Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

BIG DATA-DRIVEN SERVICE LEVEL ANALYSIS FOR A RETAIL STORE

Rie Gaku

Graduate School of Business Administration
Momoyama Gakuin University
1-1-Mamabino
Izumi, Osaka 5941198, JAPAN

Soemon Takakuwa

Department of Industrial and Systems Engineering
Chuo University
1-13-27 Kasuga, Bunkyo-ku
Tokyo 1128551, JAPAN

ABSTRACT

Using simulation technology, a procedure is proposed for a big data-driven service-level analysis for a real retail store. First, a data generator is designed to randomly select a sample of an expected number of customers or sampling data on a certain day from a large-scale dataset of sales predefined. Second, the clerk schedules are inputted into a data table created using Excel. Finally, simulation modeling mimics the service process of the retail store to examine and analyze the customer service level based on the selected data and the inputted clerk schedules. The proposed procedure for big data-driven service-level analysis shows the relations between the influencing service-level elements between the number of customers coming into stores, the frequency of customers, and the average customer service time. The procedure is generic and can easily be used to examine the service level in the remote past or to analyze and forecast the future.

1 INTRODUCTION

Innovations in ICT move business intelligence toward more large-scale business data utilization to improve business performance in retail environments. As a new trend in operation strategy, franchise chains of convenience stores in Japan are using big data collected from point of sales systems (POSs) to make decisions accurately and quickly, where the big data is defined as large-scale point of sales (POS) data that require special data-processing technologies or advanced analytical technologies in order to create business intelligence. In particular, a high level of service is a significant competitive advantage to analyze the quality of customer service with large-scale POS data. Therefore, a customizable and flexible service-level analytical tool is currently an urgent need for the operation managers of franchise convenience stores in Japan.

In this paper, a procedure of a big data-driven service-level analysis for a real retail store is proposed. First, a data generator is designed to randomly select a sample of an expected number of customers or sampling data on a certain day from a large-scale customer dataset of sales predefined. Second, the clerk schedules are inputted into a data table using Microsoft Excel. Finally, simulation modeling mimics the service process of the retail stores to examine and analyze the service level of retail stores based on the selected data and clerk schedules inputted, where the service level is defined as a percentage of customers served within a fixed time period.

Simulation is a general methodology designed to be used to evaluate and estimate effects on system performance. In the logistics and service industries, simulation is used as a powerful analyzing and design tool in performance management. Miwa and Takakuwa (2008) presented a procedure of simulation modeling for in-store merchandizing to examine customer flows in retail stores. and Miwa and Takakuwa (2010) presented a simulation analysis for optimization of staffing problems at a retail store. Takakuwa and Wijewickrama (2008) developed a discrete event simulation to optimize doctors' schedules in all departments of an outpatient hospital ward in the Nagoya University hospital. Wijewickrama and Takakuwa (2004) presented a simulation analysis of patient flow, aiming to shorten waiting time by identifying an optimum doctor mix. Liu (2009) discussed a simulation IP-based approach to optimize a retail store's personnel planning, particularly by using POS data. The objective of this study is to propose a customizable and flexible service-level analytical tool to analyze and understand the quality of customer service, particularly by using large-scale POS data.

2 CHARACTERIZATION OF CUSTOMER VISITING FREQUENCY IN THIS STUDY

The convenience store in this study is a franchise store of the FamilyMart Company Limited on the campus of Nagoya University, Japan. The shop's hours of operation are from 7AM to 11PM. Sales items are goods that are commonly sold at Japanese convenience stores, such as basic grocery items, magazines, comic books, soft drinks, and snack foods.

Through a collaborative study between Nagoya University and FamilyMart, the company has constantly provided POS data from the store from its opening until the present day for eight years. In this section, a considerable portion of the biennial POS data associated with the service level analysis is carefully reviewed to analyze the characterization of customers' visiting frequency of the convenience store. Thereafter, several frequency patterns of customers are drawn based on the data coefficients.

By analyzing large-scale POS data collected from the store in this study, the number of customers every 15 minutes for a weekday period is illustrated in Figure 1. The peak time in this store was mid-day lunchtime (approximately 12:00PM-1:00PM) and during class breaks (approximately 8:30AM-8:45AM, 10:15AM-10:30AM, 2:30PM-2:45PM, 4:15PM-4:30PM, and 6:00PM-6:15PM). Because customer waiting time is often lengthy during the peak time of weekday periods, a flexible service-level analytical tool needs to be developed to analyze and understand the quality of customer service in this store.

Figure 2 shows a part of change in the number of customers over the past years. The fourth period and the sixth period reflect the central tendency of traffic during the work week in the main semesters. Two period datasets called "Dataset for Spring Semester" and "Dataset for Fall Semester" are designed to select and create simulation entities under examination in this paper, as shown in Table 1. A portion of the results of quantifying a correlation coefficient between each number of customers at the quarter-hour is shown in Table 2. This result provides evidence for a greater degree of a linear relationship with the period datasets designed. Especially, this result is important in the simulation model because the datasets are used to generate the relative number of customer entities.

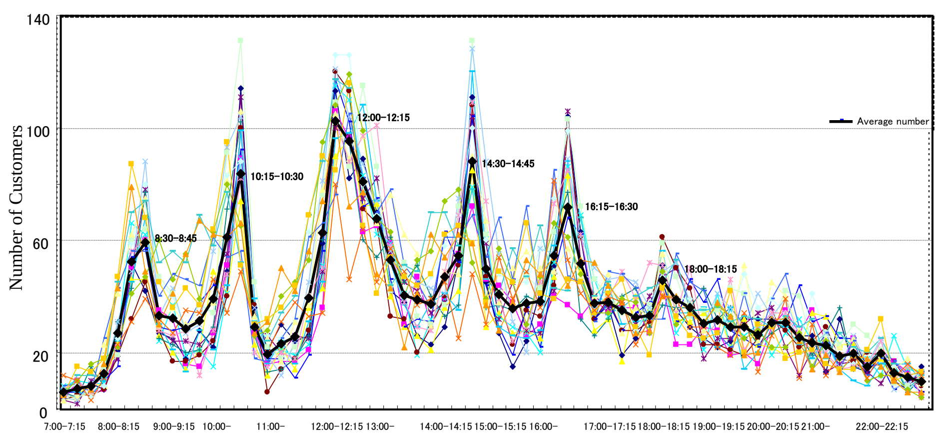

Figure 1: The waves of customers in 15-minute intervals.

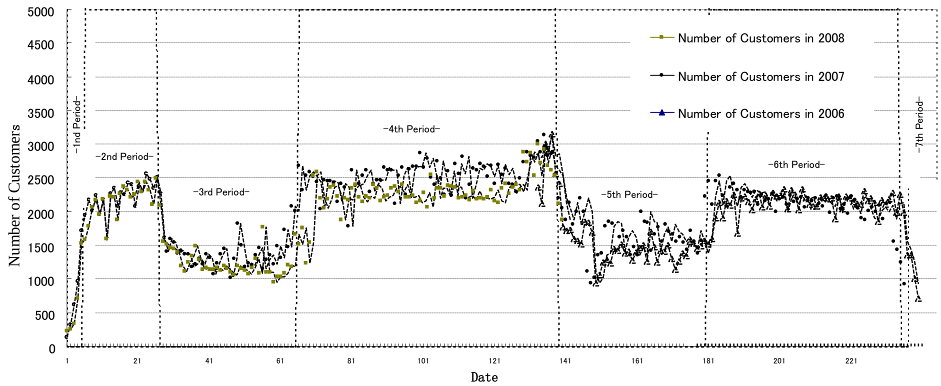

Figure 2: The change in the number of customers during the work week.

Table 1: A six-week period for datasets.

Period No.	Period	Date
1	Winter Vacation-1	1st week (e.g. 1/1~1/7)
2	Resumption of Fall Semester	2nd~5th week(e.g. 1/8~2/6)
3	Spring Vacation	6th~13th week (e.g. 2/7~4/4)
4	Spring Semester	14th~ 30th week (e.g. 4/5~7/31)
5	Summer Vacation	31th~38th week (e.g. 8/1~9/26)
6	Fall Semester	39th~51st week (e.g. 9/27~12/22)
7	Winter Vacation-2	52nd week (e.g. 12/23~12/31)

Table 2: Correlation coefficients of number of customers in 15-minute intervals.

		7/1	7/2	7/3	7/4	7/7	7/8	7/9	7/10
7/1	Pearson Correlation	1	.819(**)	.909(**)	.903(**)	.925(**)	.942(**)	.861(**)	.896(**)
	Sig. (2-tailed)		1.24896E-16	2.84318E-25	1.76833E-24	7.85172E-28	5.42117E-31	7.92101E-20	1.61424E-23
7/2	Pearson Correlation	.819(**)	1	.854(**)	.885(**)	.869(**)	.849(**)	.881(**)	.890(**)
	Sig. (2-tailed)	1.25E-16		3.06783E-19	3.09285E-22	1.38718E-20	7.57534E-19	7.45327E-22	8.75158E-23
7/3	Pearson Correlation	.909(**)	.854(**)	1	.892(**)	.896(**)	.928(**)	.867(**)	.928(**)
	Sig. (2-tailed)	2.84318E-25	3.06783E-19		4.25403E-23	1.74753E-23	2.80987E-28	1.9859E-20	2.61353E-28
7/4	Pearson Correlation	.903(**)	.885(**)	.892(**)	1	.914(**)	.912(**)	.852(**)	.904(**)
	Sig. (2-tailed)	1.76833E-24	3.09285E-22	4.25403E-23		6.55477E-26	1.04225E-25	4.33174E-19	1.36113E-24
7/7	Pearson Correlation	.925(**)	.869(**)	.896(**)	.914(**)	1	.921(**)	.892(**)	.881(**)
	Sig. (2-tailed)	7.85172E-28	1.38718E-20	1.74753E-23	6.55477E-26		4.07332E-27	4.14417E-23	7.33459E-22
7/8	Pearson Correlation	.942(**)	.849(**)	.928(**)	.912(**)	.921(**)	1	.869(**)	.917(**)
	Sig. (2-tailed)	5.42117E-31	7.57534E-19	2.80987E-28	1.04225E-25	4.07332E-27		1.44665E-20	1.66319E-26
7/9	Pearson Correlation	.861(**)	.881(**)	.867(**)	.852(**)	.892(**)	.869(**)	1	.877(**)
	Sig. (2-tailed)	7.92101E-20	7.45327E-22	1.9859E-20	4.33174E-19	4.14417E-23	1.44665E-20		2.02953E-21
7/10	Pearson Correlation	.896(**)	.890(**)	.928(**)	.904(**)	.881(**)	.917(**)	.877(**)	1
	Sig. (2-tailed)	1.61424E-23	8.75158E-23	2.61353E-28	1.36113E-24	7.33459E-22	1.66319E-26	2.02953E-21	

**. Correlation is significant at the 0.01 level (2-tailed).

3 PROCEDURES TO EVALUATE AND ESTIMATE SERVICE LEVEL

3.1 The Procedures

The objective of the procedures proposed is to evaluate and estimate the customer service level in the store using large-scale POS data. The service level is a performance measure within the store that is defined as the percentage of customers served within a fixed time period, known as the service level target. For instance, when 95% of customers were served within one minute, the service level is simply denoted as 95/60 seconds. First, the data generator is designed to randomly select a sample of an expected number of customers or sampling data on a certain day from a large-scale customer dataset of sales predefined. Here, the data generator is designed in Visual Basic for Applications (VBA) of Microsoft Excel. Some studies is related to simulation analysis making use of data generator (Takakuwa and Oyama 2003, Takakuwa and Shiozaki 2004, Takakuwa and Okada 2005). Second, the clerk schedules are inputted into a data table using Microsoft Excel. Finally, simulation modeling mimics the service process of the retail stores to examine and analyze the service level of retail stores based on the selected data and clerk schedules inputted.

As shown in Figure 3, the procedures to evaluate and estimate customer service levels in the store are itemized as follows:

[Step 1] An expected number of customers or sampling data on a certain day from “Dataset for Spring Semester” or “Dataset for Fall Semester” is selected randomly in Visual Basic for Applications (VBA) in Microsoft Excel, as shown in Figure 4. The customer dataset of sales and the customer dataset of detailed sales are shown in Tables 3 and 4. The two transaction number columns contain matching data to relate the two tables.

[Step 2] Using Microsoft Excel Power Query, detailed sales data are retrieved and stored in a new data table. Sales data and selected detailed sales data (called “complete data-set”) are used to perform a simulation model.

[Step 3] Clerk schedules data are inputted to a data table defined using Microsoft Excel, as shown in Tables 5 and 6.

[Step 4] Simulation modeling mimics the service process based on the flow of cash register service demonstrated in Figure 5. As Figure 5 shows, the service process of payment checking can be done by one salesclerk (a) or two salesclerks (b). The occasion of (b) can serve more customers than (a) in the same time period.

[Step 5] Simulation results, such as the service level, are obtained. For instance, the average customer waiting time for each interval is examined as a performance measure of simulation experiments.

[Step 6] The simulation output is checked to see whether the service level is satisfied; if not, it indicates that the clerk capacity is insufficient. Therefore, another salesclerk needs to be added to the interval in which the service level is minimum. Step 3 is repeated until the service level target is satisfied.

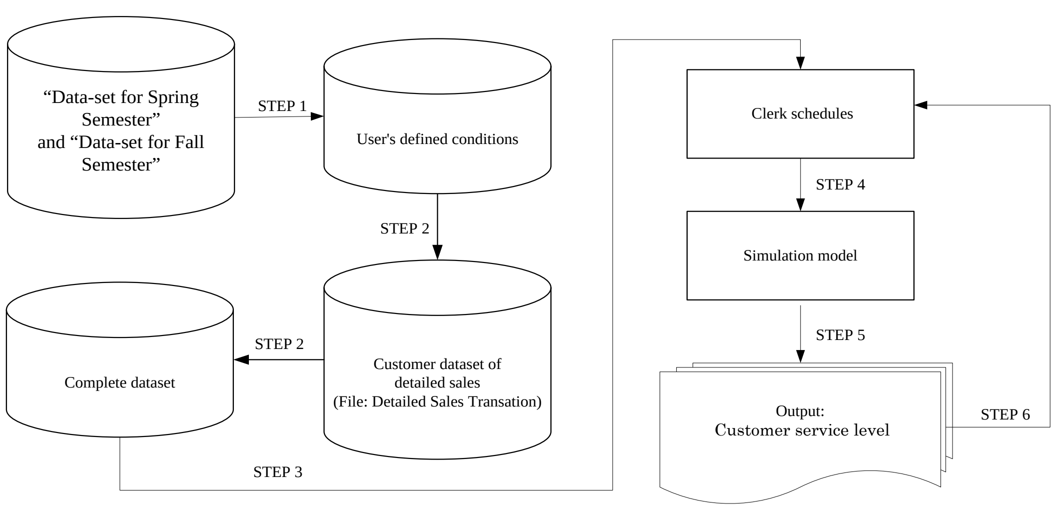

Figure 3: Procedures.

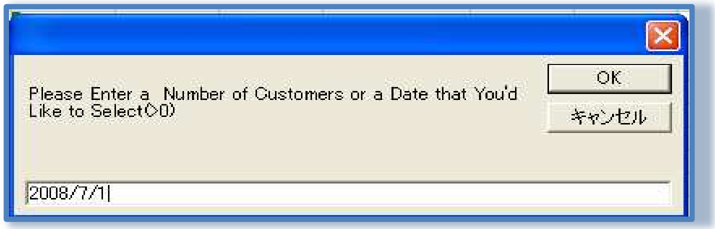

Figure 4: Excel VBA for selecting data.

Table 3: Customer dataset of sales.

Date of Sales	Store ID	Store Name	Register No.	Transaction No.	Time of Sales (hour)	Time of Sales (min.)	Customer Type	F/M, Age	No of Sales Items	Total Sum of Sales	Flag on Using Prepaid Card	Flag on Using IC Card	Name of IC Card	IC Card No.	Flag on Using Card	Transaction Code	Name of Transaction
2008-07-01	54858	Nagoya-U.	01	7	10	24	03	20's M	2	225	0	0	Non-IC	0	0	1000	Regular Registration
2008-07-01	54858	Nagoya-U.	01	8	10	24	03	20's M	1	160	0	0	Non-IC	0	0	1000	Regular Registration
2008-07-01	54858	Nagoya-U.	01	9	10	24	03	20's F	3	130	0	0	Non-IC	0	0	1000	Regular Registration
		…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
2008-07-01	54858	Nagoya-U.	01	17	10	28	03	20's M	1	144	0	0	Non-IC	0	0	1000	Regular Registration
2008-07-01	54858	Nagoya-U.	01	18	10	29	03	20's M	1	100	0	0	Non-IC	0	0	1000	Regular Registration
2008-07-01	54858	Nagoya-U.	01	19	10	30	03	20's M	1	100	0	0	Non-IC	0	0	1000	Regular Registration

* F/M: Femal / Man

Table 4: Customer dataset of detailed sales.

Date of Sales	Store ID	Store Name	Register No.	Transaction No.	Time of Sales (hour)	Time of Sales (min.)	Detail No.	Item Group ID	Name of Item Group	FM Code	Name of goods in region	Designated Goods No.	Truck-Delivery No.	Code of Detailed Sales Transaction	Name of Detailed Sales Transaction	Register ID	Name of Nominal accout	No. of Sales Items	Total Sum of Sales
2008-07-01	54858	Nagoya-U.	01	7	10	24	1	174	Chilled Juice	17432340	Vegetables &. Fruits	325320	01	2400	(Included Ta:	03	Food	1	80
2008-07-01	54858	Nagoya-U.	01	7	10	24	2	025	Sandwich-1st Delivery	02532220	Smoke Salmon &. Cheese	246103	01	2410	(Included Ta:	03	Food	1	143
2008-07-01	54858	Nagoya-U.	01	8	10	24	1	091	Sweet Bread	09116890	Chocolate D2	340610	01	2400	(Included Ta:	03	Food	1	160
2008-07-01	54858	Nagoya-U.	01	9	10	24	1	153	Sugar Candy	47558690	Cola Bittersweet	340049	01	2400	(Included Ta:	03	Food	1	130
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
2008-07-01	54858	Nagoya-U.	01	11	10	25	1	421	Coffee	42123310	Boss No Sugar 190G	266224	01	2400	(Included Ta:	03	Food	1	115

Table 5: Interval definition.

Time Interval	7:00	7:15	7:30	7:45	8:00	…	19:45
Time Interval No.	1	2	3	4	5	…	44

Table 6: Clerk schedules.

Clerk No.	Clerk name	Starting time	Over time	Working Quarterhours	Working Hours
1	Staff_1	1	33	32	8.00
2	Staff_2	1	27	26	6.50
3	Staff_3	27	44	17	4.25
4	Staff_4	22	44	12	4.00
:	:	:	:	:	:
12	Staff_12	2	19	17	4.25

3.2 Application

This section will discuss the applied procedures in an actual case. In Step 1, the sales data from 1 July, 2008 are selected in this study. The number of customers who appeared every 15 minutes on this day is shown in Figure 6. In Step 2, detailed sales data related to Step 1 are retrieved and stored in a new data table using Microsoft Excel Power Query. Selected sales data and selected detailed sales data (called "complete dataset") are used to perform a simulation model. In Step 3, clerk schedules data are inputted to a data table defined using Microsoft Excel, as shown in the Table 6 above.

The simulation models were created using Arena (Kelton, Sadowski, and Sturrock 2014). The selected simulation parameters are shown in Table 7. For Steps 4 and 5, the AS-IS clerk schedules for the peak time and the simulation output for the customer service level are shown in Figure 8 and Table 8. The simulation models are run for 50 replications and then verified and validated. Figure 7 provides an illustration of one part of the animation from the AS-IS model.

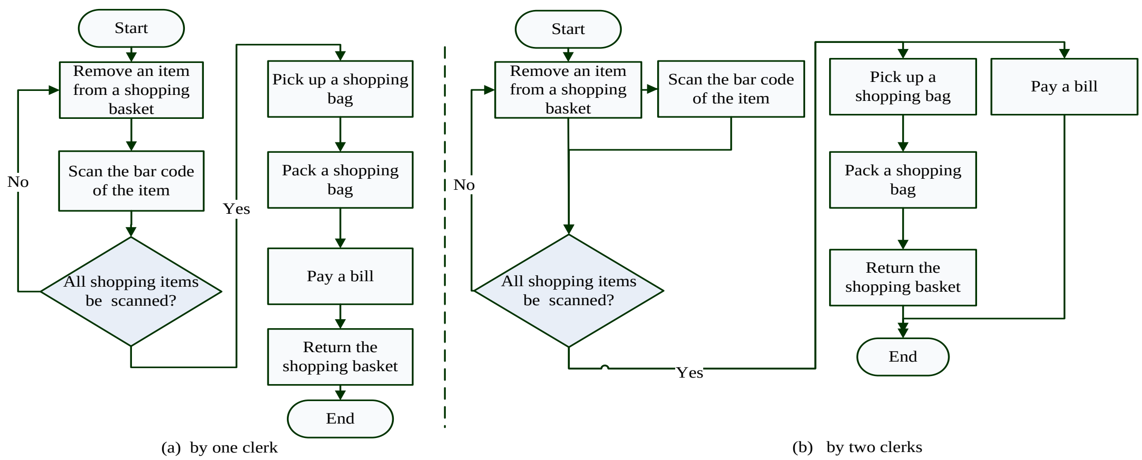

Figure 5: The flow of cash register service.

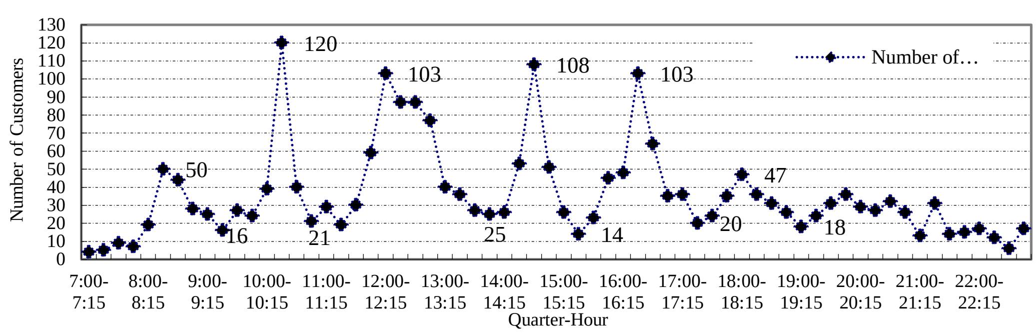

Figure 6: Number of customers appearing every 15 minutes.

Table 7: Selected simulation parameters.

Name	**Parameter**		
	(a) by one clerk	(b) by two clerks	
		1st clerk	2nd clerk
Removing Time	TRIA(4,6.61*n,6.61*n+3.04)	TRIA(2,3.17*n,3.17*n+2.48)	
Scanning Time		-	TRIA(2,3.65*n,3.65*n+2.61)
Picking Time	TRIA(1,2,3)	TRIA(1,2,3)	-
Packing Time	TRIA(2.40,4.29*n,4.29*n+2.15)	TRIA(2.40,4.29*n,4.29*n+2.15)	-
Paying Time	TRIA(8,12,20)	-	TRIA(8,12,20)
Returning Time	TRIA(1.5,2,2.5)	TRIA(1.5,2,2.5)	-

Computer simulation is a methodology that can be used to describe, analyze and predict different effects of alternatives of the performance of a complex business process without the limiting assumptions (Kelton, Smith, and Sturrock, 2014). In Step 6, to improve the customer waiting time during the peak time intervals, scenarios are designed by increasing the number of clerks on duty. For instance, if the clerk capacity is insufficient in the 12:00~13:00 interval in scenario 1, then another salesclerk member needs to be added to the interval time period in which the service level is minimal. Steps 3, 4 and 5 are repeated until the service level is satisfied as shown in scenario 3. Figure 8 shows the number of clerks for each scenario, and Table 8 shows the service level for each scenario compared with the current personnel allocation status.

The proposed procedure for the big data-driven service-level analysis shows the relation of the influencing elements of the service level between the number of customers coming into stores, the frequency of customers, and the average customers service time. In this case, the service level can be improved form 87/60 to 97/60, as shown in Figure 9.

Table 8: Summary for each scenario.

	AS-IS		Scenario 1		Scenario 2		Scenario 3	
Average Service Level	87/60		90/60		94/60		97/60	
Peak Interval	Total Number of Salesclerks	Service Level	Total Number of Salesclerks	Service Level	Total Number of Salesclerks	Service Level	Total Number of Salesclerks	Service Level
10:00-11:00	4	43/60	5	50/60	5	50/60	6	100/60
12:00-13:00	5	38/60	5	39/60	6	99/60	6	100/60
14:00-15:00	4	61/60	5	68/60	5	72/60	5	71/60

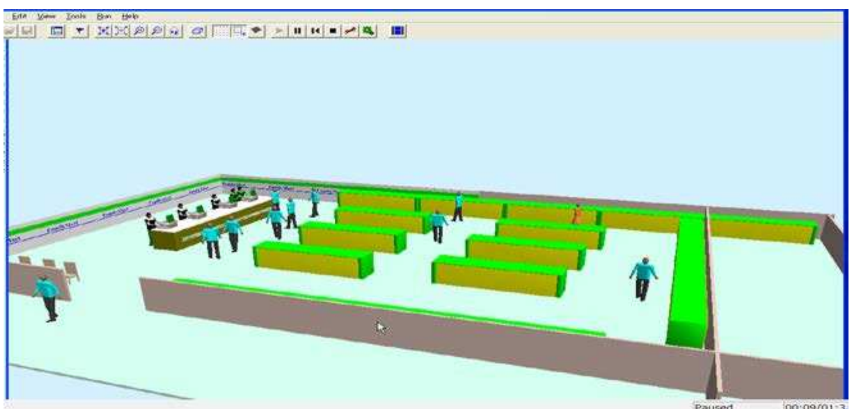

Figure 7: Part of the animation.

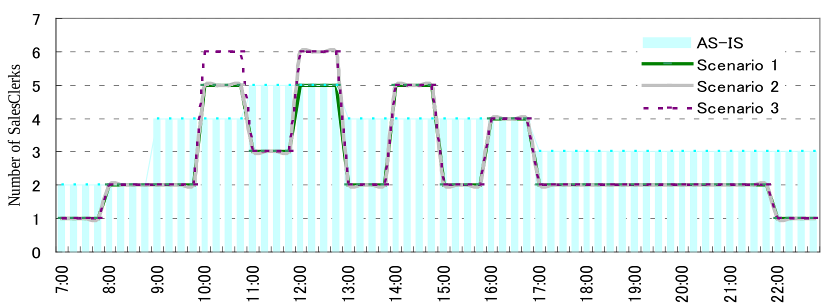

Figure 8: Number of clerks for each scenario.

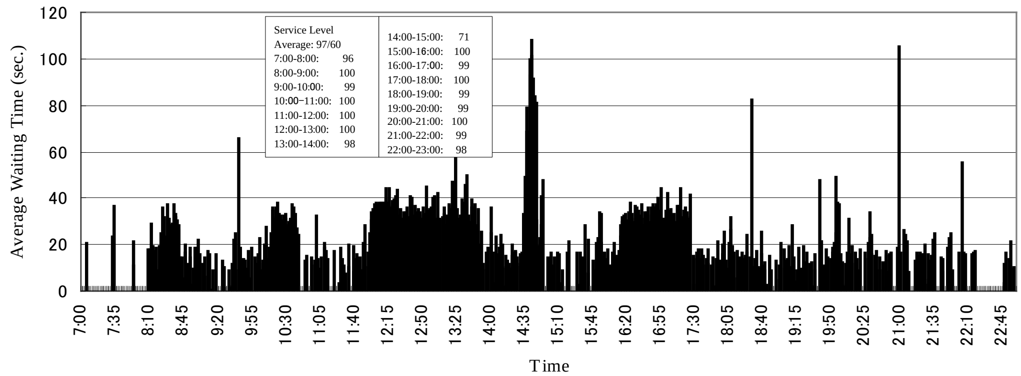

Figure 9: Service level (scenario 3).

4 CONCLUSIONS

(1) Simulation technology with a data generator designed for large-scale POS data can provide analytics to empower even more operation management flexibility in customer service systems.

(2) Procedures to evaluate and estimate service levels using simulation technology are described presenting actual examples using the customer service system of convenience stores. The procedure is generic and can easily be used to examine the service level in the remote past or to analyze and forecast the future. It can also be applied to analyze the effect of the incremental arrival of customers.

(3) It must be stressed that actual large-scale business data such as POS data with advanced simulation analytical technology are effective for analyzing and improving the complicated performance of customer service systems.

ACKNOWLEDGMENTS

The authors wish to express sincere gratitude to FamilyMart Company, Ltd. for their cooperation in completing this research. This research was supported by the Grant-in-Aid for young Scientists (B) of the Japan Society for the Promotion of Science (JSPS). The Grant Number is 23730347.

REFERENCES

Kelton, W. D., J. S. Smith, and D. T. Sturrock. 2014. *Simio and Simulation: Modeling, Analysis, and Applications*. 3rd ed. (Simio LLC). / Takakuwa S. (ed.) (2014), *Simio and Simulation: Modeling, Analysis, and Applications*. 2nd ed. (Japanese ed., Simio LLC).

Kelton, W. D., R. P. Sadowski, and N. Zupick. 2014. *Simulation with ARENA*. 6th ed.. McGraw-Hill, New York.

Liu, Y. 2009. “Personnel Planning of a Retail Store Using POS Data.” *International Journal of Simulation Modelling* 8:185-196.

Miwa, K., and S. Takakuwa. 2008. “Simulation Modeling and Analysis for In-store Merchandizing of Retail Stores with Enhanced Information Technology.” In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 1702-1710. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Miwa, K., and S. Takakuwa. 2010. "Optimization and Analysis of Staffing Problems at a Retail Store." In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-torres, J. Hugan, and E. Yucesan, 1911-1923. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Takakuwa, S., and T. Okada. 2005. "Simulation Analysis of Inbound Call Center of a City-gas Company." In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kulh, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 2026-2033. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Takakuwa, S., and T. Oyama. 2003. "Simulation Analysis of International-departure Passenger Flows in an Airport Terminal," In *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 1627-1634. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Takakuwa, S., and H. Shiozaki. 2004. "Functional Analysis for Operating Emergency Department of a General Hospital." In *Proceedings of the 2004 Winter Simulation Conference*, edited by R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 2003-2011. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Takakuwa, S., and A. Wijewickrama. 2008. "Optimizing Staffing Schedule in Light of Patient Satisfaction for the Whole Outpatient Hospital Ward." In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 1500-1508. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wijewickrama, A., and S. Takakuwa. 2004. "Simulation Analysis of Patient Flows and Staffing Schedules in an Outpatient Department." *Sri Lankan Journal of Management* 9:111-127.

AUTHOR BIOGRAPHIES

RIE GAKU is an Associate Professor in the Graduate School of Business Administration at Momoyama Gakuin University, Osaka, Japan. She received her M. Sc. and Ph.D. Degrees in Economics from Nagoya university in 2006 and 2010. Her research interests include demand forecasting using data mining technology and the simulation analysis of manufacturing and logistics systems. Her current research focuses on the optimization of logistics systems. Her e-mail address is r-gaku@andrew.ac.jp.

SOEMON TAKAKUWA is a Professor at Chuo University in Japan. He received his B.Sc. and M. Sc. degrees in industrial engineering from the Nagoya Institute of Technology in 1975 and the Tokyo Institute of Technology in 1977, respectively. He received a Ph.D. in industrial engineering from Pennsylvania State University. He also holds a Doctorate of Economics from Nagoya University and holds of a P.E. in industrial engineering. He is a corresponding member of the International Academy of Engineering in Russia. His research interests include the optimization of manufacturing and logistics systems, management information systems and simulation analysis of these systems in the context of hospitals. He has prepared the Japanese editions of Introduction to Simulation using SIMAN, Simulation with ARENA, and Simio and Simulation: Modeling, Analysis, Applications. His email address is takakuwa@indsys.chuo-u.ac.jp.