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.

MULTI-OBJECTIVE OPTIMIZATION FOR A HOSPITAL INPATIENT FLOW PROCESS VIA DISCRETE EVENT SIMULATION

Yang Wang
Loo Hay Lee
Ek Peng Chew

Department of Industrial and Systems Engineering
National University of Singapore
1 Engineering Drive 2
117576, SINGAPORE

Sean Shao Wei Lam
Seng Kee Low
Marcus Eng Hock Ong

Health Service Research and Biostatistics Unit
Singapore General Hospital
226 Outram Road
169039, SINGAPORE

Haobin Li

Department of Computing Science
Institute of High Performance Computing
1 Fusionopolis Way #16-16 Connexis
138632, SINGAPORE

ABSTRACT

This paper describes a Discrete Event Simulation (DES) model for a hypothetical inpatient flow process of a large acute-care hospital. The implementation of the Multi-Objectives Convergent Optimization via Most-Promising-Area Stochastic Search (MO-COMPASS) approach in this DES model for the identification of promising Pareto optimal solutions is also discussed. The MO-COMPASS algorithm implemented within the DES modelling paradigm demonstrates how the Multi-Objective Discrete Optimization via Simulation (MDOvS) framework can be applied to identify process improvement opportunities for a hypothetical inpatient boarding processes of a large acute care hospital.

1 INTRODUCTION

Hospital overcrowding often results in prolonged patient waiting times for boarding (Shi et al. 2015), and can significantly impact service quality and patient satisfaction levels (Lo et al. 2014). One common solution to improve inpatient waiting time is to overflow patients across non-primary wards. However, overflowing of patients are not always desirable as it may decrease service quality and increase hospital operational costs (Teow et al. 2012). Several policy levers that are crucial in the inpatient flow processes, such as bed distributions among wards, discharge distributions and overflow thresholds, can be adjusted to reduce the waiting times for boarding.

Deterministic methods have been inadequate for improving inpatient flow processes due to inherent stochasticities, system complexities and patient dynamics (Holm et al. 2013). Harper (2002) demonstrated that simulation approaches would provide better forecasts on the number of bed required than simple deterministic methods. Apart from a realistic reproduction of salient characteristics in the actual system, simulation models could be employed with optimization techniques to identify process improvement opportunities. For instance, Zhang et al. (2012) had analyzed the long-term care capacity planning using

978-1-4673-9743-8/15/$31.00 ©2015 IEEE

a combination of simulation and operations research techniques. Holm et al. (2013) had also modeled bed allocation among hospital wards, and proposed a novel allocation algorithm to minimize hospital overcrowding.

Li et al. (2015) observed that Discrete Optimization via Simulation (DOvS) had become more popular among many industrial sectors recently. DOvS requires a search algorithm to generate feasible solutions, and a simulator to predict system performance for every possible solution (Fu et al. 2005). As most solutions to real-life problems are multi-dimensional, exhaustive search for all possible solutions is not practical. Consequently, the Convergent Optimization via Most-Promising-Area Stochastic Search (COMPASS) was utilized to increase the probability of finding the most preferable ones (Hong and Nelson 2007). The COMPASS algorithm was further extended to consider multiple objective values via the Multi-Objectives COMPASS (MO-COMPASS) algorithm (Li et al. 2015).

In this paper, we described a realistic, but hypothetical, DES model of the inpatient flow processes for an acute care hospital, and demonstrated the successful application of the MO-COMPASS algorithm for the optimization of key controllable factors. Essentially, three policy dimensions were tested using the simulation-optimization framework: 1) bed allocations among wards; 2) discharge distributions, and; 3) overflow thresholds. The main outcome variables are the overflow rates and inpatient waiting times.

2 DISCRETE EVENT SIMULATION MODEL

The simulator used in this paper was built upon an existing simulation model reported in Shi et al. (2015). The existing model had been developed for a large acute care hospital in Singapore. Each ward was assumed to serve one or more specialties as the primary ward, but may receive overflow patients from other disciplines. Waiting time was defined as the duration between an admission decision (i.e., the bed request time) and the patient's arrival at the ward. Overflow rate was defined as the number of admissions to non-primary wards. The existing model provided a well-established framework, including different event types, server and customer definition, and statistic collectors, for this research. The simulator can be implemented at hourly resolutions to capture the time-of-day performance for the effective planning of bed management operations.

