# TRADE-OFF BALANCING FOR STABLE AND SUSTAINABLE OPERATING ROOM SCHEDULING 

Amin Abedini<br>University of Kentucky, a.abedini@uky.edu<br>Digital Object Identifier: https://doi.org/10.13023/etd.2019.346

Right click to open a feedback form in a new tab to let us know how this document benefits you.

## Recommended Citation

Abedini, Amin, "TRADE-OFF BALANCING FOR STABLE AND SUSTAINABLE OPERATING ROOM SCHEDULING" (2019). Theses and Dissertations--Mechanical Engineering. 141.
https://uknowledge.uky.edu/me_etds/141

This Doctoral Dissertation is brought to you for free and open access by the Mechanical Engineering at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Mechanical Engineering by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

## STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Amin Abedini, Student<br>Dr. Wei Li, Major Professor<br>Dr. Alexandre Martin, Director of Graduate Studies

# TRADE-OFF BALANCING FOR STABLE AND SUSTAINABLE OPERATING <br> ROOM SCHEDULING 

| DISSERTATION |
| :---: |

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Engineering at the University of Kentucky

By<br>Amin Abedini<br>Lexington, Kentucky

Co-Directors: Dr. Wei Li, Professor of Mechanical Engineering
and Dr. Fazleena Badurdeen, Professor of Mechanical Engineering Lexington, Kentucky 2019

# ABSTRACT OF DISSERTATION 

## TRADE-OFF BALANCING FOR STABLE AND SUSTAINABLE OPERATING ROOM SCHEDULING

The implementation of the mandatory alternative payment model (APM) guarantees savings for Medicare regardless of participant hospitals ability for reducing spending that shifts the cost minimization burden from insurers onto the hospital administrators. Surgical interventions account for more than $30 \%$ and $40 \%$ of hospitals total cost and total revenue, respectively, with a cost structure consisting of nearly $56 \%$ direct cost, thus, large cost reduction is possible through efficient operation management. However, optimizing operating rooms (ORs) schedules is extraordinarily challenging due to the complexities involved in the process. We present new algorithms and managerial guidelines to address the problem of OR planning and scheduling with disturbances in demand and case times, and inconsistencies among the performance measures. We also present an extension of these algorithms that addresses production scheduling for sustainability. We demonstrate the effectiveness and efficiency of these algorithms via simulation and statistical analyses.

KEYWORDS: Operating room, Efficiency, Scheduling, Trade-off balancing

Author's signature:

# TRADE-OFF BALANCING FOR STABLE AND SUSTAINABLE OPERATING ROOM SCHEDULING 

By<br>Amin Abedini

Co-Directors of Dissertation: Wei Li

Fazleena Badurdeen

Director of Graduate Studies:
Alexandre Martin
Date:
July 31, 2019

## ACKNOWLEDGMENTS

I am greatly grateful to my mentor and advisor, Professor Wei Li whose footprints in my life and career never be faded.

I would like to express my deepest gratitude to my dissertation committee, Professor I.S. Jawahir, Professor Fazleena Badurdeen, and Professor Yuming Zhang, for their unconditional scientific and emotional supports.

I would like to thank my forever friend Mr. Hongan Ye who has always been a great deal of help and care.

I would like to thank Mrs. Brittany L. Nett for proofreading my dissertation and her constructive comments.

## TABLE OF CONTENTS

Acknowledgments ..... iii
Table of Contents ..... iv
List of Figures ..... vii
List of Tables ..... ix
Chapter 1 Introduction and Motivation ..... 1
1.1 Problem statement ..... 3
1.2 Motivation ..... 6
1.3 Summary of Contribution ..... 6
Chapter 2 Reconciling Supply and Demand: Stochastic Optimization Models for Efficient Operating Room Planning ..... 10
2.1 Introduction ..... 11
2.2 Problem formulation ..... 13
Bi-level optimization models ..... 14
P1: Total cost to throughput ..... 14
P2: Throughput to Total cost ..... 16
P3: Trade-off balancing model ..... 18
2.3 Case study ..... 18
Data generation ..... 18
Cost ratios ..... 19
Managerial preferences ..... 19
Performance measures ..... 20
2.4 Results ..... 20
Idletime ..... 21
Overtime ..... 21
Total cost ..... 22
Throughput ..... 23
Trade-off balancing ..... 23
2.5 Conclusion ..... 24
Chapter 3 Tactical Operating Room Planning Under Surgery Type and Priority Constraints ..... 26
3.1 Introduction ..... 26
3.2 Problem Formulation ..... 28
3.3 Case studies ..... 30
3.4 Results and Discussion ..... 31
Number of ORs and Number of Setups ..... 31
Idle time and Overtime ..... 32
Utilization and Smoothness Index ..... 33
Total Cost ..... 34
3.5 Conclusion ..... 34
Chapter 4 Operational Operating Room Scheduling by Smoothing Patient Flow ..... 36
4.1 Introduction ..... 36
4.2 Problem description ..... 38
Total completion time (TCT) ..... 39
Total absolute deviations in completion times ( $T A D C$ ) ..... 40
4.3 Inconsistency between $\min (T C T)$ and $\min (T A D C)$ ..... 44
Trade-off balancing ..... 44
4.4 Optimal sequence to the trade-off function ..... 45
Bi-criteria models ..... 47
4.5 Case studies ..... 47
Evaluation schemes ..... 47
Pareto dominance ..... 47
Modern portfolio theory ..... 48
4.6 Results and Discussions ..... 48
Pareto dominance ..... 48
Modern portfolio theory ..... 49
4.7 Conclusion ..... 52
Chapter 5 An Optimization Model for Operating Room Scheduling to Reduce Blocking Across the Perioperative Process ..... 53
5.1 Introduction ..... 53
5.2 Problem Formulation ..... 56
5.3 Case Study ..... 58
Optimal MSS ..... 58
Simulation ..... 59
Statistical Process Control (SPC) ..... 61
5.4 Conclusion ..... 62
Chapter 6 Risk Management in Surgical Case Sequencing by Balancing Trade- offs Among Inconsistent Objectives ..... 64
6.1 Introduction ..... 64
Summary ..... 67
6.2 Problem description ..... 68
6.3 Case studies ..... 70
6.4 Results and Discussions ..... 71
Loss expected value(P.1) ..... 72
Loss variance(P.2) ..... 73
Conditional value at risk (P.3) ..... 75
Overall evaluation ..... 75
6.5 Conclusion ..... 76
Chapter 7 Trade-off Balancing for Sustainable Production Scheduling: A Metric- Based Approach ..... 79
7.1 Introduction ..... 79
7.2 Literature review ..... 81
Economic-oriented scheduling ..... 81
Environment-oriented scheduling ..... 83
Social-oriented scheduling ..... 83
Summary ..... 83
7.3 Problem formulation ..... 84
7.4 Case study ..... 85
Inconsistencies among objectives ..... 87
Pareto dominance ..... 87
Production scheduling sustainability ..... 88
Results and discussion ..... 89
Production scheduling Sustainability ..... 89
Sustainability at sub-cluster level ..... 92
Fundamental performance metrics ..... 94
7.5 Conclusion ..... 97
Chapter 8 Conclusions and Future Work ..... 99
8.1 Conclusions ..... 99
8.2 Direction for future works ..... 103
Chapter A Appendices ..... 104
A. 1 Proof of Lemmal3 ..... 104
A. 2 Proof of Lemma 4 ..... 105
Bibliography ..... 106
Vita ..... 119

## LIST OF FIGURES

1.1 Schematic representation of OR peri-operative process ..... 4
2.1 Overtime threshold to planned capacity ratio vs. number of planned patients ..... 16
$2.2 \quad \gamma n$ threshold vs. $n$ ..... 17
2.3 Anderson-Darling normality tests ..... 19
2.4 SPC results for Idletime ..... 21
2.5 SPC results for Overtime ..... 22
2.6 SPC results for total cost ..... 22
2.7 SPC results for the number of patients served in overtime ..... 23
2.8 Sum of average deviations from the best $T C$ and $T P$ ..... 24
3.1 Histograms of number of ORs and number of setups ..... 32
3.2 Histograms of idle time and overtime and ..... 33
3.3 Histograms of smoothness index and utilization ..... 33
3.4 Total cost with different setup costs ..... 34
4.1 $\quad V$-shaped and non- $V$-shaped processing sequences ..... 40
$4.2 \quad \Lambda$-shaped processing sequence ..... 42
4.3 A non- $\Lambda$-shaped processing sequence ..... 43
4.4 Pareto frontiers based on $D_{1}$ and $D_{2}$ ..... 49
4.5 Grand average normalized deviations of sequencing methods ..... 50
4.6 Portfolio frontiers of trade-off balancing functions ..... 50
4.7 Trade-off balancing functions Loss average and Standard deviations over all port- ..... 51
5.1 An example of MSS ..... 54
5.2 The patients' path in the peri-operative process ..... 55
5.3 A graphical representation of the $B M$ model optimal $M S S$ ..... 59
5.4 The average daily ICU occupancy ..... 60
5.5 Dotplot of blockings for the studied models ..... 60
5.6 Capabilities of the studied model for the number of blockings ..... 61
5.7 SPC charts of average weekly number of blockings ..... 62
6.1 Schematic representation of an OR peri-op process ..... 65
6.2 Average Normalized Deviations of sequencing methods ..... 71
6.3 Loss expected value of the sequencing methods for different portfolios ..... 72
6.4 Loss average and standard deviation frontiers, each point shows a portfolio ..... 73
6.5 Scheduling methods loss average and standard deviation over all portfolio ..... 74
6.6 Loss expected value and standard deviation frontiers of non-dominated methods, each point shows a portfolio ..... 74
$6.7 \quad \alpha-C V a R$ of sequencing methods ..... 77

| 7.1 | Scatter plots of normalized deviations of 66 weights, Pareto-dominant solution |
| :--- | :--- | :--- |
| are shown by red markers $. . .2 . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~$ | 88 |

7.2 PSSI Pareto frontier. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 90
$7.3 \quad P S S I$ process capability charts for alternative production schedules . . . . . . . 93
7.4 Sustainability scores at sub-cluster level . . . . . . . . . . . . . . . . . . . 93
7.5 Sub-cluster sustainability score Pareto frontier . . . . . . . . . . . . . . . . . . 94
8.1 Adaptive control structure for OR planning and scheduling . . . . . . . . . 103

## LIST OF TABLES

1.1 Summary of Contributions ..... 9
2.1 Statistics obtained from UKHealthcare data ..... 18
$2.2\left(B^{*}, n^{*}\right)$ of different combinations of different $\alpha$ and $T L$ ..... 20
$2.3\left(B^{*}, n^{*}\right)$ of different combinations of different $\beta$ and ..... 20
2.4 ANOVA results for total cost between $C 2 P-1$ and $P 2 C-15$ ..... 22
2.5 Sum of average deviations from the best $T C$ and $T P$ ..... 24
3.1 Urgency Related Groups ..... 27
3.2 Computation results for PTD and $L P T$ ..... 31
4.1 An example for demonstrating Algorithm 4 ..... 46
4.2 Statistics of average normalized deviations ..... 49
4.3 $\quad$ Statistics of loss generated by sequencing methods over all portfolios ..... 51
5.1 Historical data of the surgery groups ..... 58
5.2 The optimal $M S S$ obtained from the $B M$ model ..... 58
5.3 Daily occupancy of the post-op stage ..... 60
5.4 SPC and capability results of weekly number of blockings ..... 61
6.1 Expected loss of 13 scheduling methods with different portfolios ..... 72
6.2 Loss standard deviation of the scheduling methods with different portfolios ..... 73
$6.3 \quad \alpha-C V a R$ of sequencing methods at $\alpha \in\{0.90,0.95,0.99\}$ ..... 76
6.4 Performance of selected sequencing methods at $\alpha=0.95$ ..... 77
7.1 Metric-based hierarchical decomposition for sustainable production scheduling ..... 84
7.2 Spearman's $\rho$ among sequence for scheduling objectives ..... 87
7.3 PSSI weight assignment ..... 88
7.4 PSSI statistics of trade-off functions $\left(\min \left(z^{[\Omega]}\right)\right.$ ..... 90
7.5 Normalized deviation statistics of trade-off functions $\left(\min \left(z^{[\Omega]}\right)\right)$ ) ..... 91
7.6 PSSI for alternative production schedules ..... 91
7.7 PSSI capabilities indices for alternative production schedules ..... 92
7.8 Statistics of sustainability scores at sub-cluster level ..... 93
7.9 Average performance and control limits of fundamental performance metrics ..... 95
$7.10 L_{1}$ average performance and control limits ..... 96
7.11 Specification limits of performance indicators ..... 96
7.12 Capability results for the scheduling alternatives ..... 96
8.1 Summary of Conclusions ..... 102

## Chapter 1

## Introduction and Motivation

In this chapter, we elaborate the need for effective and efficient management strategies to improve the efficacy of operating rooms. We show how efficient operating room contributes to a larger picture in which health care quality is improved and the deficit reduction is achieved.

The business environment for hospitals has been rapidly changing due to privatization, reimbursement constraints, aging population, and social awareness. These driving forces propel hospitals administrators to improve the quality of care while reducing costs which are typically in conflicts. Health care expenditures account for approximately $18 \%$ of the United States Gross Domestic Product (GDP) [1]. Experts in economy, healthcare policy, and public finance believe that the healthcare expenditures control is the main component of the deficit reduction challenge facing the U.S. government [2]. Surgical interventions are responsible for nearly $22 \%$ of hospital stays [3] but account for more than $40 \%$ of the total expenditure [4] with average cost of $\$ 21,200$ per stay in 2012 [5]. Operating rooms (ORs) also drive many other areas within the hospital such as post anaesthesia care units (PACUs), intensive care units (ICUs), wards, etc. [6]. Statistics show that individuals aged 45-84 are accounted for approximately $52 \%$ of hospital stays and $66 \%$ of hospital costs [7], moreover, with the aging population the demand for surgical stays is sharply increasing [8]. On the other hand, the rate of OR capacity expansion is lower than the rate of increasing demand. Thus, OR managers must leverage their resources by efficient OR planning and scheduling strategies. Significant improvements can be achieved through efficient and effective OR schedules. However, OR scheduling is an extraordinarily complex and challenging task due to the conflicts among stakeholders' interests, conflicts among strategic, tactical, and operational decisions, and the highly dynamic nature of the system. Therefore, the aim of OR management is to balance the trade-offs among conflicting objectives.

