Moreover, our approach might be more appealing to people who are not "simulation experts", because in contrast to working through lines of numbers, visualization is a more striking way of handling simulation data. On the other hand it may be less precise since it is more open to interpretation. Planning decisions should always be assisted with traditional simulation studies, but our approach on covering a broadband of system behaviors can help to guideline a certain path.

Visual linkage between input/output parameters of different scale measurement levels (e.g. nominal and numerical) leads to new challenges in visualization methods. This is in due consideration that visualization should be easily understandable and interpretable. In a best case scenario, the visualization can promote intuitive investigation and knowledge discovery by leveling the tradeoff between correctness, complexity, and simplicity.

Future research is needed for deriving visualization methods and tool sets that are especially suited for visual analytics in the context of manufacturing simulation data, or even discrete event simulation data in general. For the demonstrations shown in this paper we applied a variety of toolsets, including Plant Simulation, Matlab, and Excel. For visual analytics to become commonplace, an integration of the demonstrated methods with commonly applied simulation packages will be needed.

Taking for example the problem of product mixture, we still face the challenge on how to feasibly plot one single depiction of thousands of experiments with each experiment having different mixture proportions. Furthermore, visual analytics demands interaction with data, so an easy and real-time user interaction with the visualizations are also important aspects of future research.

Other future research options include the investigation of other data mining and machine learning techniques for simulation data analysis. Promising methods are for example classification methods like decision trees or support vector machines and association rule learning like pattern and sequence mining. Furthermore, online machine learning methods (Bergmann, Stelzer, and Strassburger 2014) could enable the combination of real world factory data logging and simulation data for a real time data or near-real time data analysis.

6 REFERENCES

Bergmann, S., S. Stelzer, and S. Strassburger. 2014. "On the Use of Artificial Neural Networks in Simulation-Based Manufacturing Control." Journal of Simulation (JOS) 8:76–90.

Böhm, C., K. Kailing, P. Kröger, and A. Zimek. 2004. "Computing Clusters of Correlation Connected Objects." In Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, edited by P. Valduriez, G. Weikum, A. C. König, and S. Dessloch, 455–466.

Chen, M., D. Ebert, H. Hagen, R. S. Laramee, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. 2009. "Data, Information, and Knowledge in Visualization." IEEE Comput. Graph. Appl. 29:12–19.

Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth. 1996. "From Data Mining to Knowledge Discovery in Databases." AI Magazine 17:37–54.

Feldkamp, N., S. Bergmann, and S. Strassburger. 2015. "Knowledge Discovery in Manufacturing Simulations." In Proceedings of the 2015 ACM SIGSIM PADS Conference, 3–12.

Han, J., M. Kamber, and J. Pei. 2012. Data Mining Concepts and Techniques, 3rd ed., Amsterdam: Elsevier/Morgan Kaufmann.

Hernandez, A. S., T. W. Lucas, and M. Carlyle. 2012. "Constructing Nearly Orthogonal Latin Hypercubes for Any Nonsaturated Run-Variable Combination." ACM Transactions on Modeling and Computer Simulation 22:1–17.

Horne, G. E., and T. E. Meyer. 2005. "Data Farming: Discovering Surprise." In Winter Simulation Conference, 2005, 1082–1087.

Keim, D. A. 2002. "Information Visualization and Visual Data Mining." IEEE Transactions on Visualization and Computer Graphics 8:1–8.

Keim, D. A., F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. 2008. "Visual Analytics: Scope and Challenges." In Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, edited by S. Simoff, M. H. Boehlen, and A. Mazeika. Berlin, Heidelberg: Springer.

Kleijnen, J. P., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa. 2005. "State-of-the-Art Review: A User's Guide to the Brave New World of Designing Simulation Experiments." INFORMS Journal on Computing 17:263–289.

Law, A. M. 2014. Simulation Modeling and Analysis, 5th ed., New York, N.Y.: Mcgraw Hill Book Co.

MongoDB Inc. 2010. "Why Schemaless?" Accessed February 10, 2015. http://blog.mongodb.org/post/119945109/why-schemaless.

Sanchez, S. M. 2007. "Work Smarter, Not Harder: Guidelines for Designing Simulation Experiments." In Proceedings of the 2007 Winter Simulation Conference, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 84–94, New York, NY: ACM, IEEE.

Sanchez, S. M. 2014. "Simulation Experiments: Better Data, Not Just Big Data." In Proceedings of the 2014 Winter Simulation Conference, edited by A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 805–816.

Thomas, J. J., and K. A. Cook. 2005. Illuminating the Path. Research and Development Agenda for Visual Analytics, Los Alamitos, California: IEEE Computer Society.

Wilkinson, L., A. Anand, and R. Grossman. 2006. "High-Dimensional Visual Analytics: Interactive Exploration Guided by Pairwise Views of Point Distributions." IEEE Transactions on Visualization and Computer Graphics 12:1363–1372.

Wong, P. C., H.-W. Shen, C. R. Johnson, C. Chen, and R. B. Ross. 2012. "The Top 10 Challenges in Extreme-Scale Visual Analytics." IEEE Computer Graphics and Applications 32:63–67.

Ye, K. Q. 1998. "Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments." Journal of the American Statistical Association 93:1430–1439.

AUTHOR BIOGRAPHIES

NICLAS FELDKAMP holds bachelor and master degrees in business information systems from the University of Cologne and the Ilmenau University of Technology, respectively. He is currently working as a doctoral student at the Department of Industrial Information Systems of the Ilmenau University of Technology. His research interests include data science, business analytics, and industrial simulations. His email address is niclas.feldkamp@tu-ilmenau.de.

SÖREN BERGMANN holds a Doctoral and Diploma degree in business information systems from the Ilmenau University of Technology. He is a member of the scientific staff at the Department for Industrial Information Systems. Previously he worked as corporate consultant in various projects. His research interests include generation of simulation models and automated validation of simulation models within the digital factory context. His email is soeren.bergmann@tu-ilmenau.de.

STEFFEN STRASSBURGER is a professor at the Ilmenau University of Technology and head of the Department for Industrial Information Systems. Previously he was head of the "Virtual Development" department at the Fraunhofer Institute in Magdeburg, Germany and a researcher at the DaimlerChrysler Research Center in Ulm, Germany. He holds a Doctoral and a Diploma degree in Computer Science from the University of Magdeburg, Germany. He is further an associate editor of the Journal of Simulation. His research interests include distributed simulation, automatic simulation model generation, and general interoperability topics within the digital factory context. He is also the Vice Chair of SISO's COTS Simulation Package Interoperability Product Support Group. His web page can be found via www.tu-ilmenau.de/wi1. His email is steffen.strassburger@tu-ilmenau.de.