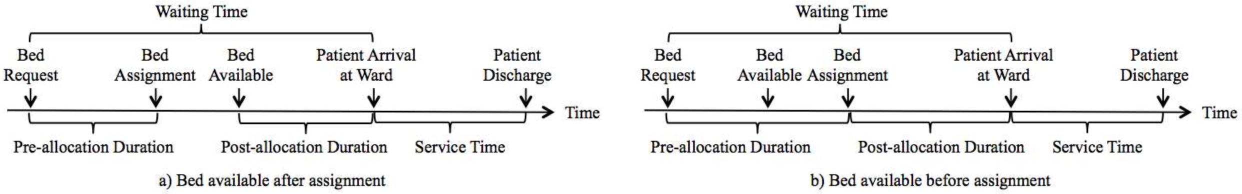
2.1 Admission Sources

The sources of admissions to the general wards (GWs) were assumed to be from the Emergency Department (ED), electives, Same-Day-Admission (SDA), specialist outpatient clinics (SOC) and Intensive Care Unit (ICU). ED patients were those who were required to continue treatment in a GW after meeting a specialist in the ED. Elective and SDA patients were those patients whose admission date was pre-scheduled. Elective patients were assumed to be admitted in the afternoon to receive elective surgeries on the next day, while SDA patients were assumed to arrive in the early morning on the same day of the surgery. SOC patients were those with urgent medical conditions who were admitted directly after consultation in the outpatient clinics. ICU patients were those who were previously admitted into an ICU-type ward from either sources, and then transferred to a GW later. Each patient was only counted when he or she was newly admitted into a GW, while future transfers in and out of GWs were not captured.

2.2 Inpatient Process Flow

The hypothetical inpatient process flow is shown in Figure 1. At the time of patient admission, a bed request was sent to the unit in charge of bed management (UBM). A member of UBM staff will start the pre-allocation process, including bed search and negotiation with wards, after which a bed was reserved for the patient. There were three types of possible bed assignments: normal, upgrade and overflow. Types of bed assignment were dictated by the availability of beds in a particular specialty and its overflow threshold. Once a bed was assigned, the patient will enter post-allocation process, including discharge from previous location, transportation to the assigned ward, etc. Upon arrival at the GW, a patient would receive treatment till his discharge from GW.

- Bed Request: This was defined as the time when a patient is admitted for inpatient treatment. For ED patients, the arrival of bed request was assumed to be a non-stationary Poisson process, since all ED admissions were ad-hoc and had a clear hour-of-day distributions. For other sources, due to the primarily non-adhoc nature, the admissions were assumed to be batched. There were essentially two factors affecting the fluctuations for non-ED admission processes: (i) the number of pre-scheduled bed requests, and; (ii) the actual time when the nurses submit the bed requests within the day.



a) Bed available after assignment

b) Bed available before assignment

Figure 1: Hypothetical Inpatient Process Flow.

- Pre- and Post-allocation Processes: The pre-allocation process referred to the time required for the UBM to search for and identify a bed. Similarly, the post-allocation process was defined as the difference between the time-stamps of bed assignments (or when the bed is available) and the actual arrival of the patient at the ward.
- Bed Assignment: The model assumed two categories of bed assignment policies: 1) normal assignments which referred to the bed assignment to the primary wards of the requested class, and; 2) overflow assignments which occurred only if there was no vacancy in the primary ward. Upgrade assignments, which was to assign a bed of higher class due to non-availability of the lower classes when requested, were not modelled. In the event when there was no primary bed of the desired class available, an upgrade opportunity within the primary ward will be sought under upgrade assignments. Overflow occurs only if there was no vacancy of any class in the primary ward. Since upgrade was less costly and not a major focus of this study, each ward was viewed as a whole and the upgrade assignment was not considered in the model.
- Overflow Policy: Each specialty was assigned a list of primary and overflow wards with a predefined prioritization scheme. Each ward may serve several specialties as primary or overflow ward. A ward was defined to be flexible if it was the primary ward of over 10 specialties. The model can make assignment decision based on the prioritization scheme and ward availability. The possible decision epochs were: the bed request time of a new patient, the departure time of a patient, and the time when an overflow is triggered.
- Overflow Threshold: This referred to a pre-specified waiting duration after which the patient will be overflowed. The threshold was assumed to be flexible over each day. Some possible mean and median thresholds tested in the scenarios for different admission sources are shown in Table 1. ICU patients were overflowed almost immediately as it is costly for a general patient to continue occupying an ICU bed. Elective and SDA patients have pre-scheduled operations and thus they could not wait long for a primary bed.
- Patient Service Time: This quantity was evaluated as follows:

$$\text{Service time} = \text{LOS (Length of Stay)} + \text{Discharge hour} - \text{Admission hour}$$

where

$$\text{LOS} = \text{Discharge day} - \text{Admission day}.$$

Table 1: Mean and Median Overflow Threshold (hours) for Different Admission Sources.

Admission Source	Mean	Median
ED	3.00	2.00
Elective	1.00	0.50
SDA	0.50	0.50
SOC	2.00	0.00
ICU	0.00	0.00

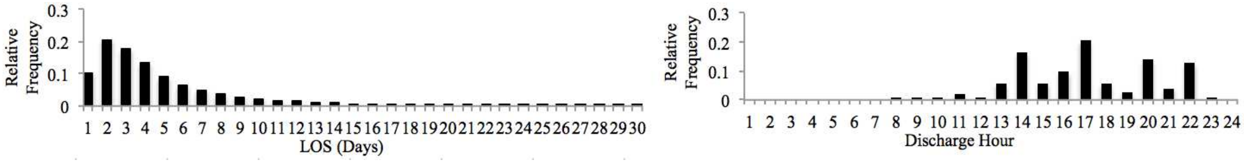


Figure 2: Histograms of LOS and discharge hour for ED patients.

The LOS and discharge hour were determined by the patient admission time, specialty and admission source. An example of the distribution of ED patients' daily LOS and hourly discharge profiles for a weekday morning is shown in Figure2.

2.3 Simulation Result - A Base Model

A base model was constructed by populating the historical data and samples from empirical distribution on admission to the DES model. The durations of service time, pre and post-allocation delays were assumed to follow hypothetical distributions. The simulation batch size was set to be 10,000 days, with the first batch utilized as the simulation warm-up period. The batch means of the subsequent 5 batches were collected. Both the daily overflow rates and average waiting times were monitored and ensured to have converged over these simulation runs.

The average waiting time as predicted by the base model was 3.27 hours for each patient, whereas the average pre- and post-allocation durations were 2.50 hours and 0.51 hour respectively. Therefore, it was evident that the two allocation processes were the dominant components of the total waiting time. The average ED waiting times, pre-allocation and post-allocation durations over the hour of the day in the hypothetical model are plotted in Figure 3. Evidently, the pattern of the ED waiting times closely followed the pre-allocation delay distribution, since the pre- and post-allocation durations constitute almost 90% of the total waiting time.

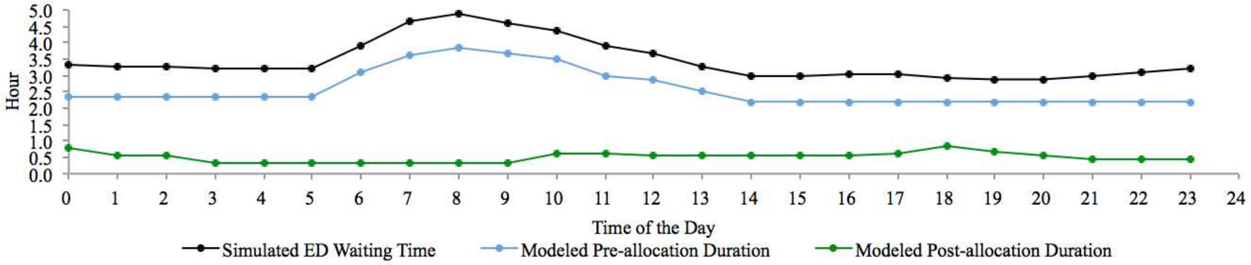


Figure 3: Average ED waiting times, pre-allocation and post allocation delays against patient admission hour.

