DEVELOPING A CALIBRATED DISCRETE EVENT SIMULATION MODEL OF SHOPS OF A DUTCH PHONE AND SUBSCRIPTION RETAILER DURING COVID-19 TO EVALUATE SHIFT PLANS TO REDUCE WAITING TIMES

Lieke R. de Groot

Belsimpel
Business Development
Waagstraat 1
Groningen, 9712JX, NETHERLANDS

Alexander Hübl

Faculty of Science and Engineering
IEM Design Group
Nijenborgh 4
Groningen, 9747AG, NETHERLANDS

ABSTRACT

Belsimpel deals with large waiting times during peak times under COVID-19 circumstances compared to non-COVID-19 times. Therefore, a discrete event simulation model of the Belsimpel shops has been developed. However, the model is not validated due to data scarcity. This paper proposes a model calibration procedure based on the idea that service times decrease during high-demanded hours and increase otherwise. The results show that the proposed procedure enables the generation of realistic Key Performance Indicator values. The calibrated simulation model can be used for analyzing the performance of possible improvements. Accordingly, the calibrated model is applied to investigate the impact of an improved employee scheduling. The results show that the mean waiting time decreases by 20–33 %, the maximum waiting time decreases by 12–20 %, and the mean service level increases by 3–11 %. These improvements enhance customer satisfaction while scheduling the same number of working hours.

1   INTRODUCTION

Waiting for service is one of the most undesirable, but inevitable, events in everyday life (Obermeier et al. 2020). Research has indicated that waiting for service negatively influences satisfaction, loyalty, and service quality perception (Voorhees et al. 2009). Unfortunately, waiting times are an unavoidable byproduct of the inability of service providers to perfectly forecast the demand for service. Belsimpel, a Dutch phone and subscription retailer, deals with the problem of long waiting times during peak times in their seven physical shops. These waiting times have increased because of COVID-19 regulations. Due to COVID-19, only three customers are allowed in the shops of the company simultaneously. Furthermore, competition from webshops has increased. Webshops have gained more attention, because online shopping is safer than in-store shopping. Moreover, webshops are also able to provide customers with shorter waiting times. In order to reduce the waiting times in shops, research is carried out on suggesting improvements for their queueing system and employee scheduling. Several improvements are suggested to reduce the waiting time. Examples of the suggested improvements are the introduction of service appointments, changing the employee scheduling, and increasing the number of workstations. To evaluate these suggested improvements, a discrete event simulation model of the Belsimpel shops has been developed. However, before being able to use the developed simulation model, it is necessary to validate the model. Therefore, this paper focuses on the calibration of a discrete event simulation model of the queueing system of Belsimpel shops during COVID-19 circumstances. This paper describes a joint research between Belsimpel and the University of Groningen.

In the past decades, much research has been performed regarding queueing systems, whereby two literature streams are dominating. On the one hand, analytical solutions are developed (Hübl and Altendorfer 2015; Altendorfer and Jodlbauer 2011; Altendorfer and Minner 2011; Buzacott and Shanthikumar 1993;

Medhi 1991) and on the other hand, those queueing problems that are most likely real-world problems are solved with discrete event simulation (Altendorfer et al. 2014; Hübl et al. 2011). Due to the stochastic nature of real-world problems, it is often the case that analytical models and techniques cannot represent a particular queueing system adequately (Kolker 2010; Shortle et al. 2018). Therefore, discrete event simulation is used more and more in recent literature dealing with real-world problems (Galankashi et al. 2016; Hübl 2015; Jodlbauer and Huber 2008; Madadi et al. 2013).

Discrete event simulation is considered to be a suitable tool to model the stochastic behavior in the Belsimpel shops. Discrete event simulation is widely used to model processes (Hübl et al. 2011; Hübl 2015; Fishman 1978). Other researchers have used discrete event simulation in retail settings as well. For instance, (Williams et al. 2002) also have presented a study in their paper in which discrete event simulation was used to specify, analyze, and improve operational processes in a retail shop. Moreover, (Eroglu et al. 2011) have developed a discrete event simulation model for a retail supply chain. Furthermore, (Scholtz et al. 2012) discuss an optimization study in which they developed a stochastic discrete-event simulation model for retail banking. Additionally, many other papers use discrete event simulation models to model COVID-19 circumstances. Examples are provided by Das (2020), Saidani and Kim (2021), Kierzkowski (2020), and Currie et al. (2020).

Since the simulation model is developed with the aim of evaluating suggested improvements, an accurate simulation model is necessary to guarantee the reliability of the results. The validity of the simulation model can be determined by comparing the Key Performance Indicators (KPIs) resulting from simulation with the observed results as proposed by Kleijnen (1995) and Sargent (2013). If the KPIs resulting from simulations do not match with the observed results, the original simulation model cannot be validated. However, model calibration can be applied to gain more insights into the lack of available data. In this research, it became evident that a lack of data regarding service times, service types, and arrival patterns exists. To collect the data, recorded videos of several days have been studied. The objective of calibrating a simulation model is to minimize the deviation between the observed data and the simulation output (De Santis et al. 2021). General discussions of calibration of simulation models are found in Hofmann (2005), Wigan (1972), and Ören (1981). In order to analyze the deviation between simulated and observed output, KPIs computed through the simulation model should be compared with the corresponding values obtained from the observed data. Therefore, it is necessary to select significant KPIs to represent the system at first (Liu et al. 2017).

A few examples of practical applications of simulation model calibration are provided by Davis et al. (2000), De Santis et al. (2021), and Quon et al. (2010). However, based on the presented literature, the calibration of real-world discrete event simulation models is still an issue. As Kleijnen (1995) suggested, the validation of a discrete event simulation model can be done by comparing the simulated and observed output, but there is still no calibration included in discrete event simulation to gain more knowledge of unknown input data. Contrary, in system dynamics it is already state of the art to do thorough model calibration (Oliva 2003).

In the next section of this paper, the simulation model is explained. Thereafter, the verification, validation, and calibration of the simulation model are discussed. Finally, some results of the simulation experiments with the calibrated simulation model are presented and the conclusions are drawn.

2 SIMULATION MODEL

In this section, the discrete event simulation model developed to model the queueing system of Belsimpel shops is discussed. First, the assumptions made to set up the model are explained. Thereafter, the simulation model is introduced with the use of a conceptual model. Subsequently, the input data are discussed.

2.1 Assumptions

The following assumptions, based on expert interviews and the input data presented in Section 2.3, are introduced to set up the model:

- For every observed day, an individual simulation model has been developed. Since there are almost no backlogs from previous days available, it was possible to focus on daily simulation models. The simulation model of one day is based on the input data from the observations of that specific day. Therefore, the model has been calibrated for all seven days individually and all the experiments are conducted with those individual models.
- Side tasks of the employees such as re-ordering products, refilling products, and cleaning the shop are not considered in the simulation models. These tasks are mainly performed in times it is not busy in the shop and, therefore, they do not directly influence the waiting time of the customers. If an employee is scheduled to perform side tasks only, this employee is not taken into account in the employee scheduling of the simulation.
- Set-up times for employees switching to another customer are not considered in the simulation. These set-up times are assumed to be very small and, therefore, not necessary to include.
- In the simulation, only the nine main categories of service are taken into account. In the observed data presented in Section 2.3, these categories make up for 99.15 % of all customers and, therefore, it seems reasonable to neglect the other types of service.
- Customers do not leave the shop when they are immediately allowed inside the shop upon arrival.
- 12.73 % of the customers that have to wait outside walk away from the queue.
- Balking is excluded from the simulation model, since it was not possible to analyze this from the video observations.
- Customers are divided among the nine categories as observed for each day in the video analysis explained in Section 2.3.
- Service times follow an empirical distribution based on the observed data (displayed in Section 2.3) per category of the specific day simulated.
- Based on the observed data, it is seen that a high fluctuation exists between service times in low-demanded hours compared to high-demanded hours. Service times are in general higher on low-demanded hours and lower on high-demand hours. Therefore, it is assumed that service times increase during low-demanded hours and service times decrease during high-demanded hours. On high-demanded hours, service times can be decreased with at most 50 % and on low-demanded hours, service times can be increased with at most 60 %.
- The waiting time of a customer is considered sufficient if the waiting time is less than or equal to five minutes. Based on this waiting time target, a service level regarding waiting time is calculated.

2.2 Simulation of Belsimpel Shops

The process of the Belsimpel shops is transferred to a discrete event simulation model in AnyLogic 8.6. The Personal Learning Edition (PLE) of AnyLogic 8.6 is used (Borshchev 2013). In Figure 1, a conceptual model is presented representing the simulation model. Basically, the simulation model consists of nine blocks. Upon arrival in Block A, a customer enters the outside queue represented by Block B. Afterwards, it is decided if there is a possibility to leave the queue in Block C. A leaving customer is directed to Block D. If the customer does not leave, the customer enters the inside queue in Block E. Subsequently, in Block F a decision is made based on the type of customer if the customer can be served with a tablet or a computer workstation is needed. In Block G, the service is provided. The service is based on the employee scheduling in Block I. Finally, the customer leaves the shop in Block H. Furthermore, with the dotted rectangle it is indicated that in Blocks E, F, and G only three customers are allowed due to COVID-19 regulations of the Dutch government.
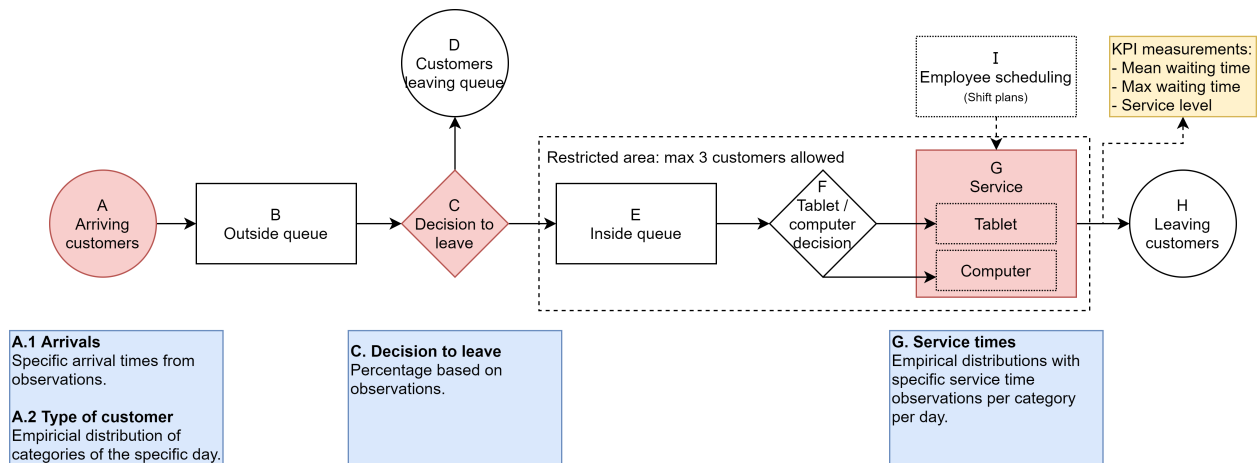
Figure 1: Conceptual model of the simulation model. Upon arrival, customers enter the outside queue, after which they enter the inside queue. Afterwards, employees provide service for the customers at a tablet or computer workstation, and finally the customers leave the shop.

In order to analyze the simulation output, KPIs are measured in the simulation model as well. Three KPIs are selected and also visualized in the right-hand block in Figure 1. The first KPI analyzed is the mean waiting time. The mean waiting time is calculated as the mean of the waiting times of all customers, where the waiting time is defined as the time between the arrival of the customer and the moment customer service begins. Moreover, the KPI maximum waiting time is analyzed, which is the maximum waiting time of all customers in the simulation. The third KPI analyzed is the service level for the waiting time. This service level is calculated by dividing the number of customers that have to wait at most five minutes by the total number of served customers.A waiting time target of five minutes is used because research conducted by a global retail technology provider has indicated that the average waiting time before customers leave, in this branch, equals 5.90 minutes (OmnicoGroup 2013). In order to serve the customers before leaving, five minutes is chosen as the reference point.

The shaded blocks in the middle row of Figure 1 represent the blocks where stochastic behavior is present. The rectangles in the bottom row explain the input data used to set up the model based on the video analysis as explained in Section 2.3. The arrivals in Block A are based on the specific arrival times from observations. The type of customer is modeled with an empirical distribution of the categories of the specific day modeled. The decision to leave, Block C, determines if there is a possibility that the customer leaves the queue. If the customer can enter the shop directly, the probability of leaving is zero and the customer enters the inside queue represented by Block E. However, if the customer cannot enter the inside queue directly, a probability is assigned that the customer will leave the queue. This probability is based on observations and modeled to be stochastic. Finally, the service times are based on empirical distributions with specific service time observations per category per day.

2.3 Input Data

In order to model the queueing system of the Belsimpel shops, input data are collected through video analysis. The recorded videos of several days of one of the shops are analyzed in which the following measurements are collected: the arrival time of the customer, the time at which the customer is allowed inside the shop, the start and end time of the service, and the type of service. With the use of these statistics, the outside and inside waiting times are calculated. Moreover, the service time is calculated. Furthermore, the company investigated has developed their own Enterprise Resource Planning (ERP) system. In this system, they also keep track of the sales in the physical shops. These data are matched with the video analysis in order to determine the specific sale of the customer. In this way, it was possible to determine

the type of customer and the division among the service types. Finally, the video observations are also connected and compared to the visitor counter recognizing customers entering or leaving the shop.

The video analysis has shown that there are nine main categories of service. The categories are provided in Table 1. Moreover, it can be concluded from the input analysis that the service times of the different types of services vary extensively. The mean, standard deviation, and coefficient of variation of the service times are provided in Table 1. On the one hand, only 2.91 minutes are needed on average to help a customer who would like to pick up a package. On the other hand, on average 17.45 minutes are used to serve a customer with a phone including an accessory. Furthermore, there are large variations within the service times of one type of service. This is visible in the relatively large standard deviations and coefficients of variation of the service times.

Table 1: The mean duration [min], standard deviation [min], coefficient of variation, and percentage of each service type based on in total $n = 626$ observations.

Service type	Mean	Standard deviation	Coefficient of variation	Percentage
Subscription	17.08	8.46	0.50	8.26 %
Ruturn or reparation	7.02	3.87	0.55	7.56 %
Phone	10.41	6.35	0.61	6.85 %
Phone & accessory	17.45	11.43	0.66	5.10 %
Package	2.91	1.96	0.67	17.05 %
Accessory	4.29	3.24	0.76	9.31 %
Only look around	5.77	5.23	0.91	1.93 %
Technical service	5.91	6.23	1.05	38.14 %
Pick-up subscription	3.98	4.24	1.07	5.80 %

Furthermore, the video analysis shows that on average 6.55 % of the total customers are walking away before they could be served in the observed days. It is noticed that, in the observed days under COVID-19 circumstances, the customers walk away from the queue only outside. Whenever a customer is allowed inside the shop, none of the customers did walk away in the observed days. Of all customers, 51.44 % had to wait outside of the shop. From these customers that had to wait outside, 12.73 % walked away.

3 VERIFICATION, VALIDATION, AND CALIBRATION OF THE SIMULATION MODEL

Since the simulation model is developed with the aim of evaluating suggested improvements, an accurate simulation model is necessary to guarantee the reliability of the results. First of all, the model should be verified. Model verification is usually defined to mean "ensuring that the computerized model represents a conceptual model within specified limits of accuracy" and, therefore, this definition is applied in this paper (Schlesinger et al. 1979). In order to verify the simulation model, structured walkthroughs have been performed with the problem owner of the company. Besides, in-depth discussions with one shop manager were conducted. Furthermore, the other shop managers were informed and had the possibility to provide their input. For verification, only one customer is placed in the simulation model and the movement of the agent is checked and compared to the real-world situation. The authors have spent a few days in the shops to make sure that they have an in-depth understanding of the queueing process in the real-world situation.

After verification, validity of the simulation model should be guaranteed. Model validation is often defined as "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" (Schlesinger et al. 1979). The validation process of the simulation model consists of two parts. On the one hand, an analysis of the simulation model and the applied assumptions is conducted with the use of expert interviews. Expert opinions are the simplest way to validate an artifact (Wieringa 2014). Expert interviews have been conducted with the problem owner, the head of the Business Development department of the company. Moreover, expert

interviews were conducted in working groups with the seven shop managers, the operation manager, and the other members of the Business Development team. On the other hand, the simulated KPIs were compared with the observed KPIs from the video analysis. The results of these comparisons are provided in Tables 2 to 4. This is done because the validity of the simulation model can be determined by comparing the KPIs resulting from simulation with the observed results as proposed by Kleijnen (1995) and Sargent (2013).

If the KPIs resulting from simulations do not match the observed results, the original simulation model cannot be validated. In this case, it was difficult to match the simulated and observed measures due to data scarcity. Due to COVID-19 regulations, the shops were closed and it was difficult to collect data. Furthermore, a high variation exists in the arrivals of customers over the day, the distribution of the type of services, and the service times. A large variation exists between the service times of the different categories, but also within one specific category a large variation between the service times exists. These circumstances combined made it very difficult to validate the simulation model. Therefore, model calibration is applied to gain more insights into the lack of available data. General discussions of calibration of simulation models are found in Hofmann (2005), Wigan (1972), and Ören (1981). A methodology is proposed to validate and calibrate the model in Section 3.1. Thereafter, the model calibration procedure is introduced in Section 3.2.

3.1 Methodology to Calibrate the Simulation Model

The methodology introduced to validate the simulation model is provided in a flowchart in Figure 2. After the set-up of the simulation model, pre-experiments are performed. Pre-experiments are experiments performed before the final experiments to test distributions, determine the number of replications, and determine which improvement has the strongest impact on the selected KPIs. From pre-experiments, it has been concluded that empirical distributions should be used, because generating standard distributions is not possible due to the lack of available data. Additionally, pre-experiments were used to determine the number of replications to use in simulations. From pre-experiments, it is also deduced that changing the employee scheduling has the strongest impact on the KPIs of all suggested improvements. Finally, pre-experiments have shown that the model is invalid based on the derived service times. Based on expert interviews and observed data, it became clear that the service time depends on the waiting time of the customer. If a customer does not have to wait, the service time is longer than in a situation where customers have to wait. This behavior is counter-intuitive, but difficult to tackle with less information about the service times. Therefore, the model is calibrated. Based on the calibrated simulation model, a feasibility test is performed. The feasibility test checks whether the values of the simulated output KPIs match the values of the observed KPIs. The feasibility test is based on confidence interval comparison with a significance level of 99.9 % and is presented in Tables 2 to 4.
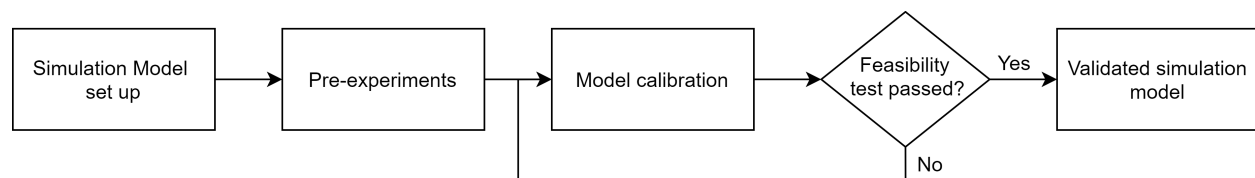

Figure 2: Flowchart describing the methodology applied to validate the simulation model.

All the simulations are performed with 200 replications. The selected number of replications is based on a trade-off between the many stochastics included in the model and the time needed for simulations and has been derived by pre-experiments. Selecting an applicable number of replications to run with the simulation model is crucial in ensuring that the desired accuracy of the results is obtained with minimal effort (Hoad et al. 2010). In order to select an appropriate number of replications, pre-experiments are performed starting with 50 replications. From these experiments, it is discovered that the confidence interval

decreased significantly by increasing the number of replications to 200. Since many behaviors in the model are stochastic, a relatively high number of replications is necessary to obtain significant results.

3.2 Calibration of the Simulation Model

From pre-experiments it is concluded that the validation of the developed simulation model requires a more sophisticated analysis. Therefore, the model is calibrated. Model calibration is a method to minimize the deviation between the observed data and the simulation output (De Santis et al. 2021). In the calibration for this paper, the focus was on minimizing the deviation between the observed and simulated mean waiting time. However, the deviation between the other observed and mean KPIs is investigated as well.

For the model calibration, the input data are as presented in the bottom-row blocks in Figure 1. This implies that the arrivals are based on the specific arrival times from the database in these simulations. In addition, it entails that the arrivals are modeled as stochastic based on historical data. Furthermore, the type of customer is selected from an empirical distribution of the specific day and the service times are based on empirical distributions with specific service times observed per category per day.

The calibration of the model is based on the basic idea that employees work faster on high-demanded hours and work slower on low-demanded hours. For instance, when not busy, an employee could take more time to explain the functionalities of a new phone or will more readily have a small talk conversation with the customer. Contrary, when busy, an employee will try to serve the customer faster. Upon arrival, the processing time of a customer is drawn from an empirical distribution based on the video observations of that specific day. Thereafter, the model determines if there is an outside queue, visualized by Block B in Figure 1. If so, the empirically assigned service time of customers already being served and customers of the inside queue are decreased by a certain percentage. Contrary, when there is no queue, the assigned service times are increased by a certain percentage. It is modeled that the service time of one customer can only be decreased and can only be increased once in order to prevent compounding decreases or increases. For every observed day, the right percentage is found to match the mean waiting time of experiments and observations. It is determined for every day individually, because different employees are working on the observed days. This implies that every day observed is used to calibrate the model of that specific day. In addition, it is assumed that the service time can be decreased by at most 50 % and can be increased by at most 60 %. Larger decreases or increases do not seem realistic. A larger upper bound for the increase percentage is chosen compared to the decrease percentage because it is easier to increase the service time than to decrease the service time. This holds because decreasing service times implies working faster while increasing service times implies taking more care of the customer's needs.

The adjustment percentages that were needed to match the simulated and observed mean waiting times are provided in Table 2. The observed and simulated mean waiting times are presented including the ratio between the simulated mean waiting time and the observed mean waiting time. Moreover, the percentage that is used to decrease the service time on high-demanded hours is presented in the column Decrease and the percentage that is used to increase the service times on low-demanded hours is presented in the column Increase. The percentages for increase and decrease applied vary per day. The percentages are dependent on the specific employees working on a specific day. Each employee represents different behavior and the combinations of employees changes over the days. Furthermore, the behavior of one specific employee can also change over days. For example, an employee can behave differently on a Monday than on a Friday.

From the column Ratio in Table 2, it can be seen that in five out of seven days the simulated mean waiting time is significant and differs by at most 4 % from the observed mean waiting time. In the other two days, the mean waiting time differs by at most 16 % from the observed values.

Table 2: Validation of calibrated simulation model: Comparison between observed mean waiting times and simulated mean waiting times. * This result is not significant based on a confidence level of 99.9 % compared to the observed output.

Day	Observed mean waiting time	Simulated mean waiting time	Ratio	Decrease	Increase
Day 1	1.80	2.08*	1.16	50 %	5 %
Day 2	0.99	1.00	1.01	30 %	5 %
Day 3	0.77	0.78	1.01	5 %	40 %
Day 4	1.36	1.16*	0.85	0 %	60 %
Day 4	1.79	1.77	0.99	20 %	10 %
Day 6	1.74	1.72	0.99	10 %	10 %
Day 7	4.59	4.76	1.04	15 %	10 %

Having matched the mean waiting time with the observations from the video analysis, it is also necessary to compare other KPIs. Therefore, the simulated maximum waiting time and mean service level are compared to the observed values. The results of the comparison of the maximum waiting time and mean service level are presented in Tables 3 and 4, respectively.

From the column Ratio in Table 3, it can be seen that the maximum waiting time differs in four out of seven days by at most 17 % of the observed maximum waiting time. For the other three days, the maximum difference equals 34 %. As the maximum waiting time is based on one customer per replication, outliers do have a large influence on the simulation waiting time. Therefore, it is expected that outliers are responsible for these large differences. Furthermore, one of the days where the simulated maximum waiting time deviates more from the observed maximum waiting time is Day 6. Day 6 was a special promotion day, which is the busiest day of the year at the shops of the company. On this day, the situation is different from the general situation for several reasons. First of all, a higher percentage of customers already knew what they wanted to buy when entering the shop. This implies that the mean time needed to serve a customer is shorter on this specific day. Moreover, more employees were scheduled on this day. As only three customers were allowed simultaneously inside the shop, employees served some of the customers while they were waiting outside the shop. For instance, the people only picking up a package could already be served outside. This adaptation from the general situation makes it difficult to compare the simulated results with the observed results of Day 6. The day is still used, because several promotion days are held during the year. The problem of long waiting times is the largest on such a promotion day. Therefore, it is considered relevant to investigate if the calibrated model is valid on such promotion days.

Table 3: Validation of calibrated simulation model: Comparison between observed maximum waiting times and simulated maximum waiting times. * This result is not significant based on a confidence level of 99.9 % compared to the observed output.

Day	Observed max waiting time	Simulated max waiting time	Ratio	Decrease	Increase
Day 1	25.03	23.03	0.92	50 %	5 %
Day 2	13.12	11.18*	0.85	30 %	5 %
Day 3	8.50	9.95*	1.17	5 %	40 %
Day 4	21.22	13.90*	0.66	0 %	60 %
Day 5	14.93	15.66	1.05	20 %	10 %
Day 6	21.33	14.28*	0.67	10 %	10 %
Day 7	38.08	27.76*	0.73	15 %	10 %

From the results displayed in the column Ratio in Table 4, it can be concluded that the simulated service level differs by at most 3 % compared to the observed service level. This implies that the simulated service

level is close to the observed service level and is, therefore, a good representation of reality. However, still two days are not considered significant.

Table 4: Validation of calibrated simulation model: Comparison between observed service level and simulated service level. * This result is not significant based on a confidence level of 99.9 % compared to the observed output.

Day	Observed service level	Simulated service level	Ratio	Decrease	Increase
Day 1	0.90	0.88*	0.97	50 %	5 %
Day 2	0.95	0.92*	0.97	30 %	5 %
Day 3	0.94	0.94	1.00	5 %	40 %
Day 4	0.91	0.91	1.00	0 %	60 %
Day 5	0.88	0.86	0.99	20 %	10 %
Day 6	0.89	0.88	0.98	10 %	10 %
Day 7	0.73	0.71	0.98	15 %	10 %

Concluding, these experimental results show that the calibrated discrete event simulation model obtained by the proposed procedure generated performance measures comparable to the observed measures, while simulation results obtained with the default input were significantly different from the observed data. Although the adjusted service times ensure that the KPIs of the discrete event simulation model are on average close to the corresponding observed values, some divergences between observed and simulated output are observed. This is the result of the problem of data scarcity, which undermines the overall accuracy of the simulation model. Nevertheless, the calibrated simulation model may be considered sufficiently reliable with respect to the investigated KPIs. Because there still is a margin for improvement, some ideas for further research will be recommended in Section 5.

4 APPLICATION OF THE CALIBRATED SIMULATION MODEL

The calibrated simulation model is used to evaluate improvement suggestions for the queueing system and employee scheduling of the Belsimpel shops. One of the suggested improvements is to change the employee scheduling. For different days, multiple improvements are suggested. In general, the more employees available, the more customers can be served at the same time and, consequently, the shorter the waiting times for the customer. However, a limit is present due to the restriction of a maximum of three customers in the shops due to COVID-19. The initial employee schedule and four improved scenarios for Saturdays are presented in Table 5. One employee is assigned per shift. Furthermore, to compare the different employee scheduling scenarios, the total number of payable hours are indicated per scenario. In this calculation, breaks of employees are not included, because the employees are not paid during the break.

Table 5: Initial and four improved scenarios of the standard employee schedule on a Saturday at the Belsimpel shops. One employee is assigned per shift.

	Initial	Saturday 1	Saturday 2	Saturday 3	Saturday 4
Shifts	10-18	10-18	10-18	10-18	10-18
	10-18	10-18	10-18	10-13:30	10-13:30
	10-18	12-17	11-17	11:30-15	13-18
		13-16	13-16	13-18	11-17:30
				14-17	
Total hours	22.5	23	24	22.5	22.5
Employees	3	4	4	5	4

The employee schedule is changed in the simulation model by changing Block I in Figure 1. All four improved scenarios are tested with the calibrated simulation model. As the results of the four improved scenarios do not significantly differ from each other, only the results of scenario Saturday 4 are provided here. In this scenario, the same number of hours is scheduled as in the initial scenario. However, one more employee is working compared to the initial scenario. The improved employee schedule is simulated on two observed Saturdays, day 5 and day 7. The simulation output of day 5 and day 7 are provided in Tables 6 and 7, respectively. For all KPIs, the value obtained from observations is listed. Next to the observed value, the value simulated initially with the calibrated model is presented. Additionally, the simulated value with the suggested improvements is provided and the improvement percentage is calculated representing a comparison between the initial and improved simulated value. All improvements are significant compared to the initially simulated value based on a significance level of 98 %.

Based on Tables 6 and 7, it can be concluded that implementing the improved employee scheduling results in a decrease of 20–33 % of the mean waiting time compared to the initial mean waiting time. The maximum waiting time decreases by 12–20 %. The mean service level increases by 3–11 % which implies that more customers can be served within five minutes. The percentages of improvement vary per day, because every day is different. Every day has a different arrival pattern, a different division over the service types, and different service times. Overall, the improvements will enhance customer satisfaction, while scheduling the same number of working hours. Therefore, the improvement can be implemented without additional costs. However, the improvements require more flexibility of the employees, as four employees are needed over the day instead of three employees.

Table 6: Results of the simulation of the improved employee scheduling (scenario Saturday 4) – Day 5.

Key Performance Indicator	Observed value	Simulated value initial	Simulated value improvement	Improvement %
Mean waiting time	1.79	1.77	1.43	19.50 %
Max waiting time	14.93	15.66	13.83	11.65 %
Mean service level	0.88	0.86	0.89	2.87 %

Table 7: Results of the simulation of the improved employee scheduling (scenario Saturday 4) – Day 7.

Key Performance Indicator	Observed value	Simulated value initial	Simulated value improvement	Improvement %
Mean waiting time	4.56	4.76	3.21	32.51 %
Max waiting time	38.08	27.76	22.15	20.22 %
Mean service level	0.73	0.71	0.79	10.94 %

5 CONCLUSION

This paper addresses one of the validating problems that frequently affect discrete event simulation models. In particular, the focus is on introducing a model calibration procedure to validate a discrete event simulation model for analyzing the queueing system of Belsimpel shops during COVID-19 circumstances. During the COVID-19 epidemic, long waiting times exist during peak times in the Belsimpel shops, since a maximum of three customers is allowed simultaneously in the shops. A discrete event simulation model has been developed in order to analyze the queueing system of Belsimpel shops during COVID-19 circumstances. However, data scarcity exists, which makes it difficult to validate the discrete event simulation model. For this purpose, a model calibration procedure is proposed to validate the simulation model. The calibration is based on the basic idea that service times increase on low-demanded hours and decrease on high-demanded hours. The results of the feasibility tests show that the proposed model calibration procedure enables the

generation of realistic values of the selected KPIs. This allows for analysing the KPIs of the queueing system. Also, it is possible to evaluate possible improvement suggestions. Accordingly, the model is applied to analyze an improved employee scheduling. Results show that the improved employee scheduling will result in a decrease in the mean waiting time of 20–33 %. The maximum waiting time decreases by 12–20 % and the service level increases by 3–11 %. These improvements enhance customer satisfaction, while scheduling the same number of working hours. For further research, it is recommended to acquire more input data. Currently, only seven days of input data are used. With more input data, it would be possible to apply a better feasibility test and to test which distributions would fit the service times of each category. Besides, more data would give a better insight in the upper bounds of the increased and decreased percentages used in the model calibration. Finally, it is recommended to acquire data from the other Belsimpel shops as only one shop is observed in this research. It would be interesting to see if there are differences among the shops.

REFERENCES

Altendorfer, K., A. Hübl, and H. Jodlbauer. 2014. "Periodical Capacity Setting Methods for Make-to-Order Multi-Machine Production Systems". *International Journal of Production Research* 52(16):4768–4784.

Altendorfer, K., and H. Jodlbauer. 2011. "An Analytical Model for Service Level and Tardiness in a Single Machine MTO Production System". *International Journal of Production Research* 49(7):1827–1850.

Altendorfer, K., and S. Minner. 2011. "Simultaneous Optimization of Capacity and Planned Lead Time in a Two-Stage Production System with Different Customer Due Dates". *European Journal of Operational Research* 213(1):134–146.

Borshchev, A. 2013. *The Big Book of Simulation Modeling: Multimethod Modeling with AnyLogic 6*. AnyLogic North America.

Buzacott, J., and J. Shanthikumar. 1993. *Stochastic Models of Manufacturing Systems*. Englewood Cliffs, New Jersey: Prentice Hall.

Currie, C., J. Fowler, K. Kotiadis, T. Monks, B. Onggo, D. Robertson, and A. Tako. 2020. "How Simulation Modelling Can Help Reduce the Impact of COVID-19". *Journal of Simulation* 14(2):83–97.

Das, A. 2020. "Impact of the COVID-19 Pandemic on the Workflow of an Ambulatory Endoscopy Center: An Assessment by Discrete Event Simulation". *Gastrointestinal endoscopy* 92(4):914–924.

Davis, P., J. Bigelow, and J. McEver. 2000. "Informing and Calibrating a Multiresolution Exploratory Analysis Model with High Resolution Simulation: The Interdiction Problem as a Case History". In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. Joines, R. Barton, K. Kang, and P. Fishwick, 316–325. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

De Santis, A., T. Giovannelli, S. Lucidi, M. Messedaglia, and M. Roma. 2021. "A Simulation-Based Optimization Approach for the Calibration of a Discrete Event Simulation Model of an Emergency Department". *arXiv preprint arXiv:2102.00945*:1–28.

Eroglu, C., B. Williams, and M. Waller. 2011. "Consumer-driven Retail Operations". *International Journal of Physical Distribution & Logistics Management* 41(5):420–434.

Fishman, G. 1978. *Principles of Discrete Event Simulation*. New York, NY: John Wiley & Sons, Inc.

Galankashi, M., E. Fallahiarezoudar, A. Moazzami, N. Yusof, and S. Helmi. 2016. "Performance Evaluation of a Petrol Station Queuing System: A Simulation-based Design of Experiments Study". *Advances in Engineering Software* 92:15–26.

Hoad, K., S. Robinson, and R. Davies. 2010, 11. "Automated Selection of the Number of Replications for a Discrete-Event Simulation". *JORS* 61:1632–1644.

Hofmann, M. 2005. "On the Complexity of Parameter Calibration in Simulation Models". *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 2(4):217–226.

Hübl, A., and K. Altendorfer. 2015. "State Probabilities for an M/M/1 Queuing System with Two Capacity Levels". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, K. Chan, I. Moon, T. Roeder, C. Macal, and M. Rosetti, 2219–2226. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hübl, A., K. Altendorfer, H. Jodlbauer, M. Gansterer, and R. Hartl. 2011. "Flexible Model for Analyzing Production Systems with Discrete Event Simulation". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. Creasey, J. Himmelspach, K. White, and M. Fu, 1554–1565. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hübl, A. 2015. *Stochastic Modelling in Production Planning*. Wiesbaden, Germany: Springer Gabler.

Jodlbauer, H., and A. Huber. 2008. "Service-level Performance of MRP, Kanban, CONWIP and DBR due to Parameter Stability and Environmental Robustness". *International Journal of Production Research* 46(8):2179–2195.

Kierzkowski, A.and Kisiel, T. 2020. "Simulation Model of Security Control Lane Operation in the State of the COVID-19 Epidemic". *Journal of Air Transport Management* 88:101868.

Kleijnen, J. 1995. "Verification and Validation of Simulation Models". *European Journal of Operational Research* 82(1):145–162.

Kolker, A. 2010. "Queuing Analytic Theory and Discrete Events Simulation for Healthcare: Right Application for the Right Problem".

Liu, Z., D. Rexachs, F. Epelde, and E. Luque. 2017. "A Simulation and Optimization Based Method for Calibrating Agent-Based Emergency Department Models under Data Scarcity". *Computers & Industrial Engineering* 103:300–309.

Madadi, N., A. Roudsari, K. Wong, and M. Galankashi. 2013. "Modeling and Simulation of a Bank Queuing System". In *Fifth International Conference on Computational Intelligence, Modelling and Simulation*, 209–215. IEEE.

Medhi, J. 1991. *Stochastic Models in Queueing Theory*. Amsterdam, The Netherlands: Elsevier Science.

Obermeier, G., R. Zimmermann, and A. Auinger. 2020. "The Effect of Queuing Technology on Customer Experience in Physical Retail Environments". In *Lecture Notes in Computer Science*, edited by F. Nah and K. Siau, Volume 12204, 141–157. Springer, Cham.

Oliva, R. 2003, 02. "Model Calibration as a Testing Strategy for System Dynamics Models". *European Journal of Operational Research* 151:552–568.

OmnicoGroup 2013. "Gone in 6 Minutes: Why Queueing In-store Should be a Thing of the Past".

Ören, T. 1981. "Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference". *Communications of the ACM* 24(4):180–189.

Quon, P., D. Vanness, A. Kansal, P. Hillemanns, V. Rémy, and S. Quilici. 2010, 11. "Calibration of a Discrete Event Simulation Model of Natural History of HPV-related Diseases". *Value in Health* 13(7).

Saidani, M., and H. Kim. 2021. "A Discrete Event Simulation-Based Model to Optimally Design and Dimension Mobile COVID-19 Saliva-Based Testing Stations". *Simulation in Healthcare* 16(2):151–152.

Sargent, R. 2013. "An Introduction to Verification and Validation of Simulation Models". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. Kuhl, 321–327. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schlesinger, S., R. Crosbie, R. Gagné, G. Innis, C. Lalwani, J. Loch, R. Sylvester, R. Wright, N. Kheir, and D. Bartos. 1979. "Terminology for Model Credibility". *SIMULATION* 32(3):103–104.

Scholtz, E., J. Bekker, and D. Du Toit. 2012. "Multi-Objective Optimisation with Stochastic Discrete-Event Simulation in Retail Banking: A Case Study". *ORiON* 28(2):117–135.

Shortle, J., J. Thompson, D. Gross, and C. Harris. 2018. *Fundamentals of Queueing Theory*, Volume 399. New York, NY: John Wiley & Sons.

Voorhees, C., J. Baker, B. Bourdeau, E. Brocato, and J. Cronin. 2009. "It Depends: Moderating the Relationships among Perceived Waiting Time, Anger, and Regret". *Journal of Service Research* 12(2):138–155.

Wieringa, R. 2014. *Design Science Methodology for Information Systems and Software Engineering*. Berlin, Heidelberg: Springer.

Wigan, M. 1972, 05. "The Fitting, Calibration, and Validation of Simulation Models". *Transactions of The Society for Modeling and Simulation International – SIMULATION* 18:188–192.

Williams, E., M. Karaki, C. Lammers, A. Verbraeck, and W. Krug. 2002. "Use of Simulation to Determine Cashier Staffing Policy at a Retail Checkout". In *Proceedings 14th European Simulation Symposium*, edited by A. Verbraeck and W. Krug, 172–176. Ghent, Belgium: SCS Europe.

AUTHOR BIOGRAPHIES

LIEKE R. DE GROOT is a graduate intern in the Business Development team of Belsimpel. She is a master student of Industrial Engineering & Management at the University of Groningen (Netherlands). Her email address is lieke.d.groot@belsimpel.nl.

ALEXANDER HÜBL is a member of the IEM Design Group at the University of Groningen (Netherlands). He holds a PhD in logistics and operations management from University of Vienna, Austria. His research interests include discrete event simulation, agent-based simulation, queuing theory, and their applications in logistics and operations management. His email address is a.hubl@rug.nl.