As Zeleny and Cochrane (1973) [9] point out, studying a system as a single-attribute problem is a search and measurement process rather than being a decision making process. The decision making problem uprises once the optimum decision is determined by multiple criteria. Hence, the main purpose of this dissertation is analyzing the conceptual and operational features of decision making process in the operating room planning and scheduling problem that involves multiple conflicting objectives.

A wide range of performance measures have been introduced to evaluate the performance of OR plans and schedules such as throughput [10, 11, 12, 13, 14], waiting time [15, 16, 17, 18, 19], utilization [19, 20, 21, 22, 23], cost [24, 25, 26, 27, 28], etc. Literature also vary according to the applied research methodologies. Majority of the problems have been formulated as combinatorial optimization models and many exact, heuristics, and scenario-based approaches have been proposed.

The majority of studies model the ORs as an isolated unit. This approach is based on the fundamental assumption that the OR is the bottleneck of the peri-operative process and the upstream and downstream resources abound. This assumption removes the resource constraints imposed on the models and reduces their complexity. However, in reality, it may shift the bottleneck from OR to other stages that results in OR blocking and/or starvation. OR blocking/starvation negatively impact the utilization of the ORs and the peri-operative process as a whole.

Guido and Conforti (2017) [29] proposed a multi-objective model for integrated OR planning and scheduling. A hybrid method based on genetic algorithms was developed to solve the proposed combinatorial model. A set of Pareto optimal solution are proposed to support the decision making process. Addis et al. (2016) [30] proposed an approach for integrated offline and online surgery scheduling. The objective is to minimize the cost of waiting times, urgency, and tardiness of the patients. The results of the integrated offline-online model and the offline model are compared. The results showed the superiority of the offline-online model, however, the computation time was significantly higher than that of the offline model. Robinson and Chen (2003) [31] studied a scheduling problem with random case times with the objective of minimizing patient waiting time and doctor's idle time. A sample average approximation approach is used to solve the stochastic linear programming model and a closed-form heuristic to set up surgery start times is proposed with worst-case performance usually within $20 \%$ of the optimal. Zhang and Xie (2015) [32] studied the scheduling of a sequence of surgeries with random case times in a multiple operating rooms context. Surgeries are assigned to ORs dynamically on a first-come, first-serve (FCFS) basis. A discreteevent framework is proposed to proactively predict the surgery start times with the objective of minimizing the total cost incurred by surgeon waiting, OR idling, and OR overtime. The problem is solved by stochastic approximation. Numerical experiments showed that the stochastic approximations converge to a unique global optima. Duenas et al. (2016) [33] proposed a compromise programming model with three objectives including minimizing operating room costs, minimizing the maximum number of required nurses, and minimizing the number of open operating rooms. They proposed a set of non-dominated solutions for the studied OR scheduling problem. Aringheiri et al. (2015) [34] studied the joint OR planning and scheduling problem. The aim is to determine, over a planning horizon, the allocation of OR time blocks to specialties together with the subsets of patients to be scheduled within each time block. The objective function is to optimize both patient utility (by reducing waiting time costs) and hospital utility (by reducing production costs measured in terms of the number of weekend stay beds required by the surgery planning). A metaheuristic is developed to solve the proposed $N P$-hard 0-1 linear programming model.

Modeling the OR peri-operative process is more complex compared to modeling the isolated ORs. Jebali and Diabat (2017) proposed a chanced-constrained stochastic
programming model in which the downstream resource constraints (ICUs) are integrated into the model. The variations in demand (emergency arrivals) are also taken into consideration. The objective is to minimize the cost to the patients, cost of OR over/underutilization, and surgery cancellation. An Average Approximation (AA) algorithm is proposed to solve the combinatorial model. Their results shows that higher scheduling robustness is achieved at the cost of lower OR utilization and higher overall cost. Marcon and Dexter (2006) [35] studied the impact of surgical case sequencing on the OR overtime and PACU staffing. Multiple sequencing methods are compared over a wide range of scenarios. They recommended against the Longest Case First ( $L C F$, also known as $L P T$ ) rule that leads to over-utilized OR and the requirement of more PACU nurses. MIX method is recommended for surgical case sequencing, $M I X$ method is a mixture of two simple rules for sequencing, the $L C F$ rule and the Shortest Case First ( $S C F$ also known as $S P T$ rule), where the longest or shortest case is evaluated by maximum or minimum of case times. However, in another study, Marcon and Dexter (2007) [36] concluded that the LCF rule did not perform worse than other methods, including the MIX method, and the overall process behaved as if there was random sequencing, given significant variations in case times from the expected ones, where random sequencing can be interpreted as the first come first serve $(F C F S)$ rule. Vissers et al. (2005) 37] proposed a mixed integer programming model for OR planning in order to generate an optimal case mix in a cardiothoracic surgery department. They studied the impacts of decreasing or increasing the number of resources including ICUs on the outcomes of the surgery department such as utilization and throughput. They proposed optimum scenarios of number resources and patient mix for the studied surgery department. Pham and Klinkert (2008) [38] modeled the surgical case scheduling problem as an extension of job shop scheduling problem. A mixed integer linear programming model is proposed with the objective of optimizing utilization. They point out that the surgical case sequencing must be performed considering all activities involved in the process, in other word, surgical case sequencing should take a holistic view of all activities and resource constraints in the OR suite instead of focusing on only an individual stage such as an OR or ICU.

In order to avoid enlarging this introductory chapter, detailed and specific literature reviews are provided in each chapter.

### 1.1 Problem statement

An OR peri-operative (peri-op) process normally consists of three sequential stages of preoperative (pre-op), intra-operative (intra-op), and post-operative (post-op), as shown by Figure 1.1, where the following activities occur: patient preparation in the pre-op stage, surgical intervention in the intra-op stage, and post anesthesia care in the post-op stage.

The stages in the OR peri-op process are tightly coupled, thus, any disturbance at any stage flows across the OR peri-op process boundaries which negatively affect the outcomes of the process. Many key performance indicators (KPIs) including utilization, patient flow, waiting time, overcrowdings, quality of care, etc. have been identified to illustrate the impacts of ORs on the overall healthcare systems. It is worth mentioning that the KPIs are frequently translated into financial terms to demonstrate the link


Figure 1.1: Schematic representation of OR peri-operative process
of healthcare systems and the national economy. The prime goal of OR planning and scheduling is to simultaneously improve the healthcare quality and hospitals solvency. Of course, the outcomes of OR peri-op process significantly depend upon the efficacy of the applied operation management strategies and techniques.

Regardless of the studied system, the operation management strategy can be expressed by a linear programming (LP) model in the canonical form as shown by Equations (1.1) to (1.3). Equation (1.1) represents the objective functions, Equations 1.2 ) and (1.3) are the constraints, where $\mathbf{x}=\left[x_{1}, x_{2}, \ldots, x_{K}\right]$ is the vector of decision variables, $\mathbf{c}=\left[c_{1}, c_{2}, \ldots, c_{K}\right]$ is the coefficients of the objective function (also known as weights), $\mathbf{A}$ is a $P \times K$ matrix, and $\mathbf{b}=\left[b_{1}, b_{2}, \ldots, b_{P}\right]$ is the vector of nonnegative constants.

$$
\begin{array}{cc}
\min & \mathbf{c}^{T} \mathbf{x} \\
\mathbf{A x}<\mathbf{b} \\
& \mathbf{x} \geq 0 \tag{1.3}
\end{array}
$$

The traditional approach for modeling the decision making process is based on the fundamental assumption that the objective function is a well-defined single-attribute function. Although, this approach is logically correct, its tenability is undermined by the fact that in reality the decision maker's goal is to achieve an optimal balance among multiple objectives, many of which are competing/conflicting. Examples for multicriteria decision making can be found almost in any area of the business environment and the daily life as well. Operating room planning and scheduling is no exception where with many stakeholders the existence of the conflicting objectives is the rule rather than an exception. Another assumption in the traditional approach is the rigidity of the constraints such that any violation is not allowed. However, in reality, it is always possible to take a certain level of violation at least in one of the constraints, for example, in the OR context, surgeons may start performing a surgery even though a portion of the operation may be done in overtime, meaning that a small amount of overtime trade offs for a large amount of idletime.

A system of linear equations is called inconsistent if it has no solutions [39] which means with inconsistencies among objectives, there is no solution that simultaneously optimizes all objectives. Thus, any solution generates some levels of trade-offs in the system. Therefore, balancing trade-offs is a meaningful and practical objective.

We use compromised programming $(C P)$, which was first proposed by Zeleny and Cochrane (1973) [9] to formulate the objective function of the trade-off balancing approach. The first step in $C P$ is to establish an 'ideal point', the coordinates of the
ideal point are given by the optimum values $\left(L B_{k}\right)$ of all objectives. It is obvious that with inconsistent objectives the ideal point is not feasible, therefore, the ideal point is only a point of reference for $C P$. The second step in $C P$ is to establish an 'anti-ideal' point. The coordinates of the anti-ideal point are given by the worst values $\left(U B_{k}\right)$ of all objectives. The objective of $C P$ is to find the closest efficient solution to the ideal point [40]. In the minimization sense, the coordinates of the ideal point and the antiideal point are $\left(L B_{1}, L B_{2}, \ldots, L B_{k}\right)$ and $\left(U B_{1}, U B_{2}, \ldots, U B_{k}\right)$, respectively. Therefore, the degree of closeness between solution $\boldsymbol{\sigma}$ to the coordinate $k$ of the ideal point is given by $y_{k}(\boldsymbol{\sigma})=\frac{\gamma_{k}(\boldsymbol{\sigma})-L B_{k}}{U B_{k}-L B_{k}}$ which is known as the normalized deviation from the best value.

To measure the distances between outcomes of the solution $\boldsymbol{\sigma}$ and the ideal point, a family of distance functions are introduced by $L_{g}\left(\alpha_{k}, \boldsymbol{\sigma}\right)=\left(\sum_{k=1}^{K}\left(\alpha_{k} y_{k}(\boldsymbol{\sigma})\right)^{g}\right)^{1 / g}$, where $\alpha_{k}$ is the weight of attribute $k$ in the objective function, without loss of generality we assume $\sum_{k=1}^{K} \alpha_{k}=1$. When $g=1, L_{1}$ measures the longest distance (geometrically speaking) between the solution and the ideal point. The best compromise or the closest solution to the ideal point is obtained by solving the linear programming (LP) model presented by Equation (1.4). Where, $F$ is the set of feasible solutions.

$$
\begin{equation*}
\min \quad L_{1}=\sum_{k=1}^{K} \alpha_{k} y_{k}, \quad \text { s.t. } \quad \mathbf{x} \in F \tag{1.4}
\end{equation*}
$$

Decision making process in operating rooms generally entails three hierarchical levels of strategic in long-term, tactical in medium-term, and operational in short-term [6]. The strategic level covers the long-term decisions such as the number and the mix of surgeries by which OR managers balance the waiting lists within the budget and time constraints. Strategic planning is "the process of reconciling supply and demand" 41]. OR planning is a complex task, because it is subjected to competing objectives, multiple stakeholders, and variations in demand. The major outcome of planning is to set a goal for the organization [42], a goal is a realistic and specific long-term aim for a specific time period. The long-terms decisions also involve the capacity dimensioning such as number of ORs, PACUs, ICUs, ward beds, nursing and support staff, etc. to match the demand with capacity.

Once the strategic decisions are made, their outputs are fed to the tactical level as inputs. At the tactical level, OR time is assigned to surgery specialties in order to produce a timetable called master surgery schedule (MSS). The MSS pairs each specialty with an OR-day in order to support the decisions made at the strategic level. According to the MSS, OR managers also produce a timetable for the number, types, and shifts of the support staff on each day.

Finally, at the operational level, surgical case sequencing occurs in order to assign patients a date, an OR, and a surgery start time. Scheduling is defined as "defining the sequence and time allocated to an operation" 41]. Scheduling allocates resources to operations in order to optimize one or multiple objectives [43]. Decisions made at the planning level directly impact OR scheduling such that OR scheduling must support the long-terms goals of resource utilization, patient mix, waiting list, etc. However, with the dynamics inherent to the system such as variations in arrivals (due to emergencies or no shows) and surgical case times, disturbances frequently occur in real time across
the peri-op process. Moreover, different stakeholders in different stages of the peri-op process have different preferences and priorities for OR management which is another source of trade-offs in the system.

In this dissertation, the goal is to design optimization models that reduce the negative impacts of trade-offs on the outcomes of the OR peri-op process. We seek for operation management strategies by proposing trade-off balancing models in the form Equation (1.4) to hedge against disturbances and inconsistencies in the system. We present novel algorithms and managerial guidelines to shift the outcomes of the system towards a more predictable state where the subjective values of the attributes are also at the nearest possible distance from their optima.

### 1.2 Motivation

Despite the fact that numerous operation management strategies and techniques exist in the industrial domain which are also applicable to the healthcare systems, the decision making process in the operating room context is heavily relied on the OR manager's experience [44] or even sometimes is on a first-come first-serve [45, 36] basis. Furthermore, the lack of systematic decision making processes leads to expediently use the scare and expensive resources to dampen the exigencies of disturbances once they occur. This may increase the expenditure and/or reduce the quality of care which negatively impact both the process and patients. Moreover, with revolving system dynamics during operations, the existing models fail to deliver the expected outcomes, thus, optimization models and managerial insights are required to shift the outcomes towards desirable targets while synchronizing the decisions made at different levels.

### 1.3 Summary of Contribution

Chapter 2 presents a novel stochastic model for OR planning with the objective of minimizing the trade-offs between throughput (number of patients) and total cost. Total cost is modeled is a function of OR utilization in terms of overtime and idletime. Newsvendor model and chance-constrained programming are used to model the studied problem under the presence of uncertainties in patient arrivals and surgery case times. The performance of the proposed models are examined using an OR historical dataset. We use statistical process control (SPC) techniques to evaluate the outcomes of different planning strategies. Our results show that minimizing trade-offs between inconsistent objectives of utilization and throughput provides the managers with a flexible tool to manage the waiting lists with lower costs. The proposed models address the strategic decision of reconciling demand and supply.