In this paper, overflow rate was defined as the number of patients overflowed to non-primary wards, and overflow proportion was defined as the proportion of patients overflowed relative to the total admission.

The average overflow rate predicted by the base model was 31.53 per day for all GW admitted patients; this was approximately 15.17% of the average daily admission rate. Figure 4 revealed the variabilities in overflow rates and proportions across various hypothetical specialties typical of a large acute-care hospital. Such a phenomenon could be explained by the fact that the number of beds allocated to each specialty can be better matched to the demand. Although the total number of beds is fixed, the bed distribution among wards can be adjusted dynamically.

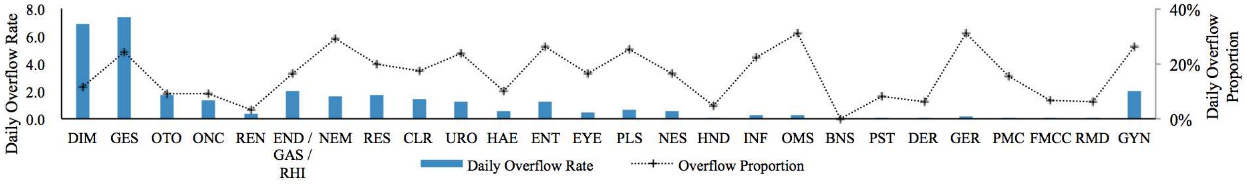


Figure 4: Daily overflow rate and overflow proportion for each patient specialty.

3 OPTIMIZATION ANALYSIS

To improve system performance, a multi-objective optimizer was developed based on the MO-COMPASS algorithm (Li et al. 2015). The MO-COMPASS is a search algorithm responsible for generating candidate solutions by random sampling within a most promising area (MPA). Generally, MPA is a subregion in the solution space in which each points have a shorter Euclidean distance to the "good" solutions (i.e., the Pareto set) rather than to the "bad" ones (i.e., dominated solutions). Li et al. (2015) developed systematic ways to implement this simple idea, proved that the algorithm has strong local convergence, and showed that it can solve benchmark problems efficiently.

Over each iteration in a simulation, the MO-COMPASS samples and evaluates feasible solutions; then it takes in the evaluated results and updates the MPA so that it increases the possibility of sampling better solutions in the next iteration.

In our simulation study, three control factors were evaluated: 1) bed allocation among wards; 2) overflow threshold, and; 3) discharge distribution. Each of these factor was first optimized with respect to the multiple objectives of minimizing overflow rate and inpatient waiting time. Subsequently, the costs associated with the Pareto solutions replaced the inpatient waiting time as one of the optimization objectives for the evaluation of discharge distributions.

3.1 On Bed Allocation

Beds in different wards may serve several specialties as primary or overflow wards, and thus can be viewed as different types of servers in a queueing network. However, as shown in Figure 4, there can be large variabilities in the overflow rates and proportions across different specialties due to the mismatch between supply and demand. Therefore, although the total number of beds was fixed, beds can be redistributed to improve system performance.

For the mathematical formulation of the problem, we first considered K decision variables representing K different general wards (GWs) (i.e., $B = \{b_1, \dots, b_K\} \in \mathbb{N}^K$). Each decision variable represents the number of beds in each GW. Hence, for a hospital with N operational beds, we have $\sum_{i=1}^{K} b_i = N$. With such an equality constraint, all feasible solutions could be located near a hyper-plane in the discretized decision space. This will complicate the sampling procedure in the MO-COMPASS algorithm as solutions could be deemed locally optimal just because its neighbors are infeasible. Considering this technical anomaly, we transformed the decision variables as $X = \{x_1, \dots, x_K\} \in [0,1)^K$ without additional constraints, and map it to B by the following function:

$$b_i = \left\lfloor N \cdot \frac{x_i}{\sum_{j=1}^{K} x_j} \right\rfloor, \forall i \in \{1, \ldots, K\} \tag{1}$$

3.1.1 Minimize Overflow Rate & Waiting Time

Daily overflow rates and waiting times were minimized in this model. For this objective function, two thousand samples were generated and tested independently by the simulator. Figure 5 revealed that the base model performance was dominated by the Pareto set, with the daily overflow rates potentially reducible from 30.73 to 2.10. Although the optimization improved the daily overflow rates by 90% in the hypothetical base model, it has marginal impact on the average waiting times (reduced by just 0.1 hour, or 3%). In addition, tradeoffs between overflow rates and waiting times were small.

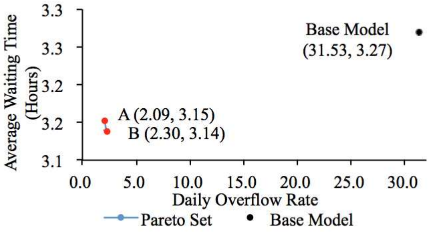


Figure 5: Pareto set of overflow rate versus average waiting time.

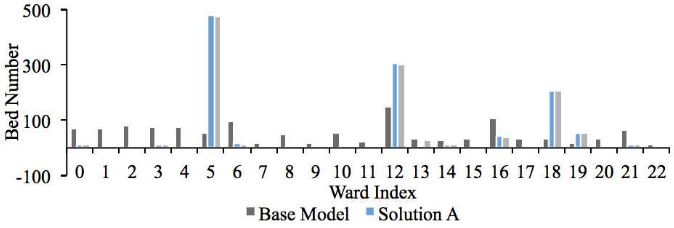


Figure 6: Bed distribution in base model and Pareto solution A and B.

Pareto solutions A and B were plotted in comparison to the base model shown in Figure 6. For these solutions, there was a clear tendency to increase the number of beds in the more flexible wards that can serve more than 10 specialties as primary ward. However, these flexible servers may come at a higher cost compared to other specialized wards, since the nurses in flexible server pools need to be crossed trained and more equipment may be required. In order to incorporate the cost of flexibility into consideration, a new cost function will be introduced in the next section.

3.1.2 Minimize Overflow Rate & Cost of Beds

In consideration of the cost differences in operating flexible beds, costs of beds were included in the objective function. The cost of beds was assumed to be positively correlated with the number of primary specialties it served. Let s_i indicate the number of primary specialties of ward i, a simple cost function can be formulated as follows:

$$C_{\text{Beds}} = \sum_{i=1}^{K} b_i \cdot s_i \tag{2}$$

The Pareto set of $2,000$ candidate solutions for this model is shown in Figure 7. The base model allocation is dominated by the solutions between Points C and D. These dominant solutions can be viewed as potential alternatives that could result in lower overflow rates. The bed distribution in the base model and at Points C and D are shown in Figure 8. Compared to the solutions proposed in Section 3.1.1, bed resources were more evenly distributed for these new solutions.

3.1.3 Minimize Overflow Rate & Cost of Change

In view of the practical difficulties in adjusting the existing bed allocation, the cost of change can be incorporated in the model. The cost of changing bed types was assumed to be incurred whenever there was an increase or decrease in the number of beds in any ward. This cost was assumed to be correlated

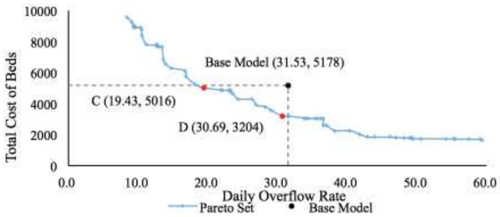


Figure 7: Pareto set of daily overflow rate versus cost of beds.

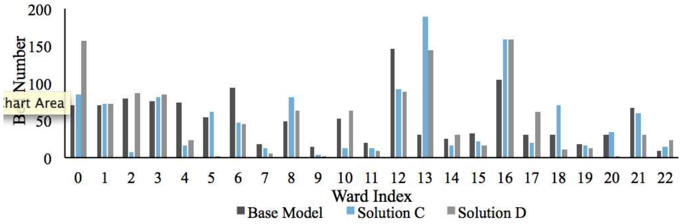


Figure 8: Bed distribution in base model, Pareto solutions A and B.

with the number of primary specialties accommodated in a ward. Let a_i denote the actual number of beds in ward i in the base model. A simple cost function can be formulated as follows:

$$C_{\text{Change}} = \sum_{i=1}^{K} |b_i - a_i| \cdot s_i \tag{3}$$

The Pareto set of $2,000$ candidate solutions is depicted in Figure 9. By changing the bed allocation from the base model to Point F, the overflow performance can be improved by 26.87% with minor costs incurred. Moreover, the flat slope of the Pareto set between Point E and Point F indicated that the cost of change was insensitive to the reduction in overflow rate.

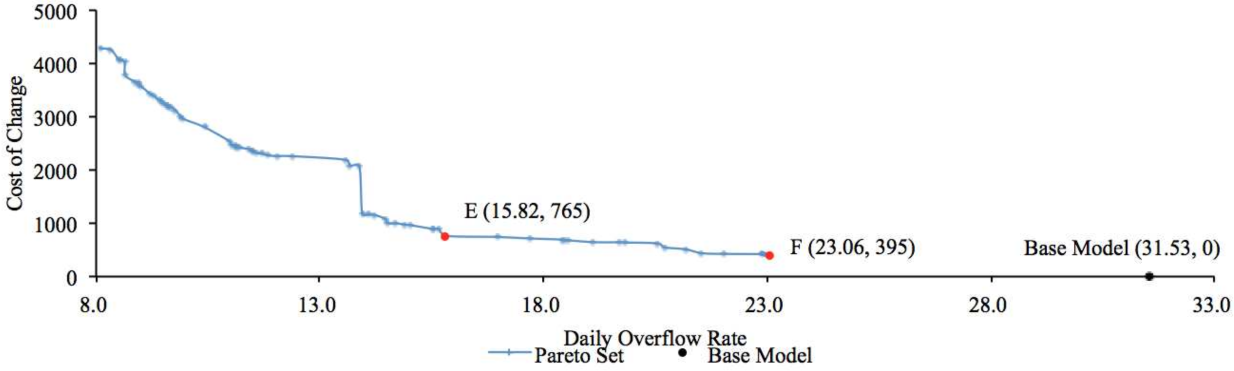


Figure 9: Pareto set of daily overflow rate versus cost of change.

3.1.4 Bed Allocation: Summary

The bed allocation performance of the hypothetical base model and Solutions A to F were summarized in Table **??**. Solutions A and B provided the lowest overflow rates and waiting times, while the cost of beds was significantly higher than all other solutions. Solutions C and D resulted in lower costs of beds, while the costs of changes were much higher than Solutions E and F.

3.2 On Discharge Distribution

Powell et al. (2012) demonstrated that the inpatient boarding performance was affected by the timing of discharges. It was thus hypothesized that moving discharges earlier before the admission peak would improve the inpatient process performance, though early discharges may come at some implementation costs (Shi et al., 2014). In order to test this hypothesis, the discharge distribution was modeled in hourly resolution considering the time period from 8:00 until 23:59. Let the decision variables $P = \{p_9, \ldots, p_{24}\} \in [0,1]^{16}$ denote the probability of patients discharged in each hour of the day. For easy implementation of MO-

Table 2: Summary of bed allocation optimization.

	Base Model	Overflow & Waiting Time		Overflow & Cost of Beds		Overflow & Cost of Change	
	Original Allocation	Solution A	Solution B	Solution C	Solution D	Solution E	Solution F
Overflow Rate	31.53	2.09	2.3	19.43	30.69	15.82	23.06
Average Waiting Time	3.27	3.15	3.14	3.26	3.21	3.18	3.23
Cost of Beds	5178	17815	17600	5016	3204	5405	5299
Cost of Change	0	15617	15374	2182	3110	765	395

COMPASS, we further transformed the decision variables to $Y = \{y_9, \ldots, x_{24}\} \in [0,1)^{16}$, and mapped it back to P by

$$p_i = \frac{x_i}{\sum_{j=9}^{24} x_j}, \forall i \in \{9, \ldots, 24\}. \tag{4}$$

The various combinations of different sets of multiple objectives can then be analyzed for different discharge distributions.

3.3 On Overflow Threshold

In this section, a model with the decision variables being overflow thresholds which can dynamically change over the day, and objectives as overflow rates and average waiting times will be discussed. In the model, we let $G = \{g_1, \ldots, g_{24}\} \in [0,10]^{24}$ be the decision variables, in which each g_i denotes the overflow threshold in the ith hour of the day.

The Pareto set of $2,000$ candidate solutions is shown in Figure 10. There was a clear trade-off between overflow rate and average waiting time with respect to the changes in overflow threshold. The thresholds proposed by Solutions L and M presented in Figure 11 tended to be conservative because the dominant components of waiting time (i.e., the pre-and post-allocation durations) were independent of the threshold changes. An aggressive threshold policy was expected to increase the overflow rates, but such a policy will not significantly reduce the average waiting time.

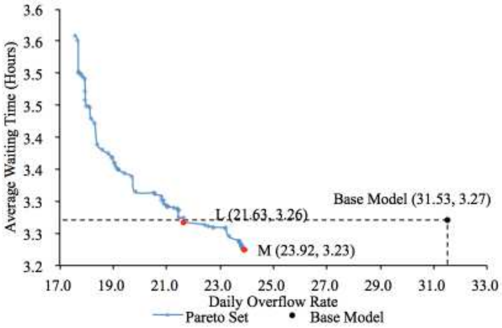


Figure 10: Pareto set of daily overflow rate vs. average waiting time.

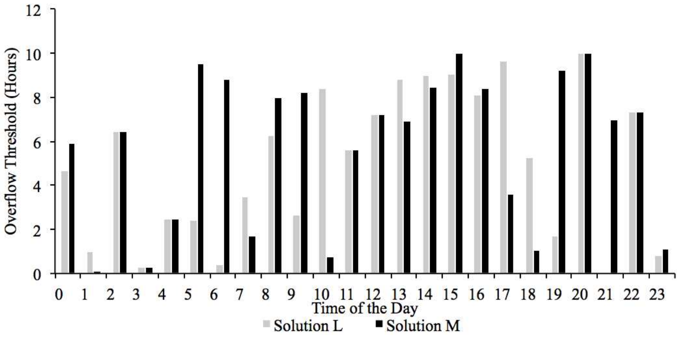


Figure 11: Overflow thresholds for Pareto Solutions A and B.

4 Conclusion

This study described the development of a hypothetical inpatient flow process model based on the DES modelling paradigm. The MO-COMPASS algorithm was successfully implemented for the optimization of three sets of paired opbjectives. The current model has made a balance between analytical tractability and fidelity for a realistic but hypothetical model. Such a model can be further generalized to include more generic complexities in inpatient boarding processes, such as different bed classes and gender assignments. Furthermore, the cost functions proposed in this research are still extremely primitive and can be further specialized based on the specific policy preferences.

ACKNOWLEDGMENTS

We would like to express our gratitude to Professor Jim Dai and Assistant Professor Shi Pengyi. This research would not have been possible without their contributions.

REFERENCES

Fu, M. C., F. W. Glover, and J. April. 2005. “Simulation Optimization: A Review, New Developments, and Applications”. In *Proceedings of the 37th conference on Winter simulation*, 83–95. Winter Simulation Conference.

Harper, P. R. 2002. “A Framework for Operational Modeling of Hospital Resources”. *Health care management science* 5 (3): 165–173.

Holm, L. B., H. Lurås, and F. A. Dahl. 2013. “Improving Hospital Bed Utilization through Simulation and Optimization: With Application to a 40% Increase in Patient Volume in a Norwegian General Hospital”. *International journal of medical informatics* 82 (2): 80–89.

Hong, L. J., and B. L. Nelson. 2007. “A Framework for Locally Convergent Random-search Algorithms for Discrete Optimization via Simulation”. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 17 (4): 19.

Li, H., L. H. Lee, E. P. Chew, and P. Lendermann. 2015. “MO-COMPASS: A Fast Convergent Search Algorithm for Multi-Objective Discrete Optimization via Simulation”. *IIE Transactions* 47:1–17.

Lo, S. M., K. T. Y. Choi, E. M. L. Wong, L. L. Y. Lee, R. S. D. Yeung, J. T. S. Chan, and S. Y. Chair. 2014. “Effectiveness of Emergency Medicine Wards in Reducing Length of Stay and Overcrowding in Emergency Departments”. *International emergency nursing* 22 (2): 116–120.

Powell, E. S., R. K. Khare, A. K. Venkatesh, B. D. Van Roo, J. G. Adams, and G. Reinhardt. 2012. “The Relationship Between Inpatient Discharge Timing and Emergency Department Boarding”. *The Journal of emergency medicine* 42 (2): 186–196.

Shi, P., M. Chou, J. Dai, D. Ding, and J. Sim. 2015. “Models and Insights for Hospital Inpatient Operations: Time-dependent ED Boarding Time”. *Management Science* (Published online: http://dx.doi.org/10.1287/mnsc.2014.2112).

Teow, K. L., E. El-Darzi, C. Foo, X. Jin, and J. Sim. 2012. “Intelligent Analysis of Acute Bed Overflow in a Tertiary Hospital in Singapore”. *Journal of medical systems* 36 (3): 1873–1882.

AUTHOR BIOGRAPHIES

YANG WANG is an undergraduate student in the Department of Industrial and Systems Engineering at National University of Singapore. Her email address is yang.wang117@gmail.com.

SEAN SHAO WEI LAM is currently the Manager of the Health Services Research Unit in Singapore General Hospital and an Associate with the Centre for Quantitative Medicine. He received his PhD and Masters in Industrial and Systems Engineering from the National University of Singapore. The primary focus of his research is on real-world healthcare operation research problems for the improvements of health services and healthcare operations. His email address is lam.shao.wei@sgh.com.sg.

HAOBIN LI is Scientist in Institute of High Performance Computing, under Agency for Science, Technology and Research (A*STAR) of Singapore. He received his B.Eng. degree (1st Class Honors) in 2009 from the Department of Industrial and Systems Engineering at National University of Singapore, with minor in computer science; and Ph.D. degree from the same department in 2014. He has research interests in operations research, simulation optimization and designing high performance optimization tools which are ready for practical industrial use. His email address is lihb@ihpc.a-star.edu.sg.

LOO HAY LEE is Associate Professor and Deputy Head in the Department of Industrial and Systems Engineering, National University of Singapore. He received his B. S. (Electrical Engineering) degree from the National Taiwan University in 1992 and his S. M. and Ph. D. degrees in 1994 and 1997 from Harvard University. He is currently a senior member of IEEE, a committee member of ORSS, and a member of INFORMS. His research interests include production planning and control, logistics and vehicle routing, supply chain modeling, simulation-based optimization, and evolutionary computation. His email address is iseleelh@nus.edu.sg.

EK PENG CHEW is Associate Professor and Deputy Head in the Department of Industrial and Systems Engineering, National University of Singapore. He received his Ph. D. degree from the Georgia Institute of Technology. His research interests include logistics and inventory management, system modeling and simulation, and system optimization. His email address is isecep@nus.edu.sg.

SENG KEE LOW is Manager of the Bed Management Unit of Singapore General Hospital. He received his Master of Business in Professional Accounting in 1998 from Victoria University. His email address is low.seng.kee@sgh.com.sg.

MARCUS ENG HOCK ONG is a Senior Consultant, Director of Research, and Clinician Scientist at the Department of Emergency Medicine in Singapore General Hospital. He is also the Director for the Health Services Research and Biostatistics Unit, Division of Research, in SGH. He serves as a Senior Consultant at the Hospital Service Division and Director for the Unit for Pre-hospital Emergency Care in Ministry of Health, Singapore. He is also Associate Professor at Duke-National University of Singapore Graduate Medical School. His research studies focus predominantly on pre-hospital emergency care, medical devices, and health services research. His email address is marcus.ong.e.h@sgh.com.sg.