Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

KNOWLEDGE DISCOVERY AND ROBUSTNESS ANALYSIS IN MANUFACTURING SIMULATIONS

Niclas Feldkamp
Soeren Bergmann
Steffen Strassburger

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

Thomas Schulze

School of Computer Science
Otto-von-Guericke-University Magdeburg
Universitätsplatz 2
39106 Magdeburg, GERMANY

ABSTRACT

Discrete event simulation is an established methodology for investigating the dynamic behavior of complex manufacturing and logistics systems. Traditionally, simulation experts conduct experiments for predetermined system specifications focusing on single model aspects and specific analysis questions. In addition to that, the concept of data farming and knowledge discovery is an ongoing research issue that consists of broad scale experimentation and data mining assisted analysis of massive simulation output data. As an extension to this approach, we propose a concept for investigating the robustness of complex manufacturing and logistic systems which are often very sensitive to variation and noise. Based on Taguchi's loss function, we developed a concept including data farming and visual analytics methodologies to investigate sources of variation in a model and the factor values that make a configuration robust. The concept is demonstrated on an exemplary case study model.

1 INTRODUCTION

Data farming and knowledge discovery in simulation data are popular and ongoing issues in current simulation methodology research. The concepts comprise broad scale simulation experimentation and the use of data mining algorithms in order to uncover unknown relationships and effects in the model to gain useful information, leading to a better understanding of the system's behavior and an increased decision support (Feldkamp, Bergmann, and Strassburger 2015b). Data farming can also be used for robustness evaluations, which enable the investigation of how prone a system is to be effected by noise and the finding of robust configurations. In the context of production and logistics simulations, finding robust configurations is often a critical issue when simulation models are used for planning manufacturing and logistics systems. Robustness means setting the controllable parameters in such a way that variance in the noise has a minimal effect on a given output parameter. Variation through noise can emerge from various sources, for example fluctuations in customer demand can lead to variation in the mixture of jobs that are dispatched in the system. This effect can increase dramatically especially at the lower tiers of the supply chain which is commonly known as the bullwhip effect. In this paper, we present an approach for robustness investigations as an extension to our approach on knowledge discovery in simulation data, based on Taguchi's loss function in combination with multidimensional visual analytics.

The remainder of this paper is structured as follows. In section 2 we introduce the related work on data farming, knowledge discovery, and robustness analysis. Section 3 discusses the concept for robustness analysis in manufacturing simulations. A case study in Section 4 demonstrates the benefits of this approach followed by concluding remarks and a discussion of future work in section 5.

2 RELATED WORK

2.1 Data Farming and Knowledge Discovery in Simulation Data

Data farming describes a methodology for using a simulation model as data generator and efficient experimental design to maximize data yield and therefore information gain (Elmegreen, Sanchez, and Szalay 2014; Horne and Meyer 2005) The farming metaphor describes how the data output yield can be optimized by experimental design like a farmer that cultivates his land to maximize his crop yield (Sanchez 2014). New approaches in design of simulation experiments manage the balance between broad scale parameter combination and variation on the one hand and manageable data volume abandoning inefficient n^k design patterns (Kleijnen et al. 2005).

Based on that, Feldkamp et al. (2015a) developed a method for finding hidden patterns and relations in large quantities of simulation output data based on broad scale experimental design and visual aided analysis called knowledge discovery in simulation data. Based on using data farming concepts for covering a large bandwidth of input factor values and therefore possible model behavior, data mining methods are applied onto this quantity of data. Knowledge can be gained through visual representations of data mining results combined with visualization of input/output relations. The actual data analysis of the generated output data and its relation to simulation input data is built around interactive visual inspection. Visualization in general is an important tool when an interpretation of data is required (Thomas and Cook 2005) and therefore dedicated techniques are commonly applied in almost any simulation study. Common visualization techniques applied in the context of discrete event simulations are animations, time-plots of outputs and graphs of certain performance indicators in a confidence interval obtained from replicating runs (Law 2014).

The approach we presented in previous papers goes beyond those commonly applied techniques by creating visualizations as the key foundation in the simulation data analysis process (Feldkamp et al. 2016; Feldkamp, Bergmann, and Strassburger 2015b). This is based on the research area called visual analytics. 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., processing huge amounts of data, with those of humans, especially 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). Unlike traditional simulation data visualization, data preprocessing is essential. In recent work we showed the application of clustering algorithms on simulation data (Feldkamp, Bergmann, and Strassburger 2015b), but other data mining tools like regression analysis are also applicable (Kallfass and Schlaak 2012). In this paper, we incorporate robustness measures for data analysis based on loss functions. In the next subsection, we give a brief review on the technical background of this topic.

2.2 Taguchi's Loss Function for Finding Robust Configurations

Genichi Taguchi, who originally came from a quality engineering background, developed a methodology to assess decision alternatives not only based on their outcome value, but also on the variability around that outcome against noise. Put another way, the best system or process configuration might not always be the one with the best resulting mean, but is the one that is most robust against variation in the noise. This is derived from a viewpoint of the relationship of factors and the underlying product or process that is shown in Figure **1**. Noise factors are considered as tolerances that cannot be controlled, but can cause a variability in the process that leads to a deviation between the target value m and the actual response f. Through an optimal setting of control factors, the variance in the response can be minimized.

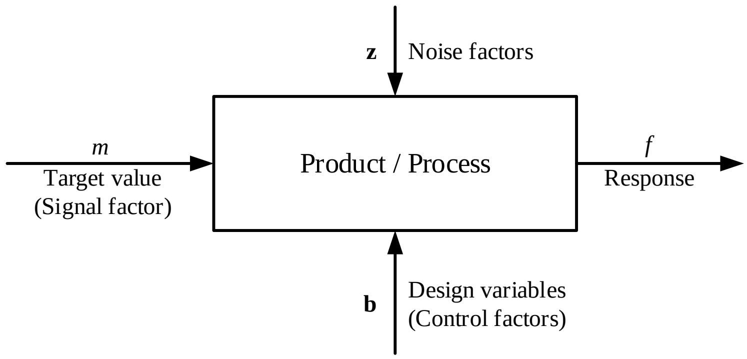

Figure 1: Block diagram of a product/process (Park et al. 2006).

Taguchi created formulas that calculate the quality loss caused by deviation from a desired value (Taguchi 1988, 1995). This worldview is a way of variance reduction. In a broader sense it can be seen as a data compression measure to assess the effect of noise on the reliability of the underlying system (Ben-Gal 2005). Taguchi's work on robustness for quality engineering had a great impact and also a lot of controversies among statisticians (Box 1988; Nair et al. 1992). Until today it is commonly applied in various applications like engineering, electric power or biotechnology (Konduk and Ucisik 1999; Rao et al. 2008; Song et al. 2017).

Table 1 shows three types of loss functions. Here, $\bar{L}$ is the average loss for a given system configuration over all noise configurations, k is a fixed constant called the *quality loss coefficient* that is used to monetarize the quality loss, $\bar{y}$ and σ^2 represent the mean and variance of an output value for each system configuration (pure y representing the output for each experiment). Depending on the characteristics of the desired output parameter, different loss functions have to be applied. The nominal-the-best loss function aims to reduce the variability around a desired output target value τ and therefore sanctions output values above and below this target, for example the required output voltage of an electrical circuit. The smaller-the-better function aims to minimize a given output, for example cost, stress or energy consumption. On the other hand, the larger-the-better loss function is used to maximize an output value like reliability, strength or efficiency. Real world examples from manufacturing engineering are minimizing radiation leakage from a microwave oven, or maximizing the bond strength of a weld point, respectively (Ben-Gal 2005; Phadke 1989). A more in-depth review on the subject of Taguchi method and other robust design concepts can be found in (Park et al. 2006).

Table 1: Loss functions for different purposes (Phadke 1989).

Type of Quality Loss Function	**Formula**
Nominal-the-best	$\bar{L} = k[\sigma^2 + (\bar{y} - \tau)^2]$
Smaller-the-better	$\bar{L} = k[y^2 + \sigma^2]$
Larger-the-better	$\bar{L} = k\left[\sum(1/y^2)\right]/n$

Taguchi's approach was transferred and applied from real world experiments to simulation experiments in several publications, since robustness design and analysis is a recurring subject in simulation methodology research. Sanchez outlined an approach for integrating the concept of robustness with response surface metamodeling for optimizing discrete event simulation models (Sanchez 1994; Sanchez 2000). Dellino et al. adopted a similar approach for simulation-based optimization of robustness issues using response surface methodology and kriging metamodels (Dellino, Kleijnen, and Meloni 2009). While metamodels usually aim to minimize experimental effort, Horne et al. already emphasized the usability of Taguchis loss function in alignment with the concept of data farming for finding robust solutions in warfare simulations (Horne et al. 2014). In our approach, we aim to combine robustness analysis based on Taguchi's loss functions with large scale experiment design and visually aided knowledge discovery methods for manufacturing simulations. In the next section, we therefore provide a basic concept for this purpose.

3 CONCEPT FOR ROBUSTNESS ANALYSIS IN MANUFACTURING SIMULATIONS

Making manufacturing systems robust against variances in the product mixture is an recurring and important issue. To solve this problem, we incorporated the concept of process robustness outlined in the previous section with our proposed data farming and knowledge discovery approach using broad scale experimentation design and data mining algorithms. We assume that a manufacturing system needs to be robust in multiple performance parameters. Since our concept incorporates the execution of a big number of simulation runs with a big number of possible input factor value combinations, it is likely that those desired configurations exist in the simulation data base. The main challenges is to find those configurations and to investigate which input factor values lead to the desired robustness, therefore the analysis of data is an essential part and is of critical importance.

In reference to the existing knowledge discovery in simulation data process, the first step is the definition of factors, hence, one must define factors that are assumed to affect the output of the system. For a robustness evaluation, it is furthermore necessary to classify factors into decision and noise factors. The decision factors are those that would be controllable to some extend in a real world system. Each combination of decision factor values is considered as a system configuration. The noise factors are not controllable and are the source of variation in the system (besides optional stochastic effects). In our case, we consider the incoming jobs and therefore the mixture of products in the manufacturing system as the noise as to which we want the manufacturing system to be robust against.

In the second step, we define two independent experiment design tables for each of the factor classes. For the decision factors, the common experiment design methods used in data farming research can be utilized, for example the *nearly orthogonal latin hypercube (NOLH)*, which is much more efficient than a default n^k-design (Vieira et al. 2011). The design of the product mixture on the other hand is much more challenging. Although the number of experiments for the product mix does not grow exponentially with n^k since factors are not independent from each other, it still grows factorially with k for an increasing number of products in the mixture and the resolution of step size from 0% to 100% (Ledi et al. 2013). In a full factorial design with five products, we would have $(5+(100/5))!/((100/5)!\,(5-1)!) = 265.650$ experiments given a 5% percent step size. Therefore, full factorial coverage of product mixes obviously is not possible for increasing k. To reduce the number of experiments, a simple approach is to use a data farming design method like *NOLH* and normalize the sum of each row to one. However, this method must be used with caution, since it can reduce the desired properties of an experimental design like balancing the distribution of factor values and the absence of correlation between input factors (orthogonality). Implementing constraints in space filling designs like latin hypercube is an tremendously challenging task and ongoing research topic (Golchi and Loeppky 2016; Petelet et al. 2010). For our purpose of proof of concept, the prior mentioned approach is sufficient for our case study model presented in the next section.

In step three, after arranging the experimental designs, both designs are crossed (crossed design), resulting in a final experimental design of every combination of system and noise configurations. The resulting experimental design can become very large, so a parallel execution of experiments and high performance computing is needed. The advantage of this approach is that after the experiments are conducted, we can arrange the result data into a matrix-like table that shows how each system configuration performs for each noise factor configuration or product mixture, respectively. Figure **2** left side shows this approach schematically.

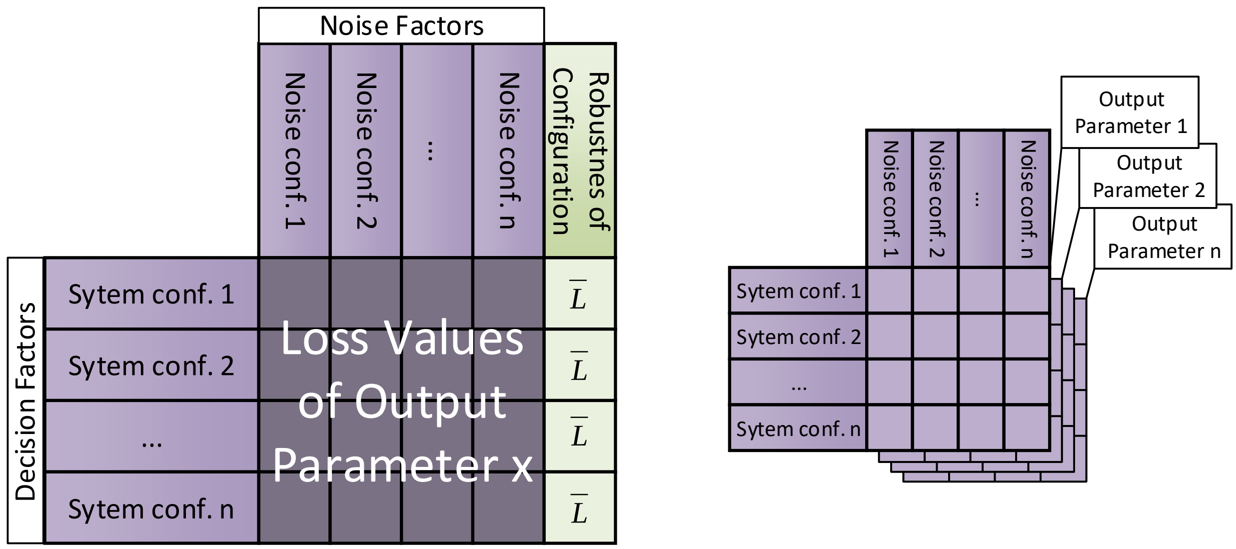

Figure 2: Matrix of crossed arrays for robustness evaluation.

Each cell of the matrix represents one simulation experiment. Cells can then be filled with their corresponding robustness value of a selected output parameter x according to the chosen loss function. If multiple replications have been conducted, either the columns have to be expanded with the numbers of replications or each cell is filled with the average loss over all replications from one experiment. For each row, we can determine the average loss representing the robustness of each system configuration. The robustness against individual noise configurations may also be interesting to investigate. Additionally, if we analyze the matrix vertically, we can also evaluate the possibility of each noise configuration to create variance in the system's output. In a complex manufacturing system, we assume that there are several output measures of interest that need to be robust against external variance. Therefore, the configuration matrix is actually multi-dimensional, which is shown in Figure **2** on the right side. In the focus of interest is finding configurations that are robust among all of the selected output parameters. In a simple approach, multiple robustness parameters can be extended with weights and summed up within a simple *use-value analysis (UVA)* that aggregates the robustness values of all output parameters to a single number. However, we do not recommend this approach since it has many disadvantages and does not justify the requirements for analyzing complex systems. For example, UVA lacks any form of sensitivity and setting weights correctly can be very difficult. For a more in-depth-analysis, we propose an interactive visually guided analysis supported by data mining, which we already demonstrated to be very suitable for the inspection of multidimensional simulation result data in our recent work (Feldkamp et al. 2016).

The main advantage of this approach is being able to find the configurations that offer the most beneficial tradeoff between output parameters in a way that a simple UVA cannot provide. For this purpose, we developed a two-step process, that is built around visually aided analysis: First, grouping system configurations into classes of similar robustness groups using classification algorithms. These algorithms can label the simulation experiments according to their class affiliation. Simulation experiments in the same class belong to system configurations that have similar values in their corresponding robustness dimensions. In the next step we use supervised learning algorithms that can train data models from that data according to the class labels and provide insights on which input factor values leads to distinct classes, preferably those with decent robustness values. Knowing how to set input factor values accordingly to get to a desired class label allows conclusions on how to make the system robust.

In the next section, we demonstrate the process of finding robust configurations through data farming by presenting a prototypical case study.

4 CASE STUDY

4.1 Model Description and Design of Experiments

For a proof of concept, we developed a simulation model of an assembly line that was implemented in Siemens Plant Simulation. Figure **3** shows a 2D and 3D-layout of this model. Here, five different part types are loaded onto workpiece carriers that are transported on a conveyor. Parts are both automatically processed on assembly stations and manually handled on up to five workplaces. At the end of the line, there is a manual quality inspection before parts get unloaded from their carrier and leave the system. The mixture of parts can vary, but arriving parts are kept in a buffer until they are cleared to get mounted on a free carrier. Some stochastic effects arise through machine reliability and a small proportion of parts that fail the quality assurance and are rescheduled for workplace manufacturing.

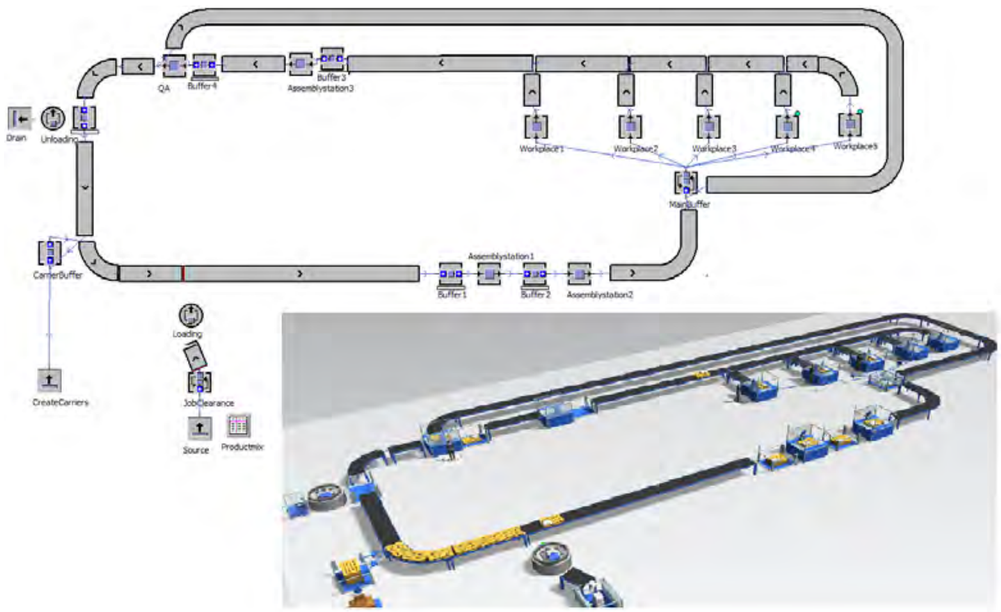

Figure 3: 2D and 3D view of the assembly line model.

The goal here is to make the line robust against the product mixture in multiple output parameters, namely throughput, workplace and carrier utilization, and job cycle time. A given number of decision factors can therefore be derived from the model's specification, which are shown in Table 2.

Table 2: Decision factors for the simulation experiments.

Factor name	**Scale**	**Description**	**Margins**
LoadingTime	Continuous	Duration for mounting parts	10-60s
UnloadingTime	Continuous	Duration for unmounting parts	10-60s
ArrivalTime	Continuous	Interval time for job clearance	100-300s
ClearanceStrategy	Categorical	Sorting strategy for jobs {fifo, lot size of: 5/10/unlimited}	1-4
BufferXCap	Discrete	Capacity for buffers (one factor for each)	1-100
#Workplaces	Discrete	Number of manual assembly work places	1-4
#Carriers	Discrete	Number of work piece carriers	1-100
WP_ProcTimeVar	Continuous	Allowed tolerance for workplace process time	100-300s
QA_ProcTimeVar	Continuous	Allowed tolerance for QA process time	100-300s

From these decision factors we derived 512 experiments in a NOLH-design. Additionally we created 40 different configurations for the product mixture from a prebuilt Latin Hypercube design. Finally both designs were crossed with each other and replicated five times. This resulted in 102.400 simulation runs. The final experiment design has been split into multiple files in order to be distributed onto ten machines. Result data was written into flat CSV-files and collected and aggregated through a dedicated. Computation of data including data mining and visualizations was performed in MATLAB and R.

4.2 Discussion of Results

As mentioned in Section 3, arranging system and noise configurations in a matrix-like table creates a profound overview on how each system configuration performs regarding any given output parameter. For an intuitive visual review, a heat map based on this matrix has been created as shown in Figure **4**. This heat map shows exemplarily the loss values for the output parameter *throughput*. The system configurations (matrix rows) have been sorted ascendingly by their total loss, so the most robust system configurations are on top. This heat map can be zoomed in and out in order to interactively review individual system and noise configurations.

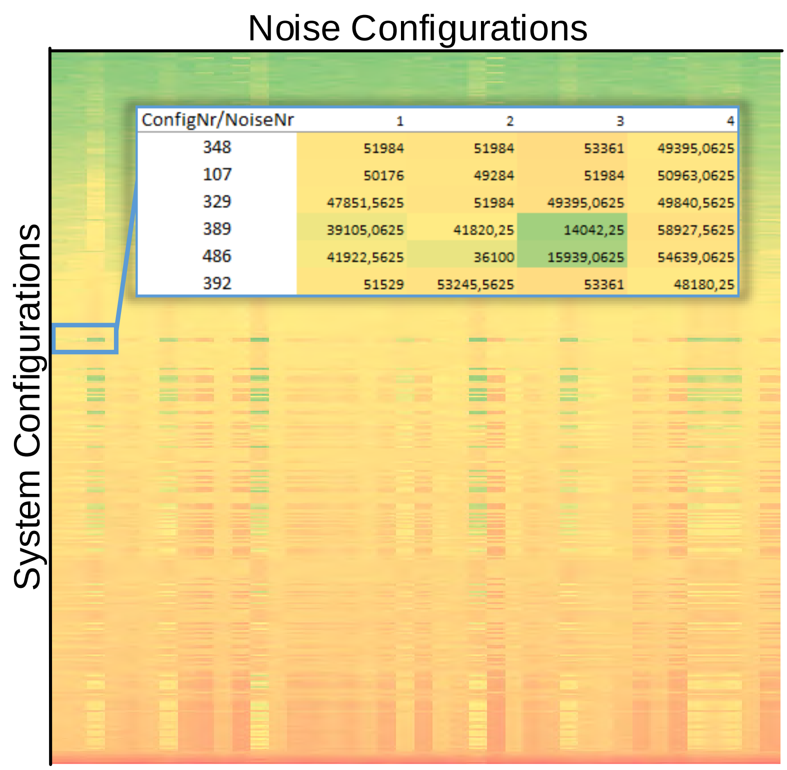

Figure 4: Heat map of loss for output parameter *Throughput* (green: small, red big loss).

In our prototypical production line, we assume that not only the throughput, but also other output parameters need to be robust against variation in the product mixture. The throughput should be nominal against 550 per day (*throughput_550*), the average cycle time should be minimized (*avg.cycleTime*) and the utilization of workplaces (*WP_Util*) and carriers (*carrier_Util*) should both be maximized. Therefore, all three different loss functions shown in Table 1 of Section 2.2 have been used here. Table 3 shows the total loss of each output dimension for the first five system configurations. It shows that the loss of each dimension is weakly correlated, so reviewing configurations and finding those that satisfy the desired robustness in all dimensions is difficult.

Table 3: Total loss of system configuration 1 - 5 for four different dimensions.

System Conf.	Loss(Throughput_550)	Loss(Avg.CycleTime)	Loss (WP_Util.)	Loss(Carrier_Util)
1	43806	1121391459	7,74	1,0045454
2	25523	614664824,7	2,33	1,0104155
3	7899	218944083,8	1,05	1,0053417
4	88430	94480746,12	1,01	1,0337554
5	201925	36723420,57	7,43	1,1315381

To support the review of system configurations, we used a clustering algorithm that classifies the simulation experiments according to their configuration's loss. This means simulation experiments in the same cluster are very similar regarding their robustness in the four selected output parameters. Figure **5** shows the result of the clustering. Simulation experiments have been grouped into ten clusters, which are indicated through different colors. The brown colored cluster (Cluster 2) consists of those simulation experiments that have the most suitable robustness among all selected dimensions. In all of the sub diagrams in Figure **5**, the loss in cluster two is rather small. Though some clusters have even better loss values in some dimensions, cluster two serves the best tradeoff among all dimensions and is therefore the target cluster for further inspection.

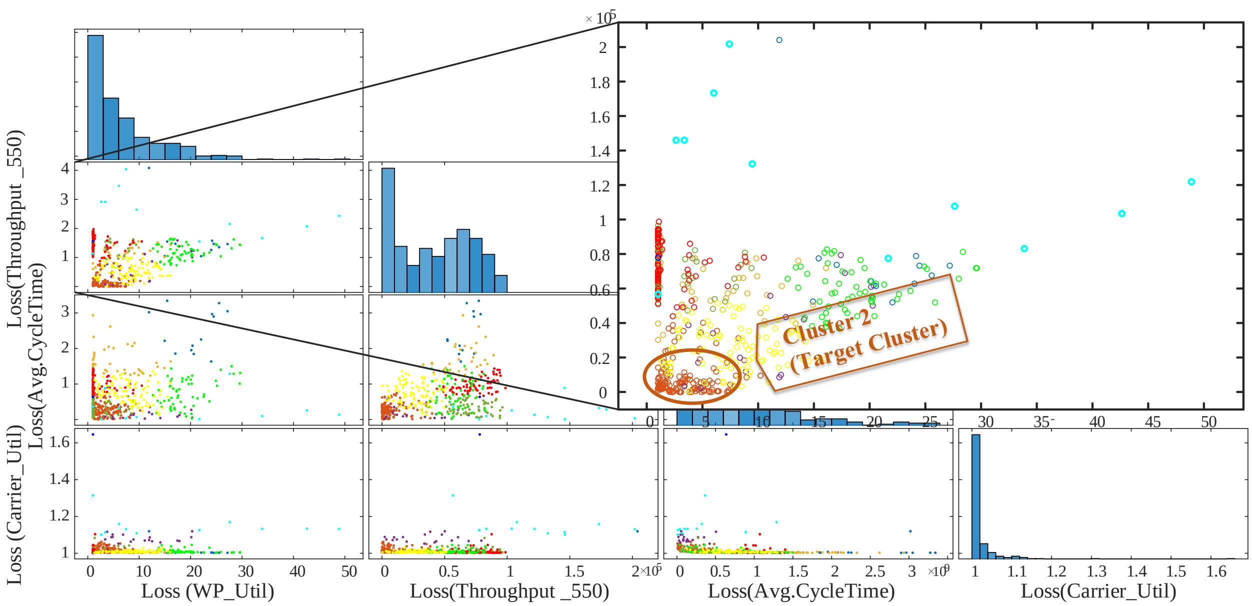

Figure 5: Clustering of output dimensions.

In next step of the analysis, we need to inspect the corresponding input factor values of simulation experiments in each cluster, especially in the target cluster. Input factor values that are very dominant in a cluster indicate a causal relationship between factors values and cluster allocation. For selected input factors that are supposed to be most influential to the loss as predicted by a correlation analysis, we investigate their distribution among each distinct cluster in a radar plot, which can be seen in Figure **6**. The plots in this figure mark the median and quartiles for each parameter. In between the quartiles are 50% of all observations or simulation experiments to be precise. So if quartiles lay close together, the corresponding parameter value is dominant in the given cluster. On the other hand, if quartiles are very broad so that a factor value is rather equally distributed among a cluster, the effects of this factor to a certain cluster allocation is presumed to be small. In our given target cluster (Cluster 2), one can see that especially the factors *ArrivalTime* and *#Workplaces* have distinct values that are dominant. Unfortunately, one cannot read off specific factor values from these radar plots. Also, we cannot draw any conclusions whether the allocation of a simulation run to cluster 2 depends on specific combinations of factor values.

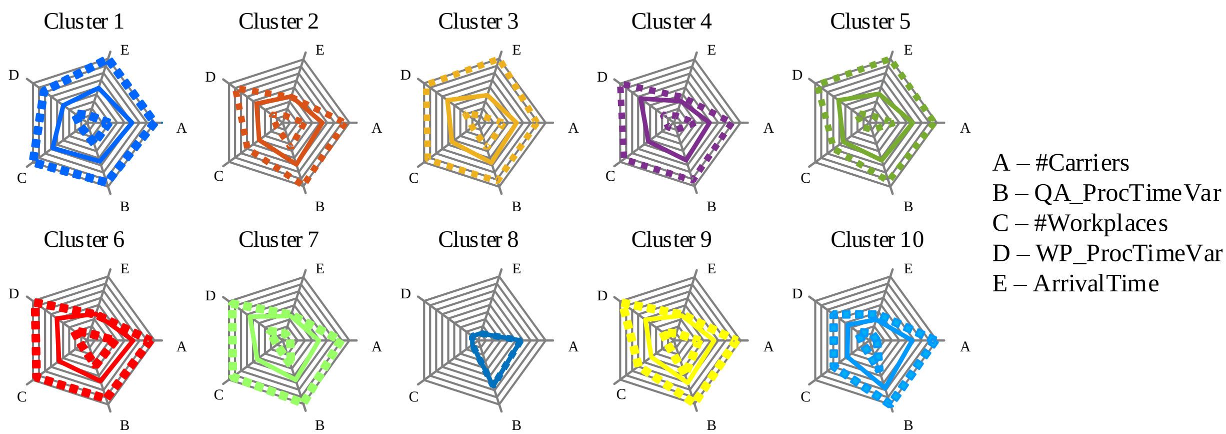

Figure 6: Radar plots of selected input factors for each cluster.

For a further inspection of the specific numbers and factor value relations, we therefore trained a binary decision tree in order to build a model that can map the relation between input factors and clusters in detail. The nodes of the tree represent a specific input factor value, the leafs or classes of the tree represent the clusters. Hence, each tree branch represents an if-then-rule that describes how to get to a certain cluster. A visualization of the tree is shown in Figure **7**.

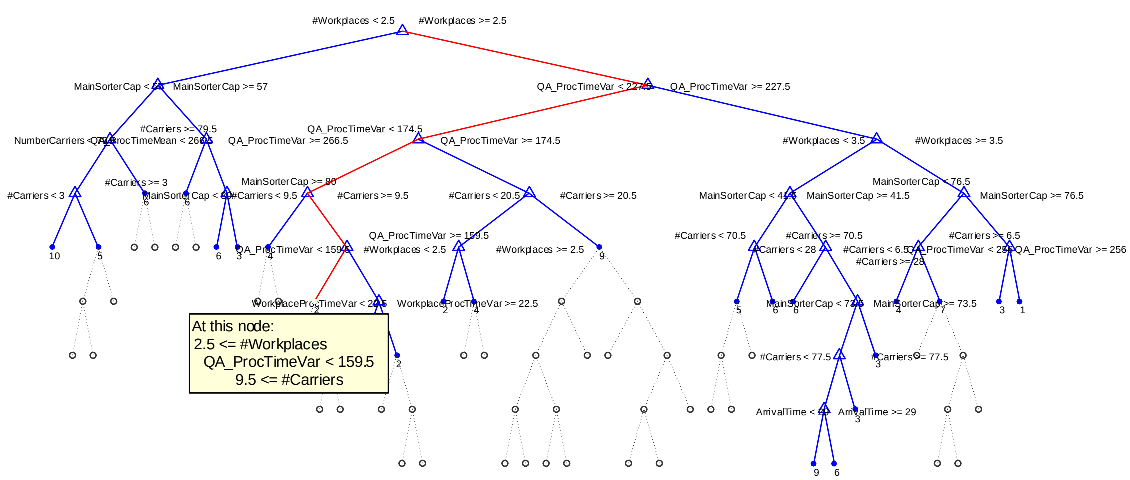

Figure 7: Visualization of the decision tree model.

The decision tree was fitted to the full data set. Overfitting issues can be ignored because the tree model is not going to be used for classifying unknown and unlabeled data, because all experiments have been already carried out at broad scale. The tree can be interactively traversed top down and each branch can be reviewed. The higher the position of node in the tree, the more importance it has for the split decision's entropy. For example, the number of workplaces, or having three or more workplaces, respectively, is the most important factor value for cluster allocation and therefore system robustness. We highlighted a specific tree branch that leads to the target cluster: 'At least three workplaces and 10 work pieces carriers, and the allowed tolerance in the process time of the quality assurance station has to be less than 160 seconds' is a configuration that leads to a system that is robust against variance in the product mixture in all desired output parameters.

To validate our findings, we created a query implementing the rule specified above and applied it on all experiments in the simulation database. Figure **8** shows mean-variance plots exemplarily for the output parameters *throughput* and *average cycle time* on all system configurations. Those configurations, that match the queried experiments are highlighted colorfully. In some configurations, very severe variance can be found, which means that the system is very sensitive to variation in the product mix. Therefore not relying on the mean only but considering the variance of configurations is indeed very important. The implemented rule derived from the decision tree fits our proposed requirements on the output dimensions very good, with a balanced tradeoff between them

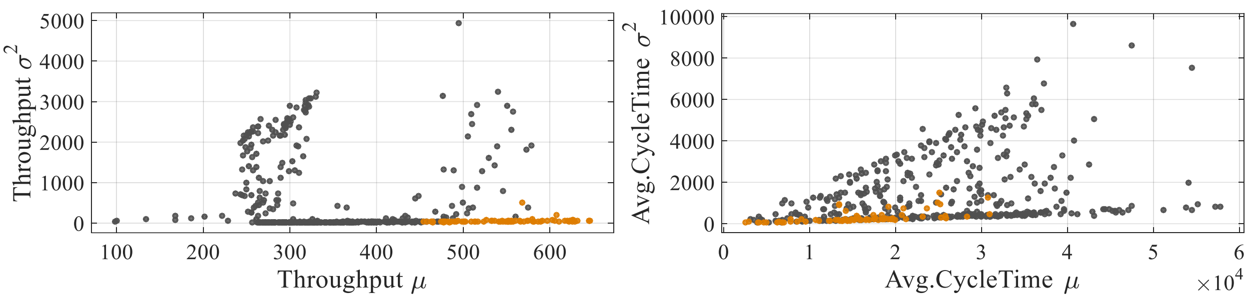

Figure 8: Mean/variance plots of selected output parameters.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrated how a visual analytics based knowledge discovery process for manufacturing simulation robustness analysis can be performed. This approach brings an additional viewpoint into the existing robustness analysis research. Having the possibility to analyze large amount of data yields a better insight into the behavior on the system. Traditional approaches for robustness analysis focus on only one dimension of robustness, so our approach allows a new level of sensitivity because we are able to investigate multiple dimensions simultaneously and evaluate the relations between them. Furthermore, our approach might be more appealing to people who are not “simulation experts”, since an interactively designed and visually aided analysis process is more user-friendly. 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 data farming and visual analytics approaches to become commonplace, a tight integration of the demonstrated methods with commonly applied simulation packages is needed. Another challenge is the experimental design of product mixtures. Creating design methods that are efficient and still deliver features like orthogonality and balancing is still an ongoing research topic.

REFERENCES

Ben-Gal, I. 2005. On the Use of Data Compression Measures to Analyze Robust Designs. *IEEE Transactions on Reliability* 54, 381–388. Piscataway, NJ: IEEE Inc.

Box, G. 1988. Signal-to-Noise Ratios, Performance Criteria, and Transformations. *Technometrics* 30, 1-17.

Dellino, G., J. P. C. Kleijnen, and C. Meloni. 2009. Robust Simulation-Optimization using Metamodels. In *Proceedings of the 2009 Winter Simulation Conference,* ed. M. D. Rossetti, R. R. Hill, B. Johansson, Dunkin A., and R. G. Ingalls, 540–550. Piscataway, NJ: IEEE Inc.

Elmegreen, B. G., S. M. Sanchez, and A. S. Szalay. 2014. The Future of Computerized Decision Making. In *Proceedings of the 2014 Winter Simulation Conference,* ed. A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 943–949. Piscataway, NJ: IEEE Inc.

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. 2015a. Knowledge Discovery in Manufacturing Simulations. In *Proceedings of the 2015 ACM SIGSIM PADS Conference,* 3–12. New York, NY: ACM.

Feldkamp, N., S. Bergmann, and S. Strassburger. 2015b. Visual Analytics of Manufacturing Simulation Data. In *Proceedings of the 2015 Winter Simulation Conference,* ed. L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 779–790. Piscataway, NJ: IEEE Inc.

Feldkamp, N., S. Bergmann, S. Strassburger, and T. Schulze. 2016. Knowledge Discovery in Simulation Data: A Case Study of a Gold Mining Facility. In *Proceedings of the 2016 Winter Simulation Conference,* ed. T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 1607–1618. Piscataway, NJ: IEEE Inc.

Golchi, S., and J. L. Loeppky. 2016. Space Filing Designs for Constrained Domains. In *International Conference on Design of Experiments,* ed. M. Aggarwal and O. E. George, 1–28.

Horne, G., B. Åkesson, T. Meyer, and S. Anderson. 2014. *Data farming in support of NATO. Final Report of Task Group MSG-088,* Neuilly-sur-Seine Cedex: North Atlantic Treaty Organisation.

Horne, G. E., and T. E. Meyer. 2005. Data Farming: Discovering Surprise. In *Proceedings of the 2005 Winter Simulation Conference,* ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 1082–1087. Piscataway, NJ: IEEE Inc.

Kallfass, D., and T. Schlaak. 2012. NATO MSG-088 Case Study Results to demonstrate the Benefit of using Data Farming for Military Decision support. In *Proceedings of the 2012 Winter Simulation Conference,* ed. C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 1–12. Piscataway, NJ: IEEE Inc.

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,* ed. 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.

Konduk, B. A., and A. H. Ucisik. 1999. Determination of Primary Parameters relevant to the Adequacy of Haemodialysis through Taguchi Method. In *Proceedings of the first Joint BMES/EMBS Conference,* ed. S. Blanchard, E. Eckstein, and J. Fouke, 627.

Law, A. M. 2014. *Simulation Modeling and Analysis,* 5th ed., New York, NY: McGraw Hill Book Co.

Ledi, T., P. Spagon, E. del Castillo, T. Moore, S. Hartley, and A. Hurwitz. 2013. e-Handbook of Statistical Methods: 5. Process Improvement. http://www.itl.nist.gov/div898/handbook/pri/pri.htm.

Nair, V. N., B. Abraham, J. MacKay, J. A. Nelder, G. Box, M. S. Phadke, R. N. Kacker, J. Sacks, W. J. Welch, T. J. Lorenzen, A. C. Shoemaker, K. L. Tsui, J. M. Lucas, S. Taguchi, R. H. Myers, G. G. Vining, and C. F. J. Wu. 1992. Taguchi's Parameter Design: A Panel Discussion. *Technometrics* 34:127.

Park, G.-J., T.-H. Lee, K. H. Lee, and K.-H. Hwang. 2006. Robust Design: An Overview. *AIAA Journal* 44:181–191.

Petelet, M., B. Looss, O. Asserin, and A. Loredo. 2010. Latin hypercube sampling with inequality constraints. *AStA Advances in Statistical Analysis* 94:325–339.

Phadke, M. S. 1989. *Quality engineering using robust design,* Englewood Cliffs, NJ: Prentice Hall.

Rao, R. S., C. G. Kumar, R. S. Prakasham, and P. J. Hobbs. 2008. The Taguchi methodology as a statistical tool for biotechnological applications: a critical appraisal. *Biotechnology journal* 3:510–523.

Sanchez, S. M. 1994. A Robust Design Tutorial. In *Proceedings of the 1994 Winter Simulation Conference*, ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 106–113. Piscataway, NJ: IEEE Inc.

Sanchez, S. M. 2000. Robust design: seeking the best of all possible worlds. In *Proceedings of the 2000 Winter Simulation Conference,* ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 69–76. Piscataway, NJ: IEEE Inc.

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

Song, J., F. Dong, J. Zhao, S. Lu, S. Dou, and H. Wang. 2017. Optimal design of permanent magnet linear synchronous motors based on Taguchi method. *IET Electric Power Applications* 11, 41–48.

Taguchi, G. 1988. *System of Experimental Design,* 3rd ed., White Plains, NY: Unipub.

Taguchi, G. 1995. Quality engineering (Taguchi methods) for the development of electronic circuit technology. *IEEE Transactions on Reliability* 44, 225–229.

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

Vieira, H., S. M. Sanchez, K. H. Kienitz, and M. C. N. Belderrain. 2011. Improved efficient, nearly orthogonal, nearly balanced mixed designs. In *Proceedings of the 2011 Winter Simulation Conference*, ed. S. Jain, R. Creasey, J. Himmelspach, K. P. White, and M. C. Fu, 3600–3611. Piscataway, NJ: IEEE Inc.

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. His research interests include distributed simulation, automatic simulation model generation, and general interoperability topics within the digital factory context. He is also an active member of the Simulation Interoperability Standards Organization (SISO). His email is steffen.strassburger@tu-ilmenau.de.

THOMAS SCHULZE is a professor in the School of Computer Science at the Otto-von-Guericke-University, Magdeburg, Germany. He received the Ph.D. degree in civil engineering in 1979 and his habil. Degree for computer science in 1991 from the University of Magdeburg. His research interests include modeling methodology, manufacturing simulation, distributed simulation with HLA and visualization. He is an active member in the ASIM, the German organization of simulation. His email address is Thomas.Schulze@ovgu.de.