Chapter 3 presents a Priority-Type-Duration (PTD) algorithm which is simple but novel to assign surgeries to ORs under the constraints of surgery priority, surgery type, and OR costs. Patient assignment problem is normally modeled as a bin-packing problem with the objective of minimizing the number of required ORs, i.e., maximizing utilization. It is known that the longest processing time first ( $L P T$ ) rule performs best in generating the initial sequence for the bin-packing problem. However, $L P T$ rule fails to address concerns such as surgery priorities and surgery types. Using $L P T$ a short
case with a high priority is scheduled at the end of the sequence which is in conflict with the concept of prioritization. On the other hand, each surgery type requires different equipment and trained staff, yet using $L P T$, different surgery types may be scheduled in the same OR that increases the number of setups and turnovers. To address these issues, we propose the PTD algorithm that orders the surgeries according to their priority, then in each priority group, $P T D$ segregates the sequence according to the surgery types. Finally, in each segment, $P T D$ sequence cases with regard to their case times using $L P T$ rule. PTD significantly decreases the total cost of OR scheduling by reducing the number of required ORs and setups while addressing the priority of surgeries. PTD is capable of addressing tactical and operational decision making processes.

In Chapter 4, we propose a novel heuristic to balance the trade-offs between patient flow mean $(P t F)$ and patient flow variance (PtFV). Minimizing patient flow mean, i.e., $\min (P t F)$ reduces the average patient waiting time but at the cost high variation in the individual waiting times. Minimizing patient flow variance, i.e., $\min (P t F V)$ ensures the uniformity of waiting times among all individuals, but at the cost of higher average waiting time. We prove that the objectives of $\min (P t F)$ and $\min (P t F V)$ are inconsistent, thus, any sequence generates some levels of trade-offs in the system. We propose a fast heuristic to solve the combinatorial mixed integer problem with the objective of balancing trade-offs between $\min (P t F)$ and $\min (P t F V)$. The proposed models address the operational level of the decision making process.

Chapter 5 presents a mixed integer programming model to reduce the number of blockings between OR and the downstream resources of ICUs and PACUs. Once a surgery case is finished in the OR, it must be removed from the OR and transformed to either an ICU or PACU unit. If all of the downstream resources are occupied, the patient is held in the OR until a unit is available. This incident is called OR blocking that negatively affects the OR utilization and patient waiting times. Our proposed model recursively update itself with the number of available downstream units, then assigns surgery cases to ORs such that the number of blockings is minimized. Our proposed model outperforms existing models in the literature.

In Chapter 6, we mark the boundaries of risk management in an operating room peri-operative process. To this aim, we elaborate risk definition, risk sources, and propose risk management strategies. We identify uncertain demand, uncertain surgical case time, and the inconsistencies among objectives as the risk sources. We identify process utilization (Util) and patient length of stay (LoS) as two major KPIs driving value flow and patient flow, respectively. We show that surgical case sequencing significantly affect the distribution of the outcomes and has the potential to shift the performance of the system towards a less risky situation in which the outcomes are more predictable. We model the OR peri-op process as a 3 -stage flow shop scheduling problem. Through extensive case studies on the historical data of UKHC, we show the efficiency and effectiveness of trade-off balancing in mitigating the risk.

Chapter 7 extends our trade-off balancing models to address sustainability concerns in production scheduling. Despite the substantial research in sustainable manufacturing, a holistic model for sustainable production scheduling is virtually absent. To address this gap, this chapter presents a metric-based model to systematically and holistically evaluate the sustainability of the production schedules. We first identify
the fundamental metrics driving different areas of production, second, we asses those metrics with respect to the triple bottom lines (TBL) including economic, environmental, and social pillars. Third, we show the inconsistencies among the fundamental performance metrics and consequently among the objectives defined in the TBL. Finally, we propose a generic model for production scheduling for sustainability based on balancing the trade-offs among the inconsistent objectives.

Tabel 1.1 presents a summary of decisions, problems, models, and contributions of the chapters of this dissertaion.

In this dissertation, the notations are defined and valid within each chapter.

Table 1.1: Summary of Contributions


## Chapter 2

## Reconciling Supply and Demand: Stochastic Optimization Models for Efficient Operating Room Planning

In this chapter, we address the strategic problem of reconciling supply and demand by balancing the trade-offs between cost and throughput. Cost is formulated as a function of resource utilization. Throughput is defined as the number of patients served over the planning horizon. Our models seek efficient solutions to manage the waiting lists while the resources are also efficiently utilized.

Within a hospital, the operating room department has the largest cost and revenue. Because of the aging population, the demand for surgical services has been increasing sharply in recent years. On the other hand, the rate of OR capacity expansion is lower than the rate of increasing demand. As the result, OR managers must leverage their resources by efficient OR planning. OR planning is challenging due multiple competing/conflicting objectives such as cost minimization and throughput maximization. Inherent uncertainty in the surgical procedures and patients arrivals complicate the decision making process even more. This increases the risk of non-realization of the system objectives. In this chapter, stochastic bi-level optimization models were formulated to optimize total cost and throughput of ORs under the presence of uncertainties in patient arrivals and case times. Newsvendor model and chance-constrained optimization method were used to optimize multiple objectives under the presence of uncertainties. Using historical data, a simulation model was established to validate the results of the optimization models. Using statistical process control (SPC) techniques, the stability of each model was investigated. Using bi-level optimization, we addressed managerial preferences for total cost and throughput. Optimizing one objective may lead to a compromise of the optimality of the other one. Using a trade-off balancing model, we found solutions that minimize the sum of deviations from the best solutions for both total cost and throughput. Trade-off balancing optimization models may lead to better solutions, compared to the traditional multi-objective optimization models. The results of this chapter appears in [46].

### 2.1 Introduction

Healthcare is one of the most important sectors of the economy, historical data show a continuous increasing trend for the healthcare expenditures in recent decades. In 2009, the U.S. healthcare expenditures exceeded $17 \%$ of the gross domestic product (GDP), but was only $4.6 \%$ of the GDP in 1950 [47]. Experts in economy, healthcare policy, and public finance believe that the healthcare expenditures control is the main component of the deficit reduction challenge facing the U.S. government [2]. Hospitals, physicians, and drugs are the prime categories of expenditures, among which, hospital costs are the largest part of the healthcare expenditures [47].

As a result of increasing costs, hospitals need to leverage their resources by using them more efficiently. Within a hospital, the operating room department is one of the most critical resources, which has the largest cost and revenue [6, 48]. Because of the aging population, the demand for surgical services has been increasing sharply in recent years [49]. Therefore, efficient OR management has the potential of offering a significant cost saving. To efficiently utilize ORs, hospitals must provide high quality care more effectively with limited resources by developing efficient OR plans and schedules [6].

OR planning is challenging because it is under the continuous pressure of competing objectives, such as cost minimization, waiting time minimization, etc. There are numerous affecting factors and various active players in an OR department. Patients, surgeons, and OR managers are some of the OR active players who may have competing/conflicting objectives, with respect to cost, waiting times, etc. A large variety of performance measures are used to evaluate the OR planning, such as throughput, waiting lists, utilization, total cost, etc. The choice among these objectives is challenging and complex, because of multiple stakeholders (i.e., patients, surgeons, OR managers, etc.) with different incentives and priorities [6]. Therefore, any decision on one objective may generate trade-offs on the other objectives.

At the planning level, waiting time is defined as "the time between the referral date and the surgery date". Waiting time is of particular importance for patients [6]. In general, patients prefer to get on schedule as soon as possible. Long waiting times may negatively affect the patients' health conditions and consequently decrease the quality of care and patients' satisfaction. On the other hand, deteriorated health condition may results in an increase in the intensity and the cost of the required care, which is not desirable for the patients and/or for the healthcare providers, and insurance companies [50, 35, 51]. Throughput, which is defined as the number of patients treated in a period of time, is of particular importance for surgeons. Surgeons prefer to perform as many surgeries as possible in their assigned OR times. In general, because of educational and research workloads, surgeons are available on limited hours/days. Therefore, any idletime is not desirable for them [52, 53, 54].

The dependency between waiting time and throughput is clearly described by Equation(2.1), which is known as Little'sLaw [6, 55]. The average work in process ( $L$ ) in the system equals the average arrivals $(\lambda)$ to the system times the average cycle-time $(W)$.

$$
\begin{equation*}
L=\lambda W \tag{2.1}
\end{equation*}
$$

In the OR planning context, $L$ can be interpreted as the average number of patients on the waiting list, $\lambda$ as the average throughput and $W$ as the average of waiting time and case time. Therefore, by increasing throughput, the waiting time indirectly decreases, if the average number of patients in the waiting list keeps the same [6]. OR utilization is of importance for OR managers. OR utilization measures the proportion of potential output that is actually realized. OR utilization is a very important operational metric, because it provides insight to the existing slack in the system. An OR department with utilization less than $100 \%$, 'theoretically' has the potential to increase the production without generating overhead costs associated with capacity expansion. OR utilization is also a very effective metric to illuminate the cost structure of the OR department, by defining under-utilization (idletime) costs and over-utilization (overtime) costs. OR utilization is one the most extensively studied OR performance measures. According to the literature, the OR utilization should be maximized to avoid under-utilization (idletime) costs. But due to the high variations in case times and patients arrivals, highly utilized ORs are unstable [6, 56].

In this chapter, two performance criteria, throughput $(T P)$ and total cost $(T C)$ are taken into consideration. These performance measures are of importance to three main stakeholders (i.e. patients, surgeons, and OR managers) and each stakeholder must be 'adequately' satisfied.

There are two types of well-defined uncertainties in the OR planning literatures: (i) uncertainty in the case times, which is the difference between expected and actual surgery duration, and (ii) uncertainty in the patients arrivals caused by emergency arrivals and patients no-show cases [6, 57]. A large body of research has been done to tackle the uncertainty in case times [50, 51, 52, 58]. On the other hand, there are a few works addressing the uncertainty in the arrival rate [51, 53, 59]. There are fewer works, if any, considering uncertainties in both case times and patients' arrivals at the same time.

In this chapter, we propose a model which takes both sources of uncertainties into consideration. Without loss of generality, we assume that the case times and patients' arrivals are normally distributed. Using joint distribution of case times and patients' arrivals, we provide theoretical properties for OR department cost function (consisting of overtime and idletime). Using the Newsvendor model, we minimize the mismatch between expected and actual cost. Then, we utilize a bi-level chance constrained optimization model to optimize $T C$ and $T P$. To this aim, we alternate the order of objectives to show the trade-offs generated by the competing objectives. Finally, we propose a trade-off balancing model, to show the effectiveness and efficiency of trade-offs balancing over the traditional optimization models.

The main contributions of this chapter are (i) a stochastic model that explicitly takes uncertainties in both case times and patient arrivals into account, (ii) bi-level optimization models in which the order of objectives is alternated to show the trade-off among objectives, and (iii) a trade-off balancing model which balances the trade-offs between competing objectives of $\max (T P)$ and $\min (T C)$. The proposed approach is unique because it provides a flexible tool for OR managers to perform OR planning more efficiently by avoiding excessive overtime/idletime cost and long waiting lists.

### 2.2 Problem formulation

This section first presents a brief introduction to the Newsvendor model and the chanceconstrained optimization method. Next, two bi-level chance constrained models for OR planning problem under the presence of uncertainties in case times and patients' arrivals are proposed. A trade-off balancing model is also presented at the end of this section.

## Notations

| $C_{o}$ | Overage cost ratio |
| :--- | :--- |
| $C_{s}$ | Shortage cost ratio |
| $D$ | Demand, $D \sim N\left(\mu_{D}, \sigma_{D}^{2}\right)$ |
| $Q$ | Inventory level |

## Newsvendor model

Newsvendor model is a mathematical model used to determine optimal inventory levels subjected to fixed cost ratios (with $C_{o}$ for the overage cost and $C_{s}$ for the shortage cost, and $\left.C_{o}, C_{s}>0\right)$ where demand $(D)$ is normally distributed, i.e., $D \sim N\left(\mu_{D}, \sigma_{D}^{2}\right)$. Before the realization of $D$, the decision maker has to make the decision for inventory level $(Q)$. Minimizing the mismatch between $Q$ and $D$ is the objective of the Newsvendor model. If $Q>D$, the overage cost occurs, which is $C_{o}(\max (0, Q-D))$. If $Q<D$, the shortage cost occurs, which is $C_{s}(\max (0, D-Q))$. The optima $Q^{*}$ minimizes the $\mathbb{E}\left\{C_{o}(\max (0, Q-D))+C_{s}(\max (0, D-Q))\right\}$. Assuming unconstrained problem and taking convexity of objective function in $Q$ into consideration, the optimal solution can be derived by the first order condition 60, 61. Therefore, $F\left(Q^{*}\right)=\Phi(z)=\frac{C_{s}}{C_{o}+C_{s}}$, where $z=\frac{Q^{*}-\mu_{D}}{\sigma_{D}}$, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. $Q^{*}$ is explicitly presented by $Q^{*}=\mu_{D}+z \sigma_{D}$.

## Chance-constrained optimization

Chance-constrained optimization method is one of the approaches to solve optimization models in the presence of uncertainty. The basic idea is to ensure that the probability of meeting certain constraints is above a predetermined level [62, 63]. In other word chance-constrained model restricts the solution feasible region to achieve higher confidence level for the solution. The general optimization model under uncertainty can be formulated as follows:

$$
\begin{align*}
& \quad \min f(\mathbf{x}, \boldsymbol{\xi})  \tag{2.2}\\
& \text { s.t. } \\
& \qquad \quad g(\mathbf{x}, \boldsymbol{\xi})=0  \tag{2.3}\\
& h(\mathbf{x}, \boldsymbol{\xi}) \geq 0 \tag{2.4}
\end{align*}
$$

Equation (2.2) describes the objective function, Equation (2.3) imposes the equality constraints, and Equation (2.4) imposes the inequality constraints. $\mathbf{x}$ is the decision variables vector and $\xi$ is the uncertainties vector. Using chance-constrained method the inequality constraints can be formulated as $P\{h(\mathbf{x}, \boldsymbol{\xi}) \geq 0\} \leq \alpha$, where $\alpha \in[0,1]$ is the predetermined probability level.

## Bi-level optimization models

## P1: Total cost to throughput

We propose a bi-level optimization model that at the first level minimizes $T C$, and then at the second level maximizes $T P$ subjected to the cost constraints imposed by the first level optimization.

## First level: Total cost

Without loss of generality, we assume that the patients randomly arrive to the OR department via a normal distribution $d \sim N\left(\mu_{d}, \sigma_{d}^{2}\right)$. This assumption is very common in the literature and it fits the actual data when the number of arrivals is large enough.

Case times are assumed to be independent, identically distributed (i.i.d) random variables with normal distribution $p \sim N\left(\mu_{p}, \sigma_{p}^{2}\right)$. This assumption is also common in the literature and it fits the actual data when the patients' population is large enough. We assume that set-up and clean-up times are factored into the case time $p$. In this chapter, we deal with the strategic OR planning problem which is a long-term one. Therefore, the assumption of having a large population of patients holds. Assuming that the case times and patients' arrivals are i.i.d, we define a new random variable named workload ( $l$ ) as the product of patients' arrival $d$ and case times $p$ (i.e., $l=d \times p$ ). It is worth noting that the product of two normal distributions is not always a normal distribution. But, under some conditions the product can be approximated to a normal distribution. Particularly, for two normal distributions with different mean $\left(\mu_{x} \neq \mu_{y}\right)$ and different variance $\left(\sigma_{x}^{2} \neq \sigma_{y}^{2}\right)$, as the inverse variation coefficient $\lambda=\frac{\mu}{\sigma}$ increases $(\lambda>1)$, the distribution of the product of two independent normal variables tends to a normal distribution 64].

Using approximation formulas proposed by Macias and Oliviera (2012) [64], we can compute mean and variance of $l$ by Equation(2.5) and Equation(2.6), respectively.

$$
\begin{gather*}
\mu_{l} \approx \mu_{d} \mu_{p}  \tag{2.5}\\
\sigma_{l}^{2} \approx \mu_{d}^{2} \sigma_{p}^{2}+\mu_{p}^{2} \sigma_{d}^{2}+\sigma_{d}^{2} \sigma_{p}^{2} \tag{2.6}
\end{gather*}
$$

We are now able to utilize the Newsvendor model to obtain the optimum workload for the planning horizon $(T)$ minimizing the total cost $T C$. To translate cost factors $C_{s}$ and $C_{o}$ into the OR planning context, we argue as follow: shortage cost occurs in OR department when the planned capacity is less than the actual realized workload. Therefore, a fraction of actual workload must be done in overtime (overtime $=\max (0$, Actual workload - Planned capacity). With this argument, shortage cost of the Newsvendor model is an equivalent for the overtime in ORs. On the other hand, overage cost occurs in ORs when the planned capacity is greater than the actual realized workload, thus, a fraction of the planned capacity sits idle (idletime $=\max (0$, Planned capacity - Actual workload). Therefore, the overage cost of the Newsvendor model is an equivalent for the idletime in ORs. Based on what was discussed above, to drive out optimal planned capacity $\left(B^{*}\right)$ in the time period of $T$, we can define the expected cost by Equation 2.7 ). Let $g(\cdot)$ and $G(\cdot)$ be the density and cumulative distribution functions.

$$
\begin{equation*}
Y(B)=C_{o} \int_{0}^{B}(B-l) g(l) d l+C_{s} \int_{B}^{\infty}(l-B) g(l) d l \tag{2.7}
\end{equation*}
$$

Because Equation 2.7) is a convex function in $B$, by applying the first derivative condition, we can derive out the optimal planned capacity $B^{*}$. Applying the Leibnize rule 65] for
differentiation under the integral sign with respect to $B$ and setting it equal to zero, it yields:

$$
\begin{gather*}
\frac{d Y(B)}{d B}=C_{o} \int_{0}^{B} 1 g(l) d l+C_{s} \int_{B}^{\infty}(-1) g(l) d l=C_{o} G(B)-C_{s}[1-G(B)]=0 \rightarrow \\
G\left(B^{*}\right)=\frac{C_{s}}{C_{o}+C_{s}} \\
G\left(B^{*}\right)=\Phi(z)=\Phi\left(\frac{B^{*}-\mu_{l}}{\sigma_{l}}\right)=\frac{C_{s}}{C_{o}+C_{s}}  \tag{2.8}\\
B^{*}=\mu_{l}+z \sigma_{l} \tag{2.9}
\end{gather*}
$$

$G\left(B^{*}\right)$ represents the probability of workload being less than or equal to $B^{*}$ (i.e $P(l \leq$ $\left.\left.B^{*}\right)=G\left(B^{*}\right)\right)$. In other words, the probability of having enough capacity to meet $l$ is $\frac{C_{s}}{C_{o}+C_{s}}$. Another interesting implication of Equation 2.9 is that for the normal case, $B^{*}$ is an increasing function of $\mu_{l}$ and $\sigma_{l}$, provided that the $z$ is positive (because $C_{s}$ and $C_{o}$ are strictly positive). Considering this fact that the cost of overtime hours is always greater (or equal) than the cost of the idletime hours, we can conclude that $\frac{C_{s}}{C_{o}+C_{s}} \geq 0.5$. Therefore, we should allocate more capacity to avoid overtime (shortage cost). $B^{*}$ is then imposed as the capacity constraints onto the second level optimization model, which maximizes throughput, i.e., $\max (T P)$.

## Second level: Throughput

In order to maximize $T P$, the OR manager can estimate the expected case times based on historical data and surgeon estimation. The OR manager can use $n \times \bar{p} \leq B^{*}$ to derive out the number of patients to be planned. Where $\bar{p}$ is the OR manager's estimation for the case times and $n$ denotes the number of patients to be planned. The drawback of this simple procedure is that it ignores the variability in the case times and patients' arrivals and it does not provide any insights to the probability of expected overtime/idletime levels. To estimate the expected overtime, stochastic constraints must be imposed to the objective function to capture the uncertainties inherent to the surgical procedures. By letting $\alpha \in[0,1]$ be the probability of overtime exceeding a threshold (tolerance on overtime) denoted by $T L$, we can formulate the probabilistic constraints by Equation 2.10).

$$
\begin{equation*}
P\left\{\left(\sum_{i=1}^{n} p_{i}-B^{*}\right)>T L\right\} \leq \alpha \tag{2.10}
\end{equation*}
$$

While avoiding the overtime, OR managers also wish to minimize the idletime to treat more patients in a given time period. Therefore, we can formulate the nonlinear optimization model for throughput $(T P)$ by Equation 2.11 , where $\mathbb{E}(\cdot)$ is the expected value.

$$
\begin{array}{ll}
\min & \mathbb{E}\left(\max \left(0,\left(B^{*}-\sum_{i=1}^{n} p_{i}\right)\right)\right) \\
\text { s.t. } & \\
& P\left\{\left(\sum_{i=1}^{n} p_{i}-B^{*}\right)>T L\right\} \leq \alpha \tag{2.12}
\end{array}
$$

Equation (2.11) is the objective function which minimizes the expected idletime. This implies that the objective function indirectly maximizes $T P$ by packing more patients into $B^{*}$. Equation 2.12 is the chance constraints which guarantees that the overtime does not exceed a
predetermined threshold of $T L . T L$ and $\alpha$ are two managerial preferences by which the OR manager can balance the waiting list. If managers experience an increasing waiting list, by adjusting a larger value for $T L$, they are capable of managing the waiting list. However, it is important to consider resource availabilities (e.g., available budget, staff availability, etc.).

Case times are assumed to be independent, identically distributed (i.i.d) random variables with normal distribution $p \sim N\left(\mu_{p}, \sigma_{p}^{2}\right)$. Therefore, overtime which is defined by $O t=$ $\max \left(0, \sum_{i=1}^{n} p_{i}-B^{*}\right)$, is also a normally distributed random variable $O t \sim N\left(\mu_{O t}, \sigma_{O t}^{2}\right)$ , where $\mu_{O t}=n \mu_{p}$ and $\sigma_{O t}^{2}=n \sigma_{p}^{2}$. The chance-constrained optimization model can be approximated to its nonlinear deterministic counterpart as follows [63]:

$$
\begin{array}{lc}
\min & \mathbb{E}\left(\max \left(0,\left(B^{*}-\sum_{i=1}^{n} p_{i}\right)\right)\right) \\
\text { s.t. } & \\
& \mu_{O t}+\Phi^{-1}(1-\alpha) \sigma_{O t} \leq T L \\
& n \mu_{p} \leq B^{*} \tag{2.15}
\end{array}
$$

Equation (2.14) guarantees that overtime does not exceed the predetermined threshold ( $T L$ ) with probability of $\alpha$. Where $\Phi(\cdot)$ represents the cumulative distribution function of the standard normal variable. Equation (2.15) imposes the notion that the long-term performance must converge to the expected one. Figure 2.1 represents the relationship among number of planned patients, ratio of overtime threshold to planned capacity, and probability level $\alpha$. This relationship provides the OR managers with managerial guidelines for managing their OR department according to their preference over the acceptable overtime threshold and the associated risk of non-realization. By packing more patients into $B^{*}$ the risk of overtime increases (as we would intuitively expect). Given $T L$ and $\alpha$, the final output of this bi-level optimization is an ordered pair $\left(B_{1}^{*}, n_{1}^{*}\right)$, which specifies the optimum planned capacity and optimum number of patients to be planned.


Figure 2.1: Overtime threshold to planned capacity ratio vs. number of planned patients

## P2: Throughput to Total cost

## First level: Throughput

In order to show the trade-offs generated by the competing objectives, we alternate the order of objectives in our bi-level optimization models. If the OR managers' preference is to meet the demand by a predetermined confidence level $\beta$, they must first find the optimum number of patients to be planned and its associated risk of non-realization. Afterwards, they have to
find the required capacity, which minimizes the total cost generated by overtime and idletime. To this aim, we formulate the first level optimization model to maximize $T P$ as follows:

$$
\begin{array}{ll} 
& \min \\
\text { s.t. } & \mathbb{E}(\max (0, d-n)) \\
& P\{(d-n)>\gamma n\} \leq \beta \tag{2.17}
\end{array}
$$

Where, $d$ is the actual number of arrivals, $n$ is the number of patients to be planned, $\gamma$ is the acceptable threshold for number patients more than $n$, and $\beta$ is the confidence level. It is worth mentioning that $\gamma$ is proportional to $n$. Patients' arrivals are assumed to be independent, identically distributed (i.i.d) random variables with normal distribution $d \sim N\left(\mu_{d}, \sigma_{d}^{2}\right)$. The chance-constrained optimization model can be rewritten to its nonlinear deterministic counterpart as follows 63]:

$$
\begin{align*}
& \quad \min \quad \mathbb{E}(\max (0, d-n))  \tag{2.18}\\
& \text { s.t. } \\
& \qquad(1+\gamma) n \geq \mu_{d}+\Phi^{-1}(1-\beta) \sigma_{d} \tag{2.19}
\end{align*}
$$

Equation 2.18 is the objective function, which minimizes $(d-n)$, the difference between the actual number of arrivals and the planned number of patients. Equation 2.19 imposes that $d-n$ does not exceed a predetermined level with the probability of $\beta$. Figure 2.2 shows the relationship among $n, \gamma$ and $\beta$.


Figure 2.2: $\gamma n$ threshold vs. $n$

## Total cost

Let $n^{*}$ be the optimal number of patients to be planned obtained from the first level optimization. At the second level of the optimization model, the objective is to minimize total cost $(T C)$ generated by overtime and idletime under the constraints of treating $n^{*}$ patients over the planning horizon. We again use the Newsvendor model, to find to optimum capacity which minimizes $T C$. We define the workload $(l)$ which is a random variable as the product of $n^{*}$ and the case times (i.e., $l=n^{*} \times p$ ). Case times are assumed to be normally distributed random variables, therefore $l$ is also a normally distributed random variable. Using standard formula for the product of a real number and normally distributed random variables, we can compute mean and variance of $l$ by $\mu_{l}=n^{*} \times \mu_{p}$ and $\sigma_{l}^{2}=n^{*} \times \sigma_{p}^{2}$, respectively. We utilize the Newsvendor model to obtain the optimum planned capacity for the planning horizon $(T)$ that minimizes $T C$. The Newsvendor model yields
$G\left(B^{*}\right)=\Phi(z)=\Phi\left(\frac{B^{*}-\mu_{l}}{\sigma_{l}}\right)=\frac{C_{s}}{C_{o}+C_{s}} \rightarrow B^{*}=\mu_{l}+z \sigma_{l}$. Given $\gamma$ and $\beta$, the final output of this bi-level optimization is an ordered pair $\left(B_{2}^{*}, n_{2}^{*}\right)$ which specifies the optimum planned capacity and optimum number of patients to be planned for.

## P3: Trade-off balancing model

Intuitively, when we optimize the OR planning problem based on different orders of objectives, it is very likely that the OR plan performs poorly with regard to the second level objective. In other words, the result of the bi-level optimization model is the global optima with regard to the first level objective, whereas it is the local optima with regard to the second level objective. Therefore, alternating the order of objectives generates trade-offs in the system.

We utilize a simulation-based trade-off balancing model to minimize the trade-offs generated by alternating the order of objectives. Let $j \in\{1,2\}$ denote the orders of objectives, where, 1 represents $P 1$ and 2 represents $P 2$. The trade-offs balancing model can be formulated as follows:

$$
\begin{equation*}
\min \quad Y_{j}=\theta \frac{\text { Best }_{T P}-T P_{j}}{\mid \text { Best }_{T P}-\text { Worst }_{T P} \mid}+(1-\theta) \frac{T C_{j}-\text { Best }_{T P}}{\mid \text { Best }_{T C}-\text { Worst }_{T C} \mid} \tag{2.20}
\end{equation*}
$$

where $\theta \in[0,1]$ is the managerial weight assigned to $\max (T P)$ and intuitively $(1-\theta)$ to $\min (T C)$, setting these weights is a subjective decision, hence, difficult to argue. Equation 2.20 ) expresses the sum of normalized deviations to tackle the fact that each performance measure is measured by a different unit and with different granularity.

### 2.3 Case study

To analyze the efficiency of our proposed models, we establish a simulation model using historical data provided by UKHealthcare (University of Kentucky healthcare). UKHealthcare hospitals perform a wide variety of surgery procedures and on average treat more than 30,000 surgery cases per year.

Different scenarios for different combination of managerial preferences for the order of objectives, $\alpha, \beta, T L$, and $\gamma$ are considered. Different performance measures including throughput, total $\operatorname{cost}$ ( $T C=C_{o} \times$ idletime $+C_{s} \times$ overtime), idletime, and overtime are computed to compare the performance of the proposed models.

## Data generation

Normality tests for patients' arrivals and case times distributions are performed on the UKHealthcare historical data of a certain surgery group. Anderson-Darling normality test shows a $p$ value of 0.3355 and 0.166 for patient arrivals and case times respectively ( $p \geq 0.05$ for both), and the data plots form a fairly straight line along the fitted line. Therefore, it appears that the normal distribution is a good fit to the data set. Detailed results of the normality tests can be found in Table 2.1 and Figure 2.3. To generate random data, we find

Table 2.1: Statistics obtained from UKHealthcare data

|  | $\mu$ | $\sigma$ |
| :---: | :---: | :---: |
| Case times | 156 | 60 |
| Patient arrivals | 72 | 33 |



Figure 2.3: Anderson-Darling normality tests
the maximum and minimum value of case time (and patients' arrivals as well), then using $R=\max -\min$, we calculate the range of data. $R$ then was discretized to 5 equal increments and the probability of each increment was calculated. Assuming $T$ (planning horizon) as equal to one week, using Monte Carlo simulation, patients' arrivals and associated case times were generated for 50 weeks. 50 replications for each week were done and the data stored to run the simulation model.

## Cost ratios

In this case study, without loss of generality, we assume that $C_{s}=2 C_{o}$. This implies that the cost of overtime is twice as the cost of idletime, although in practice these cost ratios may vary place to place, as long as $C_{s} \geq C_{o}$, it does not affect the form of the cost function. Therefore, in our case; $\Phi(z)=\frac{C_{s}}{C_{o}+C_{s}}=0.6667 \rightarrow z=0.4307$.

## Managerial preferences

## P1: Total cost to throughput

We consider $\alpha \in\{0.01,0.05,0.10,0.20,0.30\}$, and $T L \in\left\{0.1 B^{*}, 0.2 B^{*}, 0.3 B^{*}\right\}$, therefore, we have 15 different combinations of confidence levels and acceptable overtime thresholds. Using Equation(2.3), Equation(2.5), Equation(2.9), and the historical data ( $d \sim N(72,33$ ) and $p \sim N(156,60)$, we obtain $B^{*}=13194.06$ minutes. Having $B^{*}$, by using Equation 2.13) through Equation(2.15), we obtain $n^{*}$. Table 2.2 shows $n *$ for different combinations of $\alpha$ and $T L$. For the sake of brevity, a code is assigned to each combination of $\alpha$ and $T L$ as shown in Table 2.2 (e.g., $\mathrm{C} 2 \mathrm{P}-1$ for $T L=0.1$ and $\alpha=0.01$ ). It is worth mentioning that $n^{*}$ is the same for those combinations with $\alpha>0.01$ or $T L>0.1 B^{*}$, this is because Equation 2.15) sets an upper bound on the number of patients to ensure that the long term performance converges to the expected one.

## P2: Throughput to Total cost

We consider $\beta \in\{0.01,0.05,0.10,0.20\}$ and $\gamma \in\left\{0.10 n^{*}, 0.20 n^{*}, 0.30 n^{*}\right\}$, therefore, we have 15 combinations of different $\beta$ and $\gamma$. Table 2.3 shows ( $B^{*}, n^{*}$ ) of these combinations. A code is assigned to each combination of $\alpha$ and $\gamma$ as shown in Table 2.3 (e.g., P2C-1 for $\gamma=0.1$ and $\beta=0.01$ ).

Table 2.2: $\left(B^{*}, n^{*}\right)$ of different combinations of different $\alpha$ and $T L$

| $\alpha$ | TL |  |  |
| :---: | :---: | :---: | :---: |
|  | 0.1 | 0.2 | 0.3 |
| 0.01 | C2P-1 | C2P-6 | C2P-11 |
|  | (13194,84) | $(13194,87)$ | $(13194,87)$ |
| 0.05 | C2P-2 | C2P-7 | C2P-12 |
|  | $(13194,87)$ | $(13194,87)$ | $(13194,87)$ |
| 0.10 | C2P-3 | C2P-8 | C2P-13 |
|  | $(13194,87)$ | $(13194,87)$ | $(13194,87)$ |
| 0.20 | C2P-4 | C2P-9 | C2P-14 |
|  | $(13194,87)$ | $(13194,87)$ | $(13194,87)$ |
| 0.30 | C2P-5 | C2P-10 | C2P-15 |
|  | $(13194,87)$ | $(13194,87)$ | $(13194,87)$ |

Table 2.3: $\left(B^{*}, n^{*}\right)$ of different combinations of different $\beta$ and $\gamma$

| $\beta$ | $\gamma$ |  |  |
| :---: | :---: | :---: | :---: |
|  | 0.1 | 0.2 | 0.3 |
| 0.01 | P2C-1 | P2C-6 | P2C-11 |
|  | $(22008,144)$ | $(20948,137)$ | $(20190,132)$ |
| 0.05 | P2C-2 | P2C-7 | P2C-12 |
|  | $(19130,125)$ | $(18372,120)$ | $(17766,116)$ |
| 0.10 | P2C-3 | P2C-8 | P2C-13 |
|  | $(17766,116)$ | $(17160,112)$ | $(16553,108)$ |
| 0.20 | P2C-4 | P2C-9 | P2C-14 |
|  | $(16098,105)$ | $(15643,102)$ | $(15188,99)$ |
| 0.30 | P2C-5 | P2C-10 | P2C-15 |
|  | $(15491,101)$ | $(15036,98)$ | $(14733,96)$ |

## Performance measures

Several performance measures including throughput, total cost, idletime and overtime were computed to compare the performance of the proposed models.

### 2.4 Results

Statistical process control (SPC) techniques are used to monitor the performance of the proposed models. $\bar{x}$ - $R$ charts and process capability analyses are used to compare the quality of the proposed models. Process capability index $c_{p}$ is an indicator, representing if the outcomes of a process are within the user-defined specification limits, where $c_{p}=\frac{U S L-L S L}{6 \sigma}$, and $U S L$ and $L S L$ are user-defined upper specification limit and lower specification limit, respectively. The larger the $c_{p}$, the less variations in the process. Process capability index $c_{p k}$ represents the congestion of outcomes around the center line, the larger the $c_{p k}$, the more congestion of outcomes around the center line 66].
$C 2 P-1$ has the best performance on idletime, overtime, throughput and total cost among all combinations of $P 1$. As mentioned earlier, Equation 2.15 sets an upper bound on the
number of patients to ensure that the long-term performance converges to the expected one, therefore, all combinations of $C 2 P-2$ to $C 2 P-15$ have the same performance, and are not sensitive to managerial preferences. Consequently, hereafter we only review the behavior of C2P-1.

## Idletime

Idletime represents the unproductive time of ORs. In our case study, idletime is the difference between the actual workload and the planned capacity, i.e., idletime $=\max \left(0, B^{*}-\sum_{i=1}^{n} p_{i}\right)$ where $i$ is the index of cases. OR managers want to minimize the amount of idletime. For the sake of brevity, we present the best scenario of each model. Average idletime for $\mathrm{C} 2 \mathrm{P}-1$ and $P 2 C-15$ equals 3907 and 6163 minutes, respectively. $C 2 P-1$ outperforms the best $P 2 C$ by $\frac{6163-3907}{6163}=36.6 \%$.

Process capability analyses results for idletime show that the $c_{p}=0.30$ for $C 2 P-1$ and $c_{p}=0.32$ for $P 2 C-15$. This implies that the $P 2 C-15$ is more stable, but on average it generates a larger amount of idletime compared to $C 2 P-1$. For $C 2 P-1$ a big proportion of experiments (almost $1 / 3$ ) does not generate any idletime. $c_{p k}=0.27$ for $C 2 P-1$ and $c_{p k}=0.08$ for $P 2 C$ 15 , which implies that the outcomes of $C 2 P-1$ are more centered around the center line of specification limits. Figure 2.4 shows the SPC results for idletime.


Figure 2.4: SPC results for Idletime

## Overtime

Overtime represents the amount of time spent after regular hours to treat all patients. In our case study, overtime is the difference between the actual workload and the planned capacity, i.e., Overtime $=\max \left(0, \sum_{i=1}^{n} p_{i}-B^{*}\right)$. OR managers want to minimize the amount of overtime. $P 2 C-15$ has the worst performance on generating overtime among all combinations of $\beta$ and $\gamma$, i.e., generating the largest amount of overtime. Therefore, we compare the performance of $C 2 P-1$ with the performance of $P 2 C-15$. Average overtime is 1030 minutes for $C 2 P-1$ and 582 minutes for $P 2 C-15$. Therefore, $P 2 C-15$ outperforms $C 2 P-1$ by $\frac{1030-58}{1030}=$ $43.50 \%$.

Process capability analyses results for overtime show that the $c_{p}=0.09$ for $C 2 P-1$ and $c_{p}=0.14$ for $P 2 C-15$. This implies that $P 2 C-15$ generates less variations in overtime. Figure 2.5 shows the SPC results for overtime.


Figure 2.5: SPC results for Overtime

## Total cost

Total cost is defined by $T C=C_{o} \times$ idletime $+C_{s} \times$ overtime. This formula provides insight to the trade-off between overtime and idletime costs. The Newsvendor model minimizes this trade-off. $P 2 C-15$ has the best performance on total cost among all combinations of $\beta$ and $\gamma$. Therefore, we compare the performance of $C 2 P-1$ with the performance of $P 2 C-15 . C 2 P-1$ outperforms $P 2 C-15$ by $\frac{6163.36-5967.16}{6163.36}=3.18 \%$. Since this percentage is not a large number, an ANOVA was performed to determine whether the difference between $C 2 P-1$ and $P 2 C-15$ is significant. The ANOVA results demonstrates that the difference between total cost of $C 2 P-1$ and $P 2 C-15$ is statistically significant at $95 \%$ confidence level ( $p<0.05$ ). Process capability analysis shows that $C 2 P-1$ has lower mean and less variations compared to $P 2 C-15$. Detailed results of SPC and ANOVA for total cost can be found in Table 2.4. Figure 2.6 shows the SPC results for total cost.

Table 2.4: ANOVA results for total cost between $C 2 P-1$ and $P 2 C-15$

| Source | DF | SS | MS | $p$ |
| :---: | :---: | :---: | :---: | :---: |
| Factor | 1 | 48119225 | 48119225 | 0.041 |
| Error | 4998 | 63045301205 | 12614106 |  |
| Total | 4999 | 63093420430 |  |  |



Figure 2.6: SPC results for total cost

## Throughput

Throughput is defined as the number of treated patients in a given period of time. In this case study, we assume that all arriving patients must be treated. This assumption implies two points: first, the number of treated patients is the same for all scenarios, second, scenarios throughput differ from each other only by the number of patients treated in the overtime. Therefore, we use the number of patients treated in overtime as a performance indicator, to compare the throughput of different scenarios.

As we showed in previous section, $P 2 C-15$ has the lowest total cost among all combinations of $P 2 C$ (all combinations of $\beta$ and $\gamma$ ). It also has the worst performance on the number of patients served in overtime among all combinations of $\beta$ and $\gamma$. Therefore, we compare the performance of $C 2 P-1$ with the performance of $P 2 C-15$. $P 2 C-15$ outperforms $C 2 P-1$ by $\frac{6.75-3.318}{6.75}=50.84 \%$. Process capability analysis shows that $c_{p}=0.29$ for $C 2 P-1$ and $c_{p}=0.47$ for $P 2 C-15$, which implies that $P 2 C-15$ has less variations on the number of served patients in overtime. Figure 2.7 shows the SPC results for throughput (number of patients served in overtime).


Figure 2.7: SPC results for the number of patients served in overtime

## Trade-off balancing

As we showed in earlier sections, $\min (T C)$ and $\max (T P)$ are inconsistent objectives, which imply that optimizing one may lead to compromise on the optimality of the other one. Therefore, a model that has the least deviations from both objectives optima may be of interest to OR managers. We define $\theta \in\{0,0.25,0.50,0.75,1\}$, as the set of OR managers' preference for $\max (T P)$ over $\min (T C)$. Using Equation 2.20 ) and simulation results for best/worst performance, we find the most efficient model that minimizes sum of deviations from the objectives optima. We have 80 different scenarios and different $\theta$ as shown by Table 2.5.

As we intuitively expected, $C 2 P-1$ has the least deviations from the best value of $\min (T C)$ (i.e., $\theta=0$ ). Once OR managers have equal preferences over both objectives (i.e., $\theta=0.50$ ), $P 2 C-3$ has the best performance. Several combinations of $P 2 C$ have the best performance on $\max (T P)$ (i.e., $\theta=1$ ), as we intuitively expected. Figure 2.8 shows the sum of average deviations from the best values of $\min (T C)$ and $\max (T P)$.

Table 2.5: Sum of average deviations from the best $T C$ and $T P$
$\left.\begin{array}{c|ccccc}\hline & \theta=0 & \theta=0.25 & \theta=0.50 & \theta=0.75 & \theta=1.00 \\ \hline \text { C2P-1 } & 0.2882 & 0.2552 & 0.2222 & 0.1892 & \text { (Worst) }\end{array} \begin{array}{c}\text { (Worst) } \\ \\ \text { (Best) }\end{array}\right)$


Figure 2.8: Sum of average deviations from the best $T C$ and $T P$

### 2.5 Conclusion

In this chapter, we present bi-level optimization models for OR planning which captures uncertainties in patients' arrivals and case times. Minimizing total cost $\min (T C)$ and maxi-
mizing throughput max $(T P)$ are selected as objectives of the optimization models. Different managerial preferences for objectives are taken into account. By alternating the orders of objectives in our bi-level optimization models, we show that $\min (T C)$ and $\max (T P)$ are inconsistent. We propose a simulation-based trade-off balancing model to minimize the sum of deviations from the best value of each objective. Using historical data obtained from UKHealthcare, a large set of computational experiments are carried out. The simulation results show that for $\min (T C)$ our proposed $C 2 P$ has the best performance, whereas $P 2 C$ has the best performance on $\max (T P)$. The proposed models provide a flexible tool for OR managers to perform OR planning more efficiently based on their preferences and settings. The results of the study are applicable to manufacturing systems with multiple objectives under the presence of variations in demand and processing times.

## Chapter 3

## Tactical Operating Room Planning Under Surgery Type and Priority Constraints

Operating room tactical planning is critical in healthcare systems to reduce cost and improve the efficiency of ORs. The OR planning problem is complicated, involving many conflicting factors, such as overtime and idletime that both affect OR utilization and consequently affect the cost to the hospital. Allocating different types of surgeries into OR blocks affects the setup cost, whereas priorities of surgeries affect the sequence of surgeries in the OR block scheduling. Surgery durations affect both OR utilization and OR block scheduling. Traditionally, one important method for OR block scheduling is the bin-packing model, and the $L P T$ rule is the most commonly used method to generate the initial sequence for the bin-packing problem. However, in addition to case times, it is necessary in OR scheduling to consider two additional properties of the surgeries including (i) surgery priority, and (ii) surgery types. (i) means surgery with high priories should be scheduled earlier than those with low priority, and (ii) means that the same type of surgeries should be scheduled into ORs equipped accordingly. $L P T$ rule fails to address the priority of surgeries where a short case with a high priority may be scheduled at the end of the sequence. $L P T$ also does not distinguish between surgery types, thus, different surgery types may be assigned to the same OR that consequently increases the number of required setups and turnovers.

In this chapter, we propose a simple multi-step approach called priority-type-duration $(P T D)$ rule to generate the initial sequence for bin packing. The results of our case studies show that PTD rule outperforms the $L P T$ rule in terms of total cost, and priority concerns. The results of this chapter appears in [67].

### 3.1 Introduction

Tactical OR planning is a medium-term decision making process which its major output is the OR block schedules. OR block schedules specify the allocation of OR block times to surgery specialties, and the assignment of surgery specialties to daily OR slots. The tactical OR planning specifies the number of required ORs on each day, and the associated human resources (i.e., anesthesiologists, surgeons, nurses, staff, etc.), and finally, the equipment needed in each OR. Overtime, idle time, setups, and cost are of concern in OR block scheduling, thus, it is under the budget constraints.

A wide range of research methodologies were used to address, evaluate, and optimize the performance of OR block schedules [6]. Marques and Captivo (2012) 68] used integer programming to assign elective surgeries to an operating room, a day, and a specific period of
time on a weekly planning horizon in order to maximize the use of ORs. They used real data to test their approach and to compare their results with actual OR performance. Lamiri et al. (2008) [50] established a stochastic model for operating room planning for both elective and emergency cases, in order to minimize the overtime costs of ORs. They used Monte Carlo simulation and mixed integer programming to solve the proposed model. The model reduced the cost of ORs over the long-term horizon, and fulfilled the demand of emergent cases. Hsu et al. (2003) [69] proposed a tabu search approach to sequence elective cases in order to minimize the number of required nurses in PACU. Testi et al. (2007) [11] developed a hierarchical three-phase approach for scheduling of operating rooms in order to improve the overall operating theatre efficiency. At the first phase, a bin-packing problem was solved in order to select the number of surgeries to be scheduled on a weekly basis. At the second phase, a blocked booking method was used to determine the optimal time tables which defined the assignment of wards and ORs. At the third phase, the longest processing time (LPT) rule and the shortest processing time $(S P T)$ rule were used to sequence cases. Fei et al. (2010) [70] designed a weekly surgery scheduling method for an operating theatre in order to minimize the overtime cost in the operating theatre, maximize the utilization of ORs, and to minimize the unexpected idle time between surgical cases. This problem was solved in two phases. First, the planning problem is solved to assign a surgery date to each patient with regard to the availability of operating rooms and surgeons. Second, a daily scheduling problem is devised to determine the sequence of surgeries in each operating room on each day, taking into account the availability of the recovery beds.

A surgery priority indicates the severity of patient's health condition such that a surgery with high priority needs immediate care in order to prevent deteriorated health condition. I our study, we adopt the priority groups proposed by Valente et al. (2009) [71] as shown by Table 3.1.

Table 3.1: Urgency Related Groups

| $U R G^{*}$ | Clinical assessment | MTBT** <br> (Days) |
| :--- | :--- | :--- |
| A1 | Evident fast progression of disease affecting outcome by delay | 8 |
| A2 | Potential fast progression of disease affection outcome by delay | 30 |
| B | Severe pain and/or dysfunction, and/or disability, but no fast pro- <br> gression of disease affecting outcome by delay | 60 |
| C | Mild pain, and/or dysfunction, and/or disability, but no fast pro- <br> gression of disease affecting outcome by delay | 180 |
| D | No pain, dysfunction, and disability, no fast progression of disease <br> affecting outcome by delay | 360 |
| * Urgency Related Groups(URG) ** Maximum Time Before Treatment (MTBT) |  |  |

Many studies has been reported on assigning the surgery cases to ORs in order to optimize the OR efficiency from different perspectives, but there are few works considering the surgery priority, surgery type, and their impacts on the overall performance of ORs. In most cases, the surgery type is omitted at planning phase, and several surgery types are scheduled together in a single OR. This incident not only increases the number of setups for each OR, but also the increases the idle time due to the high number of required turnovers. In most studies at the planning phase, the surgeries are sequenced by the $L P T$ rule to generate the initial sequence of the bin-packing problem.

This chapter presents a novel yet simple rule to assign surgeries to ORs under the constraints of surgery priority, surgery type, and OR costs. Patient assignment problem is normally modeled as a bin-packing problem with the objective of minimizing the number of required ORs, i.e., maximizing ORs utilization. It is known that the longest processing time first $(L P T)$ rule performs best in generating the initial sequence for the bin-packing problem. However, $L P T$ rule fails to address issues such as surgery priorities and surgery types. Using $L P T$ a short case with a high priority is scheduled at the end of the sequence which is in conflict with the concept of prioritization. On the other hand, each surgery type requires different equipment and trained staff, yet using $L P T$, different surgery types may be scheduled in the same OR that increases the number of setups and turnovers. To address these issues, we propose a simple algorithm called $P T D$ (Priority-Type-Duration) that orders the surgeries according to their priority, then in each priority group, $P T D$ segregates the sequence according to the surgery types. Finally, in each segment, $P T D$ sequence cases with regard to their case times using $L P T$ rule. $P T D$ significantly decreases the total cost of OR scheduling by reducing the number of required ORs and setups while addressing the priority of surgeries at the same time. PTD is capable of addressing tactical and operational decision making processes.

### 3.2 Problem Formulation

The interest at the tactical planning level is to assign an OR to each patient while satisfying their priority under the budget and time constraints. For a fixed number of patients $n$, the objective is to minimize the total cost of the ORs. Total cost is a function of overtime, idletime, and setup costs as defined by Equation(3.1) where $j \in\{1,2, \ldots, J\}$ is the index of operating rooms. $R_{j}$ is the amount of workload performed in regular hours in OR $j$, and $C_{R}$ is the cost of regular time per minute. $O_{j}$ is the amount of workload performed in overtime in $\mathrm{OR} j$, and $C_{O}$ is the cost of overtime per minute. $I_{j}$ is the amount of idletime in $\mathrm{OR} j$, and $C_{I}$ is the cost of idletime per minute. Finally, $S_{j}$ is the number of setups in OR $j$, and $C_{S}$ is the cost of each setup.

$$
\begin{equation*}
\sum_{j=1}^{J} R_{j} C_{R}+O_{j} C_{O}+I_{j} C_{I}+S_{j} C_{S} \tag{3.1}
\end{equation*}
$$

## Notations

| $j$ | Index of operating rooms, $j=1,2, \ldots, J$ |
| :--- | :--- |
| $R_{j}$ | The amount of regular working time of OR $j(\mathrm{~min})$ |
| $O_{j}$ | The amount of overtime working time of OR $j(\mathrm{~min})$ |
| $I_{j}$ | The amount of idle time of OR $j(\mathrm{~min})$ |
| $S_{j}$ | The number of setups in OR $j$ |
| $C_{R}$ | Cost per unit of regular working time $(\$ / \mathrm{min})$ |
| $C_{O}$ | Cost per unit of overtime working time $(\$ / \mathrm{min})$ |
| $C_{I}$ | Cost unit of idle working time $(\$ / \mathrm{min})$ |


| $C_{S}$ | Cost of each set-up (\$) |
| :--- | :--- |
| $i$ | Index surgery cases, $i=1,2, \ldots, n$ |
| $p_{i}$ | Priority of case $i, p_{i} \in\{1,2, \ldots, 5\}$ |
| $t_{i}$ | Surgery type of case $i, t_{i} \in\{1,2, \ldots, T\}$ |
| $d_{i}$ | Surgery time of case $i$ |
| $H$ | Regular working time of ORs (min) |

The average cost of OR regular working time varies over a wide range from $\$ 22$ to $\$ 133$ per minutes. The actual cost depends on many factors including the region, surgery type, whether the OR cost includes the fixed overhead costs that are constant regardless of the number of surgeries performed, or if it only accounts for the variable costs, which vary according to the number of cases performed, or whether professional fees of the physician work in the OR are included [72]. We assume the cost per unit of idletime is equal to the cost per unit of regular time $C_{R}=C_{I}$, this is because that the staff monthly payment is fixed regardless of whether they are working or waiting for the beginning of a surgery. According to the Fair Labor Standards Act (FLSA) [73], overtime must be paid at a rate no less than 1.5 times regular rates after 40 hours of work in a week, ,thus, we arbitrarily assume $C_{O}=1.5 \times C_{R}$, although this ratio varies place to place, it does not affect the form of the objective function. Setup cost generally depends on the complexity of surgeries, equipment and resources used for the surgeries. Thus, this cost varies over a wide range. In summary Equation(3.1) can be expressed by Equation (3.2).

$$
\begin{equation*}
\sum_{j=1}^{J}\left(R_{j}+1.5 O_{j}+I_{j}\right) C_{R}+S_{j} C_{S} \tag{3.2}
\end{equation*}
$$

Let $p_{i} \in\{1,2, \ldots, 5\}$ denote the priority of case $i$ such that the greater the $p_{i}$ the higher the priority associated to case $i, t_{i} \in\{1,2, \ldots, T\}$ denotes the surgery type of case $i$, and finally, let $d_{i}$ denote the surgery duration of case $i$. A bin-packing model maximizes utilization and minimizes the idle time which consequently affects the cost of the planning horizon. The following model represents the general mathematical model of bin-packing for ORs.

$$
\begin{align*}
& \min \quad z=  \tag{3.3}\\
& \text { s.t. } \sum_{j=1}^{J} y_{j}  \tag{3.4}\\
& \sum_{j=1}^{{ }_{j}^{n}} d_{i} x_{i j} \leq H  \tag{3.5}\\
& \sum_{i=1}^{n} x_{i, j}=1  \tag{3.6}\\
& y_{j}= \begin{cases}1, & \text { if OR } j \text { is used } \\
0, & \text { otherwise }\end{cases} \\
& x_{i, j}= \begin{cases}1, & \text { if case } i \text { is assigned to OR } j \\
0, & \text { otherwise }\end{cases}
\end{align*}
$$

Objective function(3.3) minimizes the number of required ORs. Constraints (3.4) imposes $H$ as the total available time (capacity) of each OR. $x_{i, j}$ is an integer decision variable that equal 1 if the surgery $i$ is assigned to the OR $j$, otherwise 0 . Constraints (3.5) and (3.6) guarantee that each surgery is assigned to an OR only once. $y_{j}$ is an integer decision variable that equal 1 if the $\mathrm{OR} j$ is used over the planning horizon, otherwise 0 . The bin-packing problem is proved to be $N P$-complete [74]. Therefore, exact solutions are not computationally efficient. Multiple heuristics have been proposed to find a near-optimal solution to the binpacking problem among which first fit decreasing (FFD) and best fit decreasing (BFD) are known to perform best [75]. The initial sequence for FFD and BFD is generated by ordering the items by $L P T$ rule. However, as it was mentioned earlier in the OR context, $L P T$ rule fails to address surgery priorities and types.

We propose a simple multi-step procedure to sequence surgery cases in order to generate an initial sequence for the bin-packing model. Since the priority is the most important factor for performing a surgery, we first sequence surgeries according to their relative priorities. Thus, we have five groups of surgeries with regard to their priority. The second step is to group surgeries according to surgery types within each priority group. The third step is to sequence surgeries in each subgroup by the $L P T$ rule with regard to the surgery duration. After obtaining the initial sequence, we assign surgeries to ORs from the head of the sequence (the highest priority) to the tail of the sequence (the lowest priority), while we avoid combining different surgery types into the same OR. If there is still some slack time in an OR after assigning all cases of high priority group, we search for a compatible case from the lower priority groups with the same surgery type. If there is no compatible cases, we leave the remaining time idle. For the last case in the initial sequence, if there is no slack in any OR, instead of opening a new OR, we assign the last case to an OR even though the case is performed in overtime. This step is to reduce the idletime cost generated due to opening a new OR for only one case. The proposed procedure is named as PTD (Priority-Type-Duration) and summarized by Algorithm 1 .

```
Algorithm 1 PTD algorithm for assigning surgery cases to ORs
    Step 1. Group surgeries according to their priority.
    Step 2. Group surgeries in priority groups by their types.
    Step 3. Within each subgroup sequence surgeries according to their duration
        by \(L P T\) rule. Now we have the initial sequence for bin-packing.
    Step 4. Assign surgeries to ORs according to initial sequence.
    Step 5. If an OR has some slack time after assigning all high priority cases,
        search lower priorities for compatible cases with the same surgery type.
        If there is no compatible case, leave the rest of OR time idle.
        This step is to reduce the idle time and number of setup as well.
    Step 6. For the last case of each surgery type overtime is allowable.
```


### 3.3 Case studies

In order to examine the performance of the proposed PTD algirthm, we perform a series of case study on the OR historical dataset. We compare the performance of $P T D$ with that of $L P T$. According to the historical data from a local hospital, there are 24 surgery types. Thus the surgery types in this case study are randomly generated from a uniform distribution of $[1,24]$ as an integer value, i.e., $t_{i} \in[1,24], t_{i} \in \mathbb{Z}$. To achieve efficient OR
planning, the surgery duration must be estimated accurately, many researchers used historical data to estimate surgery duration, and some recommended log-normal distribution to estimate surgery duration [76]. In this study surgery durations are randomly generated from a uniform distribution of $[60,180]$ based on the historical data from the hospital, i.e., $d_{i} \in[60,180]$. The unit of time is minute for all time values. OR block time is set at 600 minutes ( 10 hours) for all ORs. Each surgery can be assigned to any OR, but a setup cost occurs if different types of surgeries are assigned to the same OR. In this case study $n$ elective cases with different priorities and different surgery types are selected from the waiting list.

We use several performance measures such as OR utilization, overtime, idle time, number of ORs, and number of setups to evaluate the performance of OR tactical planning. We sequence surgeries in order to minimize the total cost of the planning level which consists of maximizing the ORs utilization, reducing overtime and idletime, reducing the number of required ORs, and reducing the number of setups. We compare our proposed method with the well-known method $L P T$. The aforementioned performance measure were calculated for both methods. Smoothness Index $(S I)$ as defined by Equation(3.7) is used to compare the evenness of surgery load distribution between ORs.

$$
\begin{equation*}
S I=\sqrt{\sum_{j=1}^{J}\left(H-\sum_{i=1}^{n} x_{i, j} d_{j}\right)^{2}} \tag{3.7}
\end{equation*}
$$

Utilization is the ratio of sum of case times and the total available time of all ORs, and can be expressed by Equation $(3.8)$ where $N$ is the number of required ORs.

$$
\begin{equation*}
U t i l=\frac{\sum_{i=1}^{n} d_{i}}{N \times H} \tag{3.8}
\end{equation*}
$$

### 3.4 Results and Discussion

500 elective surgeries are randomly generated for a planning horizon of one week ( 5 days). The $P T D$ and $L P T$ rules were coded using MATLAB R2015b and the performance measures for each method were calculated, to prepare more comprehensive data this scenario was replicated 200 times that is equal to four years. Table 3.2 shows the computational results.

Table 3.2: Computation results for PTD and LPT

|  |  | $\boldsymbol{R}$ | $\boldsymbol{O}$ | $\boldsymbol{I}$ | $\boldsymbol{N}$ | $\boldsymbol{S}$ | $\boldsymbol{S I}$ | $\boldsymbol{U}$ Util |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| PTD | Min | 58197 | 0 | 3946 | 104 | 104 | 824.01 | 87.31 |
|  | Max | 61847 | 966 | 8675 | 116 | 116 | 1490.99 | 93.67 |
|  | Ave | 60024 | 286.38 | 6801 | 111.37 | 111.37 | 1195.47 | 89.83 |
|  | STD | 794 | 160.77 | 811.95 | 1.93 | 1.93 | 125.58 | 1.10 |
| LPT | Min | 58338 | 0 | 14963 | 123 | 472 | 1431.97 | 76.65 |
|  | Max | 62105 | 0 | 18346 | 133 | 494 | 1687.70 | 79.96 |
|  | Ave | 60269 | 0 | 16660 | 128.21 | 480.21 | 1553.20 | 78.34 |
|  | STD | 800 | 0 | 661.65 | 2.03 | 4.10 | 52.31 | 0.61 |

## Number of ORs and Number of Setups

From the managerial perspective, number of required ORs, i.e., $N$ should be minimized to reduce the overhead, staffing and equipment costs. Figure 3.1 a shows the histogram of
number of required ORs for $P T D$ and $L P T$. As shown in Figure 3.1. a and Table 3.2, the average number of required ORs for $P T D$ is significantly smaller than that of $L P T$. A $t$-test was performed to test the significance of difference between the two methods, with $p$-value of $p<0.001$ the difference between the two methods is statistically significant. The calculated number of ORs is for a planning horizon of 5 days. Thus, the average daily number of ORs required by $P T D$ roughly equals to 20 that is compatible with the capacity of medium size hospitals.

Number of setups represents the number of changes in equipment settings to accommodate different surgery types. PTD avoids combining different surgery types into the same OR which reduces the number of setups on average about 4.3 times compared to $L P T$. Figure 3.1. b shows the histogram of number of setups for $P T D$ and $L P T$. A $t$-test was performed to test the significance of difference between the two methods, with $p$-value of $p<0.001$ the difference between the two methods is statistically significant. As it is shown by Table 3.2, the number of setups and number of required ORs are approximately equal by $P T D$, meaning that each OR is equipped only once on each day that significantly decreases the total cost and ideltime.


Figure 3.1: Histograms of number of ORs and number of setups

## Idle time and Overtime

Idle time represents the proportion of total available time that elapsed idle either waiting for the start of the next surgery or due to lack of a compatible surgery. As Figure 3.2, a shows, on average the $P T D$ reduces the idle time almost 2.5 times compared with $L P T$. A $t$-test was performed to test the significance of difference between the two methods, with $p$ value of $p<0.001$ the difference between the two methods is statistically significant. LPT does not allow packing surgery cases (items) beyond the capacity of the ORs (bin), thus, overtime is not allowed, but PTD allows for overtime only for the last case of each surgery type in order to reduce the number of required ORs, number of setups, and idletime. Reducing the number of setup significantlly affect the total cost of the planning horizon because each setup not only causes setup cost but also results in more idle time. Figure 3.2 b shows the overtime for $P T D$ and $L P T$. As shown by Table 3.2, PTD on average generates 286 minutes of overtime for average number of ORs equal 111, meaning that PTD on average generate 2.5 minutes of overitme in each OR which is negligible. On the other hand, PTD on average generates 3946 minutes of idletime which is 3.79 times less than that of $L P T$. Therefore, we can conclude that PTD successfully balances the trade-offs between idletime and overtime.

a) Idle time

b) Overtime

Figure 3.2: Histograms of idle time and overtime and

## Utilization and Smoothness Index

Equation(3.8) defines the utilization of ORs as the percentage of total available time spent on performing surgeries. Figure 3.3. b shows the histogram of utilization for PTD and $L P T$. With average utilization of $89.83 \%$ PTD outperforms $L P T$ by roughly $10.4 \%$. The reason for the better performance of $P T D$ is the Step. 5 in $P T D$ that searches for the same surgery types from lower priority groups, and also for the last case of each surgery group overtime is allowed by PTD that slightly leads to overtime but decreases the idle time significantly. A $t$-test was performed to test the significance of difference between the two methods, with $p$ value of $p<0.001$ the difference between the utilization of the two methods is statistically significant.

Smoothness index (SI) represents the evenness of load distribution among ORs. The smaller the $S I$ the more even load distribution. Figure 3.3. a shows the histogram of $S I$ for $P T D$ and $L P T$. As shown by Table 3.2 on average $S I=1195.47$ for $P T D$, and $S I=1553.20$ for $L P T$, meaning that $P T D$ more evenly distributes the load among the ORs. A $t$-test was performed to test the significance of difference between the two methods, with $p$ value of $p<0.001$ PTD outperforms LPT.


Figure 3.3: Histograms of smoothness index and utilization

## Total Cost

Using Equation(3.2) and average values of overtime, idletime, regular time, number of setups, we calculate the average total cost for $P T D$ and $L P T$ to compare the total cost for both methods, we consider the average cost of regular time as $C_{R}=\$ 60$. To avoid prejudice in favor of the number of setups, we vary the setup cost between 0 and $\$ 2000$. As it is shown by Figure 3.4 the total cost of $P T D$ is significantly lower than that of $L P T$ regardless of the setup cost. The PTD total cost has a lower sensitivity to setup cost and ranges over [4.03, 4.26] million dollars with different setup costs, as opposed to $L P T$ total cost that varies steeply ranging over [4.61, 5.59] million dollars.


Figure 3.4: Total cost with different setup costs

### 3.5 Conclusion

Operating room tactical planning is an important phase for OR management in which efficient resource allocation is the main objective. Regular working time, overtime, number of ORs, and equipment are some indicators for capacity dimensioning at the tactical planning phase. An efficient OR planning assesses the tradeoff among these indicators, this assessment is frequently based on financial indicators. However, the cost structure of operating room is often complex. It makes the tactical planning phase more complicated. In this chapter, we proposed a multi-step procedure to assign surgeries to ORs on a weekly planning horizon. Our proposed procedure called PTD groups surgeries according to their priority, surgery type and surgery duration in order to form the initial sequence for the bin-packing problem. PTD reduces the idletime, number of required ORs, and number of setups that leads to a higher utilization and even load distribution among ORs. By taking surgery types into consideration $P T D$ reduces the number of required setups. Longest processing time first ( $L P T$ ) rule is the most common rule in bin-packing, thus, we compare the performance of PTD with those of $L P T$. The $L P T$ rule is merely based on the surgery duration and fails to address the surgery priorities and surgery types concerns. Priority is the level of urgency of a surgery, thus, surgeries with higher priorities must be performed earlier than those of lower priorities. PTD successfully addresses the priorities concerns. From the cost perspective, PTD compared to $L P T$ significantly reduces the idletime and number of required setups, leading to a higher utilization and significant lower total cost. Although PTD generate more overtime comapred
to $L P T$, it reduces the idletime significantly that leads to a lower total cost. In other word, $P T D$ balances the trade-offs between idletime, overtime, and number of required setups.

## Chapter 4

## Operational Operating Room Scheduling by Smoothing Patient Flow

In operational level of operating room scheduling, patient flow time mean (PtF) and patient flow time variance ( $P t F V$ ) are two key performance indicators driving many areas within a hospital. Surgical case sequences significantly affect $P t F$ and $P t F V$. Sequencing to $\min (P t F)$ reduces patients' waiting time and overcrowding between the preoperative stage and the OR. However, $\min (P t F)$ results in a rushed flow to the post-operative stage generating overcrowding between the OR and the post-op stage that negatively affect the quality of care. Sequencing to $\min (P t F V)$ ensures an approximately the same waiting time for all patients, and balances the OR intake and discharge flows. However, $\min (P t F V)$ negatively impacts PtF and consequently the staff and patients' satisfaction. By modeling OR as a single-machine production system, we show that $\min (P t F)$ and $\min (P t F V)$ are inconsistent objectives. Therefore, any sequence generates some levels of trade-offs in the system. We propose balancing trade-offs between $\min (P t F)$ and $\min (P t F V)$ as an alternative objective function, we also propose a fast algorithm to find the optimal sequence to minimizing tradeoffs in $O(n \log n)$. Through extensive case studies, we demonstrate the superiority of our trade-off balancing models over the singe-attribute and existing bi-criteria models. Balancing trade-offs enables OR managers to shift the outcomes of ORs towards a more predictable yet near optimal state.

### 4.1 Introduction

An OR peri-operative process normally consists of three sequential stages of pre-op, intra-op, and post-op, where the following activities occur: patient preparation in the pre-op stage, surgical intervention in the intra-op stage, and post anesthesia care in the post-op stage. Patient flow $(P t F)$ is the movement of the patients across the boundaries of the OR periop process. Smoothing patient flow can reduce the waiting times, overcrowding, and poor handoffs that consequently improves the quality of care [77]. Studies show that surgical case sequencing can significantly affect the outcomes of OR peri-op process [4, 78, 79, 80, 35]

With ORs being the most expensive resource in the OR peri-op process [81], the managerial efforts have been focused on optimizing ORs efficiency [35]. Minimizing OR patient flow time i.e. $\min (P t F)$ that minimizes patient's waiting time i.e. "the time interval from the time that the patient is available to the time that the surgery starts" 35,82 is one the most studied objectives. The common practice to $\min (P t F)$ is to sequence surgical cases in the non-decreasing order (i.e. shortest processing time first $(S P T)$ ). Sequencing with
$S P T$ rule minimizes the patients' waiting time and reduces overcrowdings between pre-op and intra-op stages. However, scheduling by $S P T$ rule results in a rushed patient flow to the post-op stage at the beginning of the schedule when a large number of patients with short surgical case times are discharged from the OR generating overcrowding in the post-op stage. Studies show that rushed OR patient flow may negatively affect the outcomes of downstream resources such as the ICUs and the performance of support staff such as laboratory [83, 84]. Baker et al.(2009) [85] pointed out that ICU overcrowding may result in a premature discharge from the ICU that increases the risk of readmission to the ICU by 2.34 times. Marcon and Dexter(2006) [35] showed the negative impacts of overcrowding in the post-op stage on the PACUs staffing and OR overtime due to delays in PACU admissions. Therefore, smoothing the patient flow across the OR peri-op process is imperative to improve the quality of care.

Patient flow variance ( $P t F V$ ) is a measure to quantify the uniformity of patient flow time. Minimizing patient flow time variance i.e. $\min (P t F V)$ finds application whenever it is desirable to provide the patients with approximately the same treatment such that each patient's waiting time is the same as every other patient. Minimizing PtFV in ORs results in a balance between the flow to the ORs and the discharge flow from the ORs. Therefore, $\min (P t F V)$ is able to balance the overcrowdings between pre-op and intra-op stages, and between intra-op and post-op stages. However, as we show in 4.3, the objectives of $\min (P t F)$ and $\min (P t F V)$ are inconsistent, meaning optimizing one may be at the cost of the other. As we later show, schedules to $\min (P t F)$ have high fluctuations in patient flow times in which the longest surgical cases that commonly are more critical procedures suffer the highest waiting times. At the other hand, schedules to $\min (P t F V)$ have high average patient flow time that conflicts with the preferences of both surgeons and patients. Therefore, balancing the trade-offs between these two extremes is necessary to improve the patient flow across the OR peri-op process, staff and patients' satisfaction, and eventually the quality of care.

An operating room can be modeled as a single-machine production system [86, 87], where the surgical cases are treated as jobs, surgical case times as the job processing times, and the OR is thought of as a machine. In an $n$-job single-machine system, $C_{j}=\sum_{l=1}^{j} p_{l}$ is the completion of time of job $j$ in the sequence where $p_{l}$ is the processing time of job $l . C_{j}$ can be written as $C_{j}=\sum_{l=1}^{j-1} p_{l}+p_{j}$ where $\sum_{l=1}^{j-1} p_{l}$ is referred as the waiting time of job $j$ equivalent to the waiting time of patient $j$. Total completion time is the sum of all jobs completion times and defined as $T C T=\sum_{j=1}^{n} C_{j}$. Average completion time i.e. $A C T=T C T / n$ represents the time that a job on average spends in the systems which is equivalent to patient flow time $(P t F)$, thus, $\min (A C T)$ is a surrogate objective function for $\min (P t F)$. For a fixed $n$, minimizing average completion time $(\min (A C T))$ is the same as minimizing total completion time i.e. $\min (T C T)$, thus, hereafter, we use $\min (T C T)$ as the surrogate objective function for $\min (P t F)$. The optimal sequence to $\min (T C T)$ is obtained by ordering the jobs in a non-decreasing order (i.e. by $S P T$ rule) [43].

Completion time variance is defined as $C T V=\left(\sum_{j=1}^{n}\left(C_{j}-A C T\right)^{2}\right) / n$ and represent the fluctuations in completion times. Minimizing completion time variance i.e. $\min (C T V)$ in a single-machine production system has been extensively studied [88, 89, 90, 91. Kubiak(1993) [89] proved that $\min (C T V)$ in a single-machine systems is $N P$-hard. Multiple algorithms have been proposed to find the optimal sequence to $\min (C T V)$ [88, 90, 92]. Kanet(1981) [90] proposed the total absolute differences in completion times defined by $T A D C=\sum_{j=1}^{n} \sum_{i=1}^{n}\left|C_{j}-C_{i}\right|$ as an alternative measure to completion time variance, and proved that TADC has the same complexity as $C T V$ but the optimal solutions to $\min (T A D C)$ are easier to find. Kanet(1981) [90] also proposed a fast heuristic to find the optimal solutions to $\min (T A D C)$. Because of the advantages of $T A D C$ over $C T V$, we adopt
$\min (T A D C)$ as the surrogate function for $\min (P t F V)$.
Surgical case sequencing significantly affects the outcomes of operating rooms and the associated upstream and downstream resources. Patient flow time mean and patient flow time variance are two inconsistent key performance indicators in operating room scheduling that derive many other areas within a hospital. However, with inconsistencies among objectives high levels of trade-offs are generated in the system. We propose trade-off balancing models to reduce the negative impacts of inconsistencies on the operating room outcomes. Through extensive case studies, we demonstrate the efficiency and effectiveness of trade-off balancing models in shifting the OR outcomes towards a more predictable state with minimum deviation for the 'ideal point' at which all objective are at their optima, with inconsistencies among the objectives the ideal point is infeasible and only used as a point of reference.

The remainder of this paper is organized as follows: in Section 4.2, we present the mathematical formulations of our trade-off balancing models. In Section 4.5, we presents the details of the case studies and the evaluation schemes to systematically evaluate the performance of sequencing methods on the University of Kentucky Healthcare (UKHC) OR historical data. In Section 4.6, we present results and discussions. Finally, in Section 4.7, we draw conclusions and present future research directions.

### 4.2 Problem description

As it was mentioned in Section 4.1, we model an operating room as a single machine production system, where the objectives of $\min (T C T)$ and $\min (T A D C)$ are surrogate objectives for $\min (P t F)$ and $\min (P t F V)$, respectively. In this section, given the vector of processing times $\mathbf{P}$, we present the mathematical formulations to balancing trade-offs between $\min (T C T)$ and $\min (T A D C)$ in the form $Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\alpha y_{1}+(1-\alpha) y_{2}$, where $\alpha$ is weight of $\min (T C T)$ in the trade-off function, $\boldsymbol{\sigma}$ is a sequence of jobs, $y_{k}=\frac{\gamma(\boldsymbol{\sigma}, \mathbf{P})-L B_{k}(\mathbf{P})}{U B_{k}(\mathbf{P})-L B_{k}(\mathbf{P})}$ is the normalized deviations from the best (minimum) possible value of attribute $k, U B_{k}(\mathbf{P})$ and $L B_{k}(\mathbf{P})$ are the best (minimum) and the worst (maximum) possible value of attribute $k$, respectively. It is worth mentioning that in order to distinguish between scalars and vectors, we use bold font to represent vectors (e.g. $\boldsymbol{\sigma}$ vs. $\sigma(i)$ ).

In 4.2 and 4.2 , we present methods for calculating $L B_{k}$ and $U B_{k}$. In 4.3, we prove the inconsistency between the objectives of $\min (T C T)$ and $\min (T A D C)$. Finally, in 4.3, we propose a heuristic to balance trade-offs between $\min (T C T)$ and $\min (T A D C)$.

## Notations

| $n$ | Number of jobs |
| :--- | :--- |
| $p_{j}$ | Processing time of job $j, j=1,2,, n$ |
| $\mathbf{P}$ | Vector of job processing times, $\mathbf{P}=\left[p_{1}, p_{2}, \ldots, p_{n}\right]^{T}$ |
| $\boldsymbol{\Omega}$ | Decision space, $\|\boldsymbol{\Omega}\|=n!$ |
| $\boldsymbol{\sigma}$ | $\boldsymbol{\sigma} \in \boldsymbol{\Omega}$, a permutation of of $n$ jobs |
| $i$ | Index of positions in $\boldsymbol{\sigma}, i=1,2, \ldots, n$ |
| $\sigma(i)$ | Job that occupies position $i$ in sequence $\boldsymbol{\sigma}=[\sigma(1), \sigma(2), \ldots, \sigma(n)]$ |
| $k$ | Index of attributes, $k=1,2$, with $k=1$ for $T C T$ and $k=2$ for $T A D C$ |

$$
\begin{array}{ll}
\gamma_{k}(\boldsymbol{\sigma}, \mathbf{P}) & \text { A function describing attribute } k \text { given } \boldsymbol{\sigma} \text { and } \mathbf{P} \\
w_{k}(i) & \text { Weight of position } i \text { in objective function } \gamma_{k} \\
\mathbf{W}_{k} & \text { Vector of positional weights in } \gamma_{k}, \mathbf{W}_{k}=\left[w_{k}(1), w_{k}(2), \ldots, w_{k}(n)\right] \\
x_{j, i} & \text { Binary decision variable, } x_{j, i}=1 \text { if } \sigma(i)=j, \text { otherwise } 0 \\
\mathbf{X} & n \times n \text { assignment matrix where } \mathbf{X}(j, i)=x_{j, i} \\
L B_{k}(\mathbf{P}) & \text { The minimum possible value for attribute } k \text { given } \mathbf{P} \\
U B_{k}(\mathbf{P}) & \text { The maximum possible value for attribute } k \text { given } \mathbf{P} \\
y_{k}(\boldsymbol{\sigma}, \mathbf{P}) & \text { Normalized deviation from the best value of attribute } k \text { given } \boldsymbol{\sigma} \text { and } \mathbf{P} \\
\alpha & \text { Weight of } \min \left(\gamma_{1}\right) \text { in the objective function, } \alpha \in[0,1]
\end{array}
$$

## Total completion time ( $T C T$ )

Given $n$ jobs, there are $n$ ! possible sequences i.e. $|\boldsymbol{\Omega}|=n$ !, where $\boldsymbol{\Omega}$ is the decision space. Let $\boldsymbol{\sigma} \in \boldsymbol{\Omega}$ be a sequence of $n$ jobs, and $i \in\{1,2, \ldots, n\}$ be the index of positions in $\boldsymbol{\sigma}$ i.e. $\boldsymbol{\sigma}=\{\sigma(1), \sigma(2), \ldots, \sigma(n)\}, \sigma(i)=j$ if job $j$ is assigned to position $i$ in $\boldsymbol{\sigma}$. For a deterministic single-machine production system with $n$ jobs, we can formulate total completion time (TCT) by Equation 4.1, where $C_{\sigma(i)}$ is the completion time of the job in position $i$. Note that the positional weight $w_{1}(i)=(n-i+1)$ is independent of $p_{\sigma(i)}$ and can be written in the vector form of $\mathbf{W}_{1}$, where $\mathbf{W}_{1}(1, i)=w_{1}(i)$.

$$
\begin{equation*}
\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})=\sum_{i=1}^{n} C_{\sigma(i)}=\sum_{i=1}^{n}(n-i+1) p_{\sigma(i)}=\sum_{i=1}^{n} w_{1}(i) p_{\sigma(i)} \tag{4.1}
\end{equation*}
$$

By introducing the binary decision variable $x_{j, i}=1$, if $\sigma(i)=j$, otherwise 0 , we can write $\boldsymbol{\sigma}$ in the form of an assignment matrix of $\mathbf{X}$, where $\mathbf{X}(j, i)=x_{j, i}$. Thus, Equation(4.1) is written in the matrix form as $\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})=\mathbf{W}_{1} \mathbf{X P}$. The mixed integer programming (MIP) models to find $L B_{1}(\mathbf{P})$ and $U B_{1}(\mathbf{P})$ are presented by Equation 4.2) and Equation 4.6), respectively. $L B_{1}(\mathbf{P})$ and $U B_{1}(\mathbf{P})$ can be solved in $O(n \log (n))$ by ordering jobs in the nondecreasing order $\left(\boldsymbol{\sigma}_{1}(\mathbf{P})=S P T\right)$, and non-increasing order (Longest Processing Time first, $L P T$ rule) $\left(\overline{\boldsymbol{\sigma}}_{1}(\mathbf{P})=L P T\right)$, respectively. Therefore, the range of $T C T$ is calculated as $R_{1}(\mathbf{P})=U B_{1}(\mathbf{P})-L B_{1}(\mathbf{P})=\gamma_{1}\left(\overline{\boldsymbol{\sigma}}_{1}, \mathbf{P}\right)-\gamma_{1}\left(\stackrel{+}{\boldsymbol{\sigma}}_{1}, \mathbf{P}\right)$ in $O(n \log n)$ as well.

$$
\begin{gather*}
L B_{1}(\mathbf{P})=\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})\right)=\min _{\mathbf{X} \in F}\left(\mathbf{W}_{1} \mathbf{X P}\right)  \tag{4.2}\\
\text { s.t. } \sum_{i=1}^{n} x_{j, i}=1  \tag{4.3}\\
\sum_{j=1}^{n} x_{j, i}=1  \tag{4.4}\\
x_{j, i} \in\{0,1\}  \tag{4.5}\\
U B_{1}(\mathbf{P})=\max _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})\right)=\max _{\mathbf{X} \in F}\left(\mathbf{W}_{1} \mathbf{X P}\right)  \tag{4.6}\\
\text { s.t. } \\
4.3), 4.4,4.5
\end{gather*}
$$

## Total absolute deviations in completion times (TADC)

For a deterministic single-machine production system with $n$ jobs, we can formulate total absolute differences in completion times ( $T A D C$ ) by Equation (4.7).

$$
\begin{equation*}
\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})=\sum_{i=1}^{n} \sum_{l=1}^{n}\left|C_{\sigma(i)}-C_{\sigma(l)}\right|=\sum_{i=1}^{n}(i-1)(n-i+1) p_{\sigma(i)} \tag{4.7}
\end{equation*}
$$

The positional weight of $w_{2}(i)=(i-1)(n-i+1)$ is independent of $p_{\sigma(i)}$, thus, can be written in the vector form of $\mathbf{W}_{2}$, where $\mathbf{W}_{2}(1, i)=w_{2}(i)$. By using the assignment matrix X, Equation $\sqrt{4.7})$ is rewritten as $\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})=\mathbf{W}_{2} \mathbf{X P}$. The mixed integer programming (MIP) models to find $L B_{2}(\mathbf{P})$ and $U B_{2}(\mathbf{P})$ are presented by Equation 4.8) and Equation 4.9), respectively.

$$
\begin{align*}
& L B_{2}(\mathbf{P})=\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)=\min _{\mathbf{X} \in F}\left(\mathbf{W}_{2} \mathbf{X P}\right)  \tag{4.8}\\
& \text { s.t. }4.3), 4.4), 4.5 \\
& U B_{2}(\mathbf{P})=\max _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)=\max _{\mathbf{X} \in F}\left(\mathbf{W}_{2} \mathbf{X P}\right)  \tag{4.9}\\
& \text { s.t. } 4.3,4.44
\end{align*}
$$

Kubiak (1993) [89] proved that the completion time variance minimization problem is $N P$ hard. Intuitively, we can conclude that problems (4.8) and (4.9) are also $N P$-hard. Kanet (1981) [90] showed that the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ is $V$-shaped. In a $V$-shaped sequence, jobs before and after the job with the shortest processing time are ordered by LPT and $S P T$ rules, respectively. Figure 4.1. a shows an example of a $V$-shaped sequence.


Figure 4.1: $V$-shaped and non- $V$-shaped processing sequences

Kanet's [90] heuristic to find $\stackrel{+}{\sigma}_{2}(P)$ the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ is presented by Algorithm 2 .

## Lemma 1.

$\stackrel{+}{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ must have the following properties [90]:

1. $\stackrel{+}{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ is $V$-shaped.
2. The longest job must be scheduled first

Algorithm $2 \stackrel{+}{\boldsymbol{\sigma}}_{2}(\mathbf{P})$, the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right), 90$
Let:
$U$ be the of unscheduled jobs, $B, A$ be empty sets;
For $j=1: n$
Remove the longest jobs from $U$ and label it $J_{j}$ if $j$ is odd place $J_{j}$ in the last empty position of $B$ else
place $J_{j}$ in the first empty position of $A$

$$
\begin{aligned}
& \quad \text { End } \\
& \stackrel{+}{\boldsymbol{\sigma}}_{2}(\mathbf{P})=\{B, A\}
\end{aligned}
$$

Proof of Lemma 11:
If $\boldsymbol{\sigma}$ is the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ and is NOT $V$-shaped, then there are three consecutive jobs $(i, j, k)$ such that $p_{j}>p_{i}$ and $p_{j}>p_{k}$, as depicted in Figure 4.1. b .

We prove that $\boldsymbol{\sigma}$ cannot be the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ because exchanging either $i$ and $j$ or $j$ and $k$ will result in a smaller $\gamma_{2}(\mathbf{P})$. Let $\boldsymbol{\sigma}^{\prime}$ be the sequence generated by exchanging jobs $i$ and $j$. Similarly, let $\sigma^{\prime \prime}$ be the sequence generated by exchanging jobs $j$ and $k$. We need to prove the following:

$$
\begin{aligned}
& \text { I. } \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)<0 \Rightarrow \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)>0 \\
& \text { II. } \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)<0 \Rightarrow \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)>0
\end{aligned}
$$

Proof for II. is similar to the proof for T., therefore, we only prove I. Let $e$ be the position of job $i$ in sequence $\boldsymbol{\sigma}$, using Equation(4.7), we have:

$$
\begin{gather*}
\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})=\sum_{h=1}^{e-1}(h-1)(n-h+1) p_{h}+(e-1)(n-e+1) p_{i}+(e)(n-e) p_{j}+(e+1)(n-e-1) p_{k}+ \\
\sum_{h=e+3}^{n}(h-1)(n-h+1) p_{h}  \tag{4.10}\\
\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)=\sum_{h=1}^{e-1}(h-1)(n-h+1) p_{h}+(e-1)(n-e+1) p_{j}+(e)(n-e) p_{i}+(e+1)(n-e-1) p_{k}+ \\
\sum_{h=e+3}^{n}(h-1)(n-h+1) p_{h}  \tag{4.11}\\
\gamma_{2}\left(\boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)=\sum_{h=1}^{e-1}(h-1)(n-h+1) p_{h}+(e-1)(n-e+1) p_{i}+(e)(n-e) p_{k}+(e+1)(n-e-1) p_{j}+ \\
\sum_{h=e+3}^{n}(h-1)(n-h+1) p_{h} \tag{4.12}
\end{gather*}
$$

Equation (4.14) states that $\boldsymbol{\sigma}^{\prime \prime}$ is strictly better than $\boldsymbol{\sigma}$, therefore, $\stackrel{\rightharpoonup}{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ must be $V$ shaped.

## Proof of Lemma 1. 2

Assume $\boldsymbol{\sigma}$ is the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ and the longest job with processing time $p_{L}$ is NOT scheduled first i.e. $p_{\sigma(1)} