Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

USING ACCURACY MEASUREMENTS TO EVALUATE SIMULATION MODEL SIMPLIFICATION

Igor Stogniy

Institute of Applied Computer Science
Technische Universität Dresden
Dresden, 01062, GERMANY

Wolfgang Scholl

Infineon Technologies Dresden
Königsbrücker Straße 180
Dresden, 01099, GERMANY

ABSTRACT

Infineon Technologies Dresden has long used discrete event simulation to optimize production planning for its fully automated front end manufacturing lines. There are needs to reduce maintenance efforts, to increase transparency for validation and verification, and to improve flexibility for the fast simulation of scenarios with a focus on qualitative statements. Less detailed models will be utilized where components could be omitted. This paper considers a simplification of the process flows through operation substitution for constant delays. The main idea is to consider the accuracy measurements used to evaluate simplification. It is shown that standard accuracy measurements (e.g. mean absolute error, correlation coefficient, etc.) produce rather poor performance. It is suggested instead to use measurements based on lot cycle time distributions (e.g. goodness-of-fit tests). Nine types of simplification sieve functions were analyzed, with analyses based on the MIMAC dataset 5 model.

1 INTRODUCTION

Van der Zee (2019) presented an overview of the simulation model simplification and noted that simplification “is still very much a green field”. Mönch et al. (2018) provided several other papers relevant to simplification for supply chain simulation needs and links on existing testbeds. Unfortunately, research in this area often describes evaluation measurements which are only suitable for the rather simple examples which the authors considered in their works. More appropriate methods do exist, e.g. goodness-of-fit tests (Law 2015). However, these methods are utilized less often, most likely due to historical traditions in simulation, where the main focus is placed on precise estimation of the mean and quantile values of the simulation model results in comparison with reality (Alexopoulos and Kelton 2017), but not on the comparison of the dozens of simulation models with each other.

The research presented in this paper seeks to compare 83 simplified models with a base model and with each other. This is done as the problem of gradual simplification of the process flows through operation substitution for constant delays is being considered. There is also a need to compare the different experiment series of 83 simplified models with each other in order to identify the most effective simplification method (see Table 2). The main idea of the experiments presented here is to investigate how many tool sets could be substituted by delays, and with which heuristic (sieve function), without a big loss of accuracy of the simplified model. Here, the accuracy measurements are the main focus.

As a continuation of previous work (Stogniy and Scholl 2019), the current paper makes the following contributions: 1) gradual simplification with substitution of a one tool set per step (this allows oscillations in the accuracy measurements to be observed); 2) the consideration not only of tool set, but also process step based delays (providing an improved simplified model); 3) considerations of 9 (not one and a half) sieve functions with a focus on standard accuracy measurements (mean absolute error, correlation coefficient, etc.) as opposed to simply comparing the results of sieve functions, an approach that highlighted, in contrast to other findings in the literature, the unsuitability of the measurements for simplification tasks; and 4) a dramatically increased data volume for which R (R Core Team 2019) was used for calculations, as opposed to the electronic spreadsheets used in previous research.

As a basis for the research, the MIMAC datasets (1997) were chosen. MIMAC dataset 5 was decided upon because it includes the largest quantity of process flows. For the fab simulation model, AutoSched AP from Applied Material, version 11.5, was used.

This paper is organized as follows. Several ideas from related works are presented in Section 2. The designs of the experiments, including sieve function, delay types, and accuracy measurements description, are discussed in Section 3. The diagrams, including measurements and their critiques, are described in Section 4. Section 5 includes summarized experimental results for sieve function comparison.

2 RELATED WORK

This research was inspired mostly by two papers: Johnson et al. (2005) and Völker and Gmilkowsky (2003). Following Johnson et al. (2005), gradual simplification was implemented in this paper, but with a much more detailed model (MIMAC dataset 5 and not a tandem-10 M/M/1 or M/M/10 model). Johnson et al. (2005) used correlation and autocorrelation to estimate the accuracy of simplified models. In the current this paper it was determined that these estimations would not be sufficient (see Secion 4.1). Völker and Gmilkowsky (2003) was found to be the most detailed paper in the field of simplification. The authors used three sieve functions to choose the operations for substitution. In comparison, the work presented in the current paper uses nine sieve functions, each with two variations. The focus was put on tool sets and not on the operation substitution, as this does not require capacity reservations. However, information about each particular operation was also used for the calculation of process step based delays. Völker and Gmilkowsky (2003) also used an estimation based on lot cycle time mean absolute error (MAE) and found “pathological” configurations when the error function is not monotonous. The research presented below also found such configurations and a logical explanation for this phenomenon is presented in Section 4.1. Moreover, it was also found that MAE is inadequate as an estimation tool for simplified models. Similar to the Völker and Gmilkowsky, an operation ratio was used here to quantify the degree of simplification.

Another consequential paper is that of Rose (1999), from which several design features were adopted: using First In First Out (FIFO) and Critical Ratio (CR) dispatching rules, considering lot cycle time distribution as an accuracy criterion, and investigating bottleneck problems. Rose (1999) did not present formal criteria to estimate differences in lot cycle distributions, but used them as illustrations. In a previous paper (Stogniy and Scholl 2019), we suggested using a Kolmogorov-Smirnov test and summarized absolute divergence (SAD). Within the previous review process, our attention was directed to a similar SAD criterion in (Ewen et al. 2017). In the current work, these and other accuracy measurements are also considered. It is also necessary to emphasize that one of the more popular statistical tests, namely, the t-test (see for example Piplani and Puah (2004)), is not suitable for the simplification sought in our case. A literature review of papers on bottlenecks produced two helpful ideas: a sieve function ζ_3 based on active period method (Roser and Nakano 2015) and using standard deviation to estimate sieve function fluctuation.

3 DESIGN OF EXPERIMENTS

For this research a special automated experimental environment was developed. It allowed for 4,000 detailed model runs to calculate warm up period and delays (2000 seeds x2: FIFO and CR dispatching rules). Subsequently, 44,820 automated experiments for accuracy measurements calculations were carried out: 3 system configurations: ε_1 – FIFO, ε_2 – CR, ε_{2b} – CR with consolidated delays at the beginning; 9 sieve functions; 2 variants of sieve function fluctuation (with and without); 2 types of delays (process step and tool set based); and 83 tool sets substitutions (5 seeds for each substitution). An experiment includes a single run of the simplified simulation model. The experiment number signifies the number of substituted tool sets (from 1 to 83).

A steady state simulation was considered. Each simulation run was 114 weeks. Welch's procedure to determine the warm up period was used (Law 2015), with the warm up being set at 10 weeks. Tool set substitution for constant but not random delays was considered. Explanations for this decision can be found in (Stogniy and Scholl 2019). In the current paper distributed delays (ε_1 and ε_2) were applied. However, because the previous research showed a significant difference for the CR rule when calculating a consolidated delay for each process flow and placing it already at the beginning, an ε_{2b} system configuration was examined as well.

3.1 Sieve functions ζ

For tool set substitution, workcenter criticality indexes were used, referred to here as sieve functions following (Völker and Gmilkowsky 2003). An order of tool sets according to a particular sieve function was created to substitute a tool set one by one. The term "bottleneck" is often used in simplification papers. However, the term "tool importance" is used in the current work to emphasize the focus being placed on identifying an important tool set when considering substitution, as opposed to simply locating bottlenecks.

The following model statistics based on weekly standard model reports were used to build sieve functions: *IDLE%/IDLE#* – the percent of time/the number of times a tool set entered the idle state; *PROC%* – the percent of time a tool set entered the processing state; BS_{AVG} – the average of batches processed (batch size); BS_{MAX} – the maximum quantity of pieces allowed in a batch; QT_{AVG} – the average time lots waited at the tool set (queue time); QL_{AVG} – the average number of pieces in front of the tool set (queue length); PT_{AVG} – the average of the lot processing time for the tool set; CT_{AVG} – the average lot cycle time for the tool set ($CT_{AVG} = PT_{AVG} + QT_{AVG}$); and CT_{SD} – standard deviation of the cycle time for the tool set ($CT^2_{SD} = PT^2_{SD} + QT^2_{SD}$).

The following sieve functions were used: $\zeta_1 = IDLE\%$; $\zeta_2 = IDLE\% + PROC\% - PROC\%(BS_{AVG} / BS_{MAX})$; $\zeta_3 = (100 - IDLE\%) / IDLE\#$; $\zeta_4 = QT_{AVG}$; $\zeta_5 = QT_{AVG} / PT_{AVG}$; $\zeta_6 = QL_{AVG}$; $\zeta_7 = QL_{AVG} / BS_{MAX}$; $\zeta_8 = CT_{SD}^{total}$; and $\zeta_9 = CT_{SD}^{total} / CT_{AVG}^{total}$. All ζ (except ζ_3) were calculated based on weekly report data of 2,000 independent experiments with detailed models. ζ_3 was calculated based on 104-week period reports of the 2,000 experiments. This was done to reduce the cases where $IDLE\# = 0$. In the case of 104-week period reports, $IDLE\# = 0$ appeared for only one tool set in about 500 of 2,000 experiments. In this case it was substituted with $IDLE\# = 1$ for the calculation of ζ_3. To calculate CT_{SD}^{total} (ζ_8), the following formula was used (the law of total variance (Fewster 2014)):

$$\mathrm{CT}_{SD}^{total} = \sqrt{\frac{\sum_{i=1}^{n}(CT_{SD})_i^2 \cdot p_i}{\sum_{i=1}^{n} p_i} + \frac{\sum_{i=1}^{n}\left((CT_{AVG})_i - CT_{AVG}^{total}\right)^2 \cdot p_i}{\sum_{i=1}^{n} p_i}},$$

where i is the report period (a week), $n = 104$ weeks, and p_i is the number of pieces which finished processing at the tool set during the i^{th} report period.

Several papers about bottlenecks (e.g. Roser and Nakano (2015)) mentioned that taking momentary bottlenecks into account is important, and that advice was followed here. Therefore sieve function fluctuations were considered. As long as ζ is calculated based on large amounts of data, a distribution of its value and calculated mean value (μ) and standard deviation (σ) is available. There are two sieve function variants: $\zeta_i = \mu$ and $\zeta_{if} = \mu \pm \sigma$ (where i = 1, 2, …, 9 and f – fluctuation). To understand the difference between the two variants and which sign ("+" or "–") should be used in the second variant, consider the following illustrative example (Figure 1). There are three tool sets (A, B, and C) with the following case to consider: "tool importance" → max, when ζ → min. This means that an ascending order of the tool sets should be created. Taking into account only mean values (μ) results in an order of B<A<C. But taking both mean and standard deviation values into consideration suggests that μ – σ be used to obtain a conservative estimation. This results in A<B<C. If the following case is considered, "tool importance" → max, when ζ → max, then μ + σ should be used.

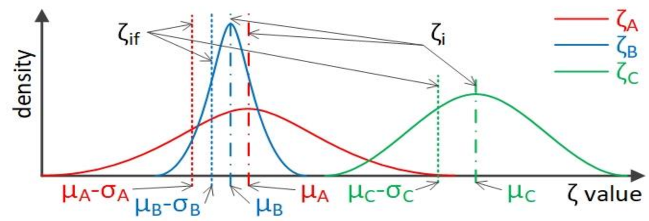

Figure 1: An illustrative example of ζ fluctuation. Two variants: ζ_i and ζ_{if}.

3.2 Delay types η

Two types of delays are considered here: process step (η_1) and tool set (η_2) based. The delays are calculated as averages based on 104 weekly reports of 2,000 independent simulation runs with detailed models. The first type is calculated based on process step reports and the second type based on tool set reports (Figure 2). A delay for a particular week is equal to the sum of average processing and average queueing time for a process step or a tool set. For system configurations FIFO (ε_1) and CR (ε_2), distributed delays were used: each delay substitutes each operation exactly at the place where this operation appears in the process flow. For the case CRb (ε_{2b}), the sum of all the delays for a process flow was calculated and placed at the beginning of the process flow.

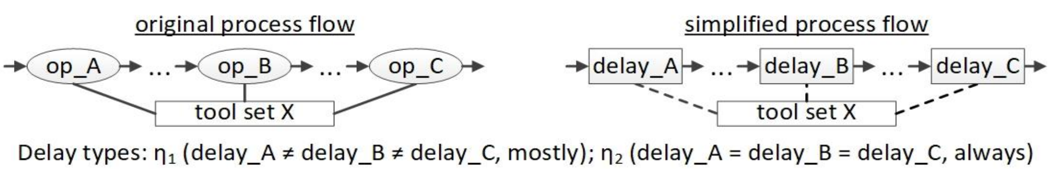

Figure 2: Delay types: η_1 – process step based; η_2 – tool set based.

Thus, for each of the nine sieve functions there are four experiment configurations: (ζ_i, η_1), (ζ_i, η_2), (ζ_{if}, η_1), and (ζ_{if}, η_2). Altogether there are 36 experiment configurations for each of the three system configurations (ε_1 – FIFO, ε_2 – CR, ε_{2b} – CRb). When referring to individual simulations within this paper, a triad (ε, ζ, η) is used to label a particular experimental configuration.

3.3 Accuracy measurements

In this paper, the following simulation run data measurements were used: simulation time; operation ratio (OP_r = 100 * substituted operations/all operations); Work-In-Process ratio (WIP_r =100 * WIP of substituted tool sets/WIP of all tool sets); Work-In-Process in the queue ratio (WIPq_r = 100 * WIP of the queue in front of substituted tool sets/ WIP of the queue in front of all tool sets). WIP values are calculated based on detailed models. Four typical accuracy measurements from simplification papers were used: lot cycle time (CT) mean absolute error (MAE), correlation coefficient, autocorrelation function, and t-test. Additionally, following the advice of a reviewer, the Wilcoxon rank sum test is also considered as an alternative to the t-test.

During a preliminary experimental analysis, the above-mentioned common accuracy measurements were found to be unsuitable (see "pathological" in Section 4.1). Therefore, based on literature analysis (Law 2015, Rose 1999, Ewen et al. 2017) it was decided to apply the following accuracy measurements based on lot CT distribution (Dowd 2018): Kolmogorov-Smirnov (KS), Anderson-Darling (AD) and Cramer-von Mises (CVM), Two Sample Test (DTS), Wasserstein distance (WASS), and Summarized Absolute Divergence (SAD). For the KS test, the standard function from the R-package "stats" was used, and for SAD a novel R-code was written. The formulas (Table 1) were taken from the package description (Dowd 2018).

Table 1: Accuracy measurements based on lot cycle time distributions.

Designation	R package / function	Formula
KS	stats/ks.test	$\max\lvert E(x) - F(x)\rvert$
AD	twosamples / ad_stat	$\sum_x \left(E(x) - F(x)\right)^2 / [G(x)(1 - G(x))]$
CVM	twosamples / cvm_stat	$\sum_x \left(E(x) - F(x)\right)^2$
DTS	twosamples / dts_stat	$\sum_x \lvert E(x) - F(x)\rvert / [G(x)(1 - G(x))]$
WASS	twosamples / wass_stat	$\sum_x \lvert E(x) - F(x)\rvert$
SAD	own code	$\sum_x \lvert PE(x) - PF(x)\rvert$

Table 1 displays the designation, source, and formulas used for calculating the statistics. In the formulas: E(x) is the empirical cumulative distribution function (ECDF) of sample 1 (detailed model); F(x) is the ECDF of sample 2 (simplified model); G(x) is the ECDF of the joint sample; PE(x) is the probability density function (PDF) of sample 1; and PF(x) is the PDF of sample 2. Because the R-package "twosamples" uses randomization to create p-values (bootstrapping), it takes significant time to produce meaningful numbers. Therefore, p-values were not calculated for AD, CVM, DTS, and WASS. For SAD it is not possible to calculate a p-value, because it is not a statistic in a classical meaning, but it does represent a suitable measurement for the purpose (see Section 4.2).

4 EXPERIMENTS

Five independent experiments with different seeds were carried out for each tool set substitution. Subsequently, averages of output parameters (simulation time, correlation coefficient, etc.) were calculated to compare the various experiments. Figure 3a illustrates the calculation for simulation time. The calculations were carried out on a server which is used for numerous tasks simultaneously. It can be assumed that fluctuations in simulation time were caused by fluctuations in server load resulting from other tasks being served at the same time.

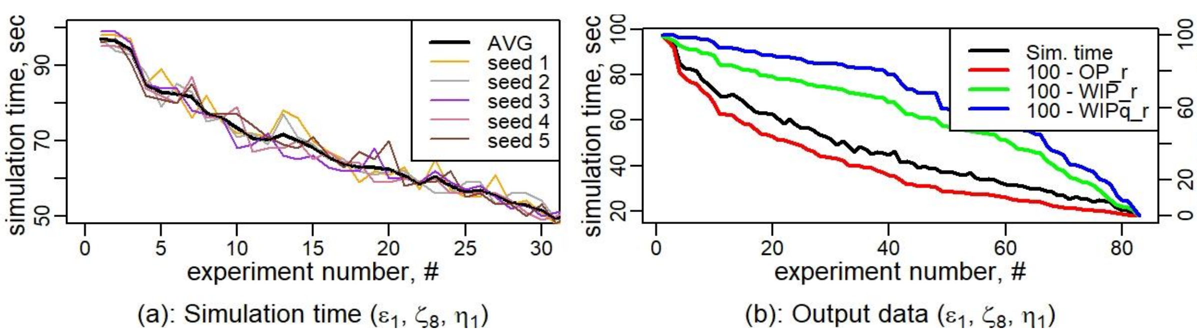

Figure 3: Simulation output.

Figure 3b demonstrates relations between simulation time, operation ratio, and WIP ratios (here and later, all curves represent the average of the five seeds). The resulting curves appear different for different experiment configurations. Often they are relatively close to each other, but for some cases, such as (ε_1, ζ_8, η_1), only operation ratio repeats the simulation time curve convexity. Moreover, Figure 3b shows that simulation time is not a monotonically decreasing function, in contrast to an operation ratio. Therefore, operation ratio was used as a measurement of simplification to compare different experiment configurations at the end of this paper (Figure 9). On the other hand, using the experiment number on the x-axis results in more intuitive and understandable diagrams. Therefore, this configuration of the x-axis is used for the following figures.

4.1 Typical accuracy measurements

A common measure of model accuracy is lot cycle time (CT) mean absolute error (MAE). Völker and Gmilkowsky (2003) used a measurement based on MAE and found some configurations which they called "pathological", where an error function does not grow monotonously, but increases at the beginning and decreases towards the end. As an illustration, consider (ε_1, ζ_1, η_1): MAE increases until experiment number 60 and falls after 70 (Figure 4a). At first it was assumed that this behavior was caused by a shift of the overall lot CT mean. Initial analysis provided confirmation of this (Figure 4b). However, a further example (ε_1, ζ_7, η_1) was found where a similar overall mean shift exists, but there is no "pathological" behavior. Therefore, lot CT standard deviation (SD) was also considered (Figure 4b). An average SD based on each process flow SD was calculated. The average SD decreases in both cases, but more sharply in the first case. Figure 4c and 4d illustrate this behavior: the exp#70 distribution for (ε_1, ζ_1, η_1) is much more narrow than for (ε_1, ζ_7, η_1), but there is a significant shift to the left at the right peak for both cases. From this it was concluded that "pathological" behavior exists when both lot CT mean shift and a strong decrease of average SD appear (ε_1, ζ_1, η_1). If there is only the shift and a mild decrease of average SD, then no "pathology" is observed (ε_1, ζ_7, η_1). Several other examples of "pathological" but less expressed behavior were observed: e.g. (ε_1, ζ_3, η_1), (ε_1, ζ_4, η_1), and (ε_1, ζ_5, η_1).

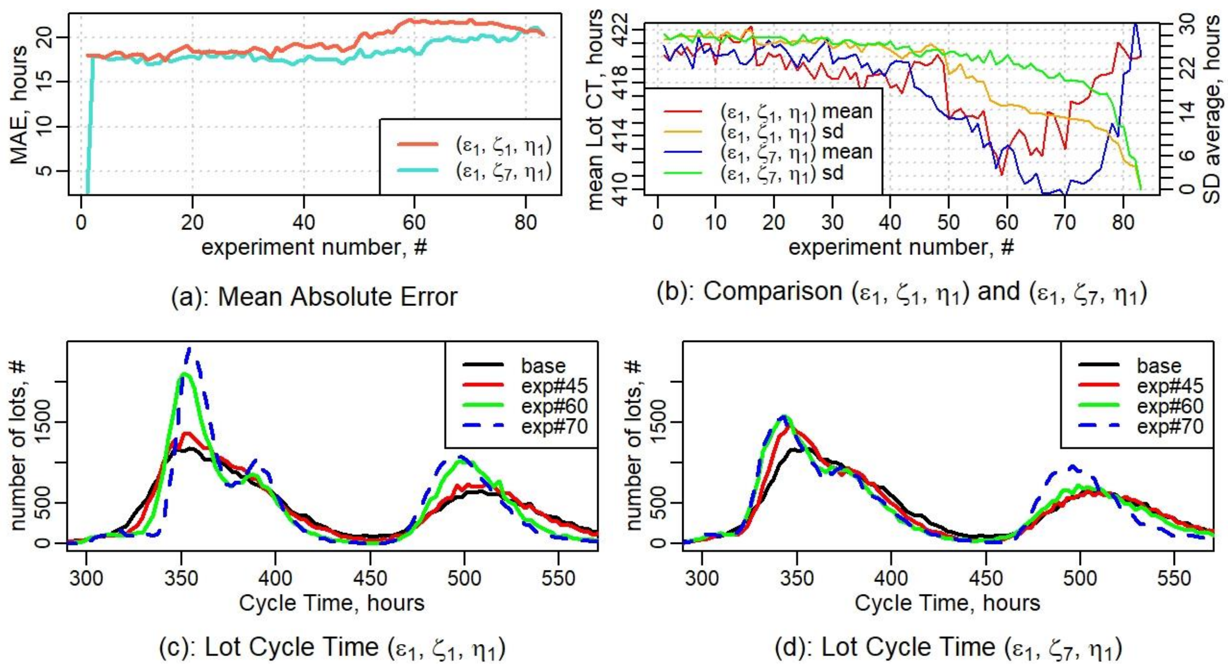

Figure 4: (a) MAE; (b) mean lot CT and average SD; (c) and (d) lot CT distributions.

Lot CT mean shift and decrease of average SD were a result of substitution for constant delays. If the SD reduction is logical, due to the use of constant and not random delays, the shifting of the lot CT mean indicates that the calculation of constant delays as a sum of averaged processing and queueing time is possibly not correct. For process step based simplification, processing time is a constant value and the impact is caused by averaged queueing time. What is more interesting is that the shift has negative ((ε_1, ζ_1, η_1) exp# 40-70) and positive ((ε_1, ζ_1, η_1) exp# 70-80) values. Identifying a more accurate way to calculate the delays would be a worthwhile endeavor for future research.

Another interesting observation about MAE is that, for some configurations (e.g. (ε_1, ζ_1, η_1) see Figure 4a), the MAE value is high (18.08) already in the first experiment, and there is minimal change (20.3) through the last experiment. This suggests that the first substitution of a tool set with IDLE% =

97.14 ((ε_1, ζ_1, η_1) exp#1) leads to significant changes in MAE (from 0 to 18.08), but further substitutions incur only small changes (from 18.08 to 20.3). For other configurations (e.g. (ε_1, ζ_7, η_1), this can be observed after the second substitution. Two possible reasons for this exist: 1) the MAE calculation is based on every single lot; 2) even small changes in the production system could lead to big lot rearrangements. We could call it a "butterfly effect". The scatter plot in Figure 5b illustrates this phenomenon for exp#1 (ε_1, ζ_1, η_1).

A second typical measurement for model accuracy is correlation coefficient. Figure 5 demonstrates how the correlation is a less effective measurement for the simplified model. We can see that, for the configuration ($\varepsilon1$, ζ_7, η_1), the correlation oscillates from exp#2 to exp#77 between 0.95 and 0.96 (Figure 5a). Moreover, the previously discussed "pathological" behavior can be observed for the configuration (ε_1, ζ_1, η_1). Thus, it can be assumed that the correlation coefficient is indeed not sensitive to the simplification. The scatter plots in Figure 5 illustrate the reason for this: despite the existence of a point cloud in the scatter plot, the cloud is distributed along the bisector. Moreover, the cloud narrows to parallel lines as the experiment number increases (compare b with e and c with f). Figure 5d shows the scatter plot for the extremely reduced model (exp#83). The correlation coefficient is 0.947 in this case.

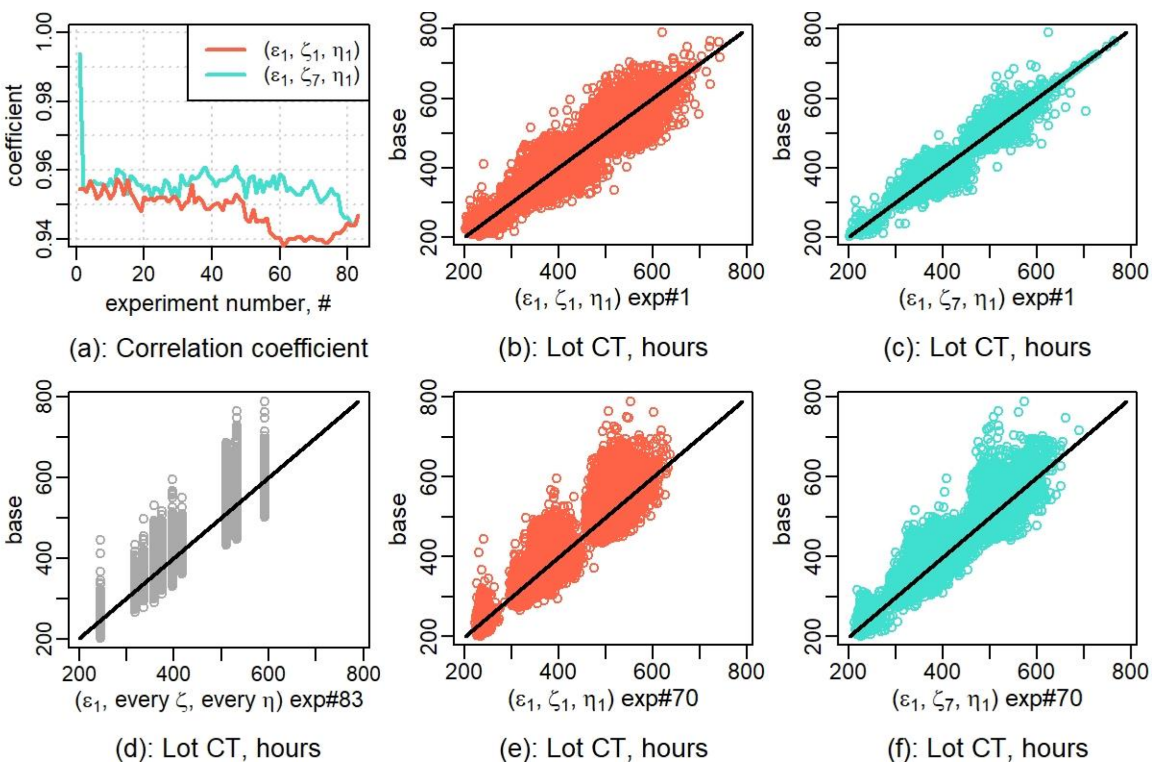

Figure 5: Correlation coefficient and scatter plots.

The third typical criteria to determine the accuracy of simplified models is the value of the autocorrelation function of lot CT for model output. Rose (1999) and Johnson et al. (2005) noted that autocorrelation should decrease with simplification. This is true when considering the very simplified model of exp#80, where all except 3 tool sets were substituted for delays (see Figure 6). However, attempting to use autocorrelation as an accuracy measurement of simplification seemed to be counterproductive, as some configurations (e.g. (ε_1, ζ_7, η_1)) saw increases instead of decreases, at least until exp#70. Moreover, three different products were chosen with different demands (product 14 – demand 35 lots/week, 12 – 5 lots/week, 1 – 2 lots/week) to show that autocorrelation varies for different products depending on the quantity of the product lots in the model.

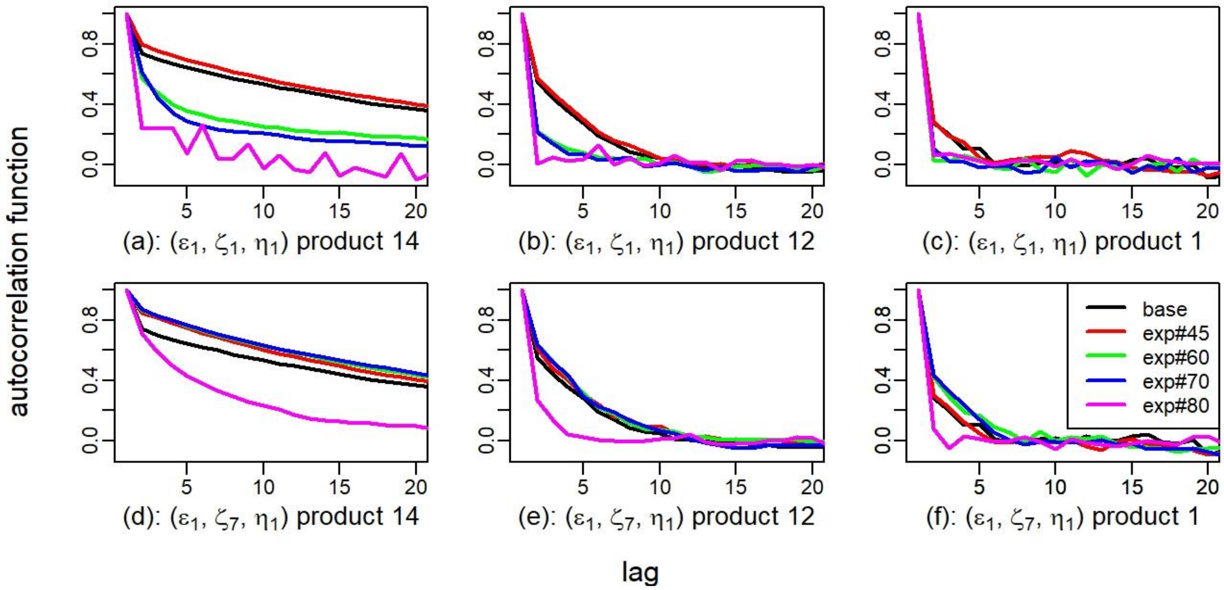

Figure 6: Autocorrelation function (base case, experiments: 45, 60, 70, 80).

The forth typical measurement is a t-test (Piplani and Puah 2004). T-test is usually recommended if there is a normal distribution. A testing of the normality using R package normtest (Gavrilov and Pusev 2014) proved unsuccessful. Nevertheless, the results of t-tests within this research (Figures 7a and 7b) may provide some useful insights.

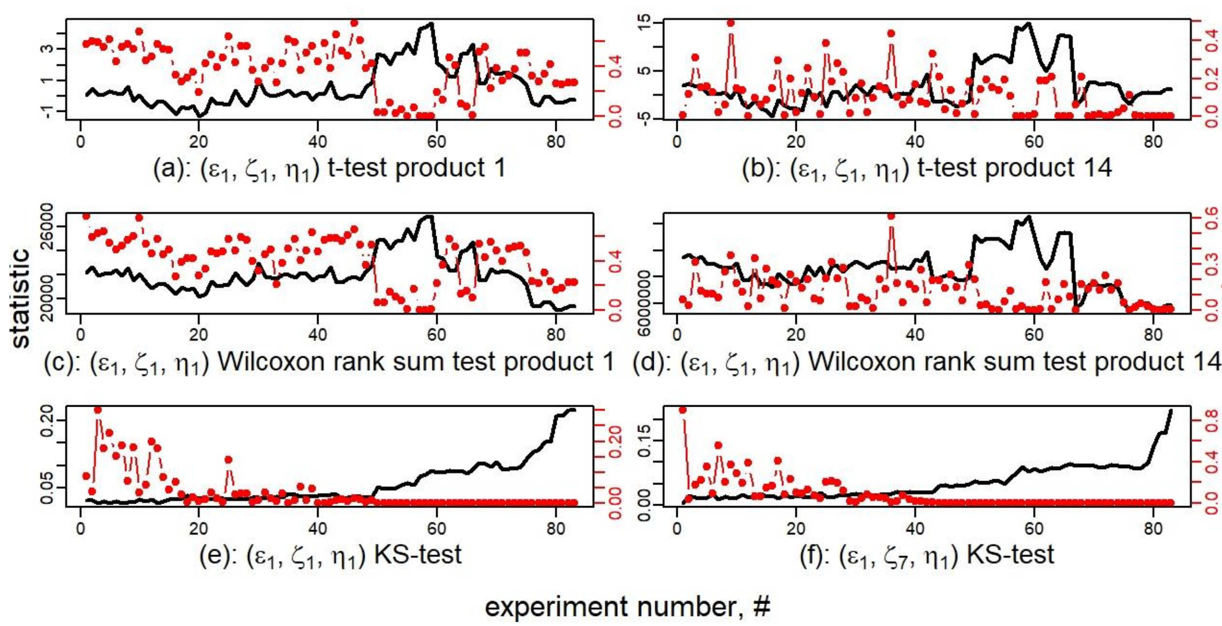

Figure 7: (a) and (b) t-test, (c) and (d) Wilcoxon rank sum test, (e) and (f) Kolmogorov-Smirnov test: statistic (black) and p-value (red).

For the configuration (ε_1, ζ_1, η_1) for both products, the t-statistics can be observed first increasing and then eventually falling. This is unsurprising due to the shift of lot CT mean (see Figure 4b). Moreover, the p-value was above 0.2 for product 1 even for exp#83 and there were significant oscillations of the p-value from exp#1 to exp#83. In the end, it can be concluded that t-test is not suitable for the research being carried out in this paper. During the review process, it was suggested that the Wilcoxon rank sum test may provide a useful alternative to t-test (Figures 7c and 7b). The results were found to be similar to the t-test.

4.2 More suitable accuracy measurements

Figures 7e and 7f present the results of a classical KS-test. Here it is clear that the KS-statistics also demonstrate some fluctuations, but they are much smaller than the t-statistics and, more importantly, the KS-statistics do not grow monotonously but quite strongly. The p-value also has oscillations but it tends to zero quite quickly. It is assumed that using measurements based on lot CT distribution is more suitable for the simplification task. First of all, they are not sensitive to the "butterfly effect", which occurs for exp#1 (see Figure 5b), because they consider all lots as a whole (distribution) as opposed to just looking at a single one. Second, the "pathological" (non-monotonous) behavior due to lot CT mean shift is not as expressed (see Figure 4b), because the measurements are not as sensitive to the shift.

In the Section 3.3, six measurements based on lot CT distribution were introduced. Figures 8 and 9 present the results of these measurements based on the same configurations as before: (ε_1, ζ_1, η_1) and (ε_1, ζ_7, η_1). In Figure 8 it can be seen that SAD has the least fluctuations in comparison with other measurements. This occurs due to the use of the probability density function (PDF), which is built as a histogram with equidistant bins (Figures 4c and 4d). In this case, small changes inside one bin do not have an influence on the SAD value. On the other hand, we could see that SAD is less sensitive in the area exp#1-70 in comparison with others. The KS test shows moderate fluctuations and is more sensitive to the changes in the area exp#1-70. This happens as a result of the empirical cumulative distribution function (ECDF), which contains data about every single lot. But the KS-statistic is based on a measurement only of the maximum of the difference between the ECDFs, meaning the measurement of the difference is made only at one point. Before the experiments it was assumed that this would not be adequate and the preference was to include the whole ECDF. Such measurements include AD, CVM, DTS, and WASS. It is interesting to see that AD and CVM were less sensitive than KS in the area exp#1-40, and DTS and WASS were somewhat more sensitive than KS in the same area. Moreover, AD and CVM have larger oscillations in the area exp#70-83 than other accuracy measurements. These differences between AD/CVM and DTS/WASS are caused by using a factor of 2 in the formulas (Table 1).

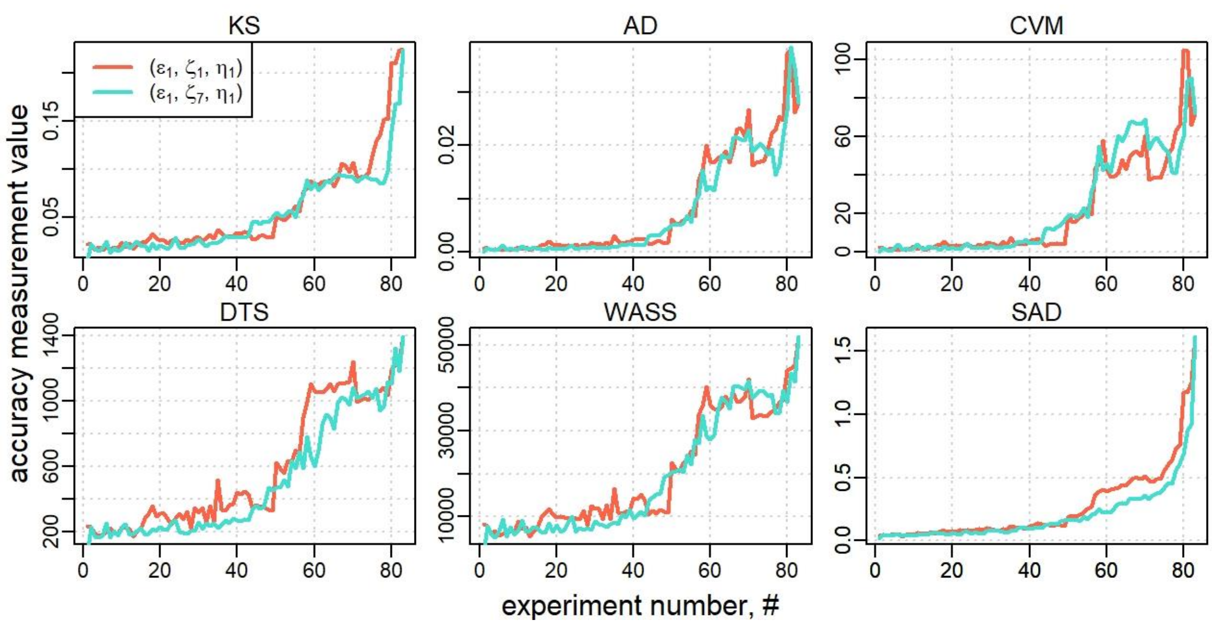

Figure 8: Six accuracy measurements in dependence on experiment number.

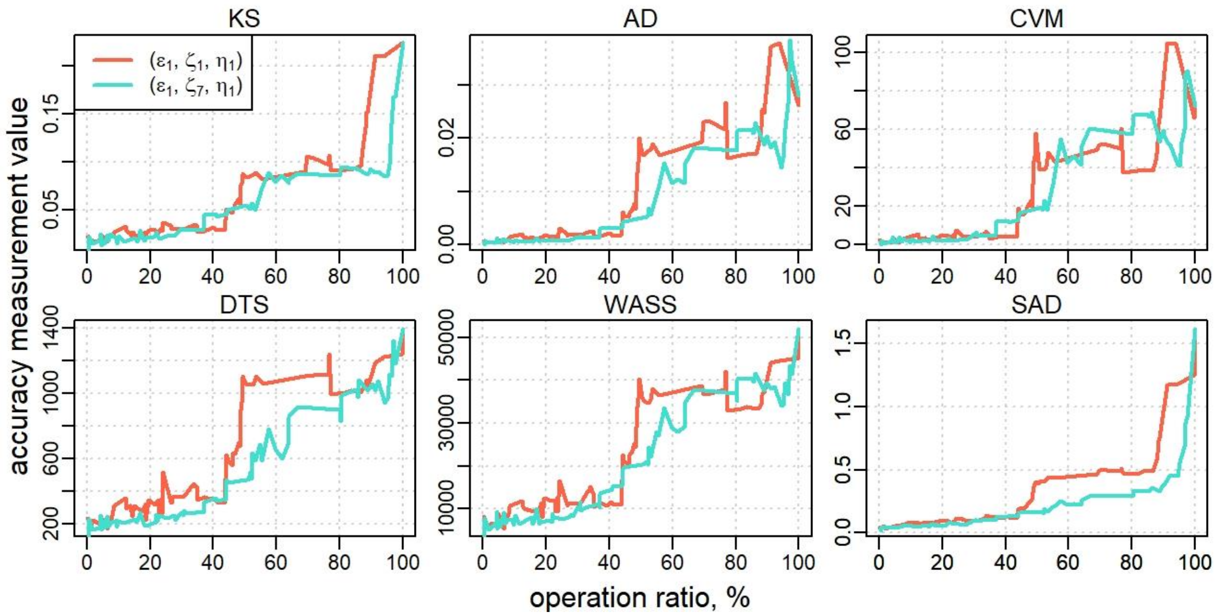

Figure 9: Six accuracy measurements in dependence on operation ratio.

Moreover, AD/CVM become much more sensitive to the lot CT mean shift as exp# grows. At the end (exp#80-83), even small changes in lot CT mean cause significant changes in AD/CVM (compare with Figure 4b). It is assumed that KS and SAD are more reliable measurements for simulation simplification, but further research is needed to prove this. Figure 8 represents a picture of the quality. For the calculations, another x-axis was needed – operation ratio (see Figure 9).

As mentioned above, the simulation time curve varies for different configurations (Figure 3b). This means that experiment numbers from one configuration are not equivalent to other configurations in the sense of simplification. As already mentioned, it was also observed that the curve is not monotonous. Therefore, it could not be used for the comparison of configurations. On the other hand, it was found that simulation time and operation ratio curves matched sufficiently. Thus, it was decided within this paper to use the operation ratio as a basis for the comparisons. Figure 9 presents the accuracy measurements with operation ratio as the x-axis. Here the difference between two configurations is even more clear than in Figure 8.

5 EXPERIMENTAL RESULTS

To compare various configurations, measurements were made along the axis (accuracy measurement, operation ratio). Those were then approximated with lines to produce the curves (see Figure 9), which were then numerically integrated to obtain a single value (area under the curve). This value represents a score of the particular configuration in a particular measurement. All configurations were ordered according to the score for each accuracy measurement. To compare two system configurations, ε_2 and ε_{2b}, a single list with those configurations was created. Table 2 presents the results for the first eighteen items in each list. For ε_2, a gray background is used. Other colors in the table are used only to easily distinguish different configurations.

Table 2 shows that there are different orders for different accuracy measurements. Nevertheless, it can be seen that configurations (ζ_x, η_1) mostly outperform (ζ_x, η_2): process step based delays usually perform better than tool set-based. However, the results of (ζ_x) and (ζ_{xf}) are less clear. It seems that those configurations do not differentiate from each other much, meaning it is difficult to evaluate which configuration is superior. For ε_1 and ε_{2b} configurations, (ζ_4, η_1), (ζ_{4f}, η_1), (ζ_8, η_1), and (ζ_{8f}, η_1) demonstrated the best accuracy overall. But for ε_2 there were other leaders, namely (ζ_7) and (ζ_{7f}). Based on our previous experience (Stogniy and Scholl 2019), it was assumed that ε_{2b} would always outperform ε_2. The data in Table 2 suggest otherwise. Further research is needed to determine the underlying reasons for this.

Table 2: Ordered configurations.

	FIFO (ε_1)													CR (ε_2) and CRb (ε_{2b})											
#	KS		AD		CVM		DTS		WASS		SAD			KS		AD		CVM		DTS		WASS		SAD	
	ζ	η	ζ	η	ζ	η	ζ	η	ζ	η	ζ	η		ζ	η	ζ	η	ζ	η	ζ	η	ζ	η	ζ	η
1	4	1	8f	1	8f	1	4f	1	4	1	8	1		8f	1	8f	1	8f	1	8	1	8	1	8f	1
2	4f	1	8	1	8	1	4	1	4f	1	8f	1		8	1	8	1	8	1	8f	1	8f	1	8	1
3	8	1	4	1	4	1	8f	1	8f	1	4	1		4f	1	4f	1	4f	1	7f	1	4f	1	8	2
4	8f	1	4f	1	4f	1	8	1	8	1	4f	1		4	1	4	1	4	1	4f	1	4	1	4f	1
5	4	2	5	1	9	1	4	2	4	2	8	2		8	2	7f	1	7f	1	4	1	8	2	8f	2
6	4f	2	5f	1	9f	1	5f	1	9	1	8f	2		8f	2	7	1	7f	1	8	2	8f	2	4	1
7	8	2	9	1	5	1	4f	2	4f	2	4f	2		7f	1	7f	1	7	1	8f	2	7f	1	4f	2
8	8f	2	9f	1	5f	1	5	1	8f	2	4	2		7f	1	7	1	8f	2	7f	2	7f	2	4	2
9	5	1	8f	2	8f	2	8f	2	8	2	5f	1		4f	2	8f	2	7	1	7	1	7f	1	7f	1
10	5f	1	8	2	8	2	8	2	9f	1	5	1		4	2	7f	2	8	2	7	2	7	2	7	1
11	6f	1	4f	2	4	2	9	1	5f	1	6f	1		7	1	6f	1	7f	2	7	1	7	1	7f	2
12	9	1	4	2	4f	2	9f	1	5	1	7f	1		6f	1	8	2	6f	1	7f	1	7	1	7f	1
13	9f	1	6f	1	3f	1	3f	1	3f	1	9	1		5f	1	7	2	7	2	2f	2	2f	2	5f	1
14	6	1	3f	1	3	1	3	1	3	1	6	1		7f	2	2f	1	2f	2	2f	1	6f	1	6f	1
15	3f	1	3	1	6f	1	6f	1	6f	1	3f	1		7	1	5f	1	6	1	5	1	2f	1	7f	2
16	3	1	6	1	6	1	5f	2	5f	2	3	1		5	1	5	1	5f	1	8f	1	4f	2	8f	1
17	6f	2	5f	2	5f	2	5	2	3f	2	9f	1		5f	1	2f	2	2f	1	6f	1	2	2	5f	1
18	6	2	6f	2	6f	2	3f	2	3	2	7	1		2f	2	6	1	5f	1	5f	1	7	2	8	1

6 CONCLUSIONS

In this paper, a simplification of process flow operations through substitution for constant delays was considered. Numerous experiments with three system configurations were carried out: First In First Out, Critical Ratio, and Critical Ratio with consolidated delays at the beginning. For the first two system configurations, it was found that queue time or cycle time standard deviation performed better as sieve functions for the substitution. For the third variant, queue length divided by batch size was the best choice.

Different accuracy measurements were used to analyze the results of experiments. Based on the presented analysis, it can be concluded that the following measurements can not be effectively used to estimate simplified models: lot cycle time mean absolute error, correlation coefficient, autocorrelation function, t-test, and Wilcoxon rank sum test. Accuracy measurements based on lot cycle time distributions are more suitable. It is assumed that a classical two sample Kolmogorov-Smirnov test and summarized absolute divergence (based on probability density function) outperform others as they have proven to be more stable. The experiments presented here showed that delays calculated based on average processing and queueing time are not precise enough and could lead to overall lot cycle time shift. This causes oscillation within the accuracy measurements. If that problem were to be solved, it can be assumed that accuracy measurements based on empirical cumulative distribution functions would show better performance.

In future research, we plan to analyze methods for more precise constant delay calculations and to consider other dispatching rules (Earliest Due Date, Operation Due Date). We will enhance our approach for variable workload scenarios and combine with the artificial process flow concept, which was presented earlier. Eventually, these approaches will be implemented in a real Infineon simulation model.

ACKNOWLEDGMENTS

A part of the work has been performed in the project iDev40. The iDev40 project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 783163. The JU receives support from the European Union's Horizon 2020 research and innovation programme. It is co-funded by the consortium members, grants from Austria, Germany, Belgium, Italy, Spain, and Romania.

REFERENCES

Alexopoulos, C., and W. D. Kelton. 2017. "A concise history of simulation output analysis". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 115-130. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Dowd C. 2018. twosamples: Fast Permutation Based Two Sample Tests. R package version 1.0.0. https://CRAN.R-project.org/package=twosamples, accessed 20th April 2020.

Gavrilov I., and R. Pusev. 2014. normtest: Tests for Normality. R package version 1.1. https://CRAN.R-project.org/package=normtest, accessed 20th April 2020.

Ewen, H., L. Mönch, H. Ehm, T. Ponsignon, J.W. Fowler, and L. Forstner. 2017. "A Testbed for Simulating Semiconductor Supply Chains". *IEEE Transactions on Semiconductor Manufacturing* 30(3): 293-305.

Fewster R. 2014. Course notes. STATS 325. Stochastic Processes Department of Statistics University of Auckland. https://www.stat.auckland.ac.nz/~fewster/325/notes/325book.pdf, accessed 20th April 2020

Johnson, R. T., J. W. Fowler and G. T. Mackulak. 2005. "A Discrete Event Simulation Model Simplification Technique". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 2172–2176. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Law, A. M. 2015. *Simulation Modeling and Analysis*. 5th ed. New York: McGraw-Hill Education

MIMAC Datasets. 1997. http://p2schedgen.fernuni-hagen.de/index.php?id=296, accessed 20th April 2020.

Mönch L., R. Uzsoy and J.W. Fowler. 2018. "A survey of semiconductor supply chain models part I: semiconductor supply chains, strategic network design, and supply chain simulation". *International Journal of Production Research* 56(13): 4524-4545.

Piplani, R. and S. A. Puah. 2004. "Simplification Strategies for Simulation Models of Semiconductor Facilities". *Journal of Manufacturing Technology Management* 15(7): 618-625.

R Core Team 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, accessed 20th April 2020.

Rose, O. 1999. "Estimation of the Cycle Time Distribution of a Wafer Fab by a Simple Simulation Model". In *Proceedings of the Semiconductor Manufacturing Operational Modeling and Simulation conference,* San Francisco, CA, USA,133-138.

Roser, C., and M. Nakano. 2015. "A quantitative comparison of bottleneck detection methods in manufacturing systems with particular consideration for shifting bottlenecks". In *Proceedings of the IFIP International Conference on Advances in Production Management Systems*. 273-281. Springer, Cham.

Stogniy I. and W. Scholl. 2019. Using delays for process flow simplification. In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 2372 – 2383. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Van der Zee, D. J. 2019. "Model Simplification in Manufacturing Simulation – Review and Framework". *Computers & Industrial Engineering*, 127: 1056-1067.

Völker S. and P. Gmilkowsky. 2003. "Reduced Discrete-Event Simulation Models for Medium-Term Production Scheduling". *Systems Analysis Modeling Simulation* 43(7): 867-883.

AUTHOR BIOGRAPHIES

IGOR STOGNIY is a PhD-student at the Technische Universität Dresden (Germany). He received his M.Sc in Automation and Control in 2007 from Bauman Moscow State Technical University (Russia). Since 2012 he has been working at the TU Dresden as a scientific researcher in the area of simulation. His email address is igor.stogniy@tu-dresden.de

WOLFGANG SCHOLL works as a Senior Staff Expert for modeling and simulation for Infineon Technologies in Dresden (Germany). He studied physics at the Technical University of Chemnitz (Germany) and graduated in solid-state physics in 1984. From 1984 to 1995 he worked as a process engineer for ZMD in Dresden. In 1996 he joined Infineon Technologies (formerly SIMEC) and worked in the field of capacity planning. Since 2003 he has been responsible for fab simulation. He supervises development and roll-out projects and is also a member of the Supply Chain Simulation community. His email address is wolfgang.scholl@infineon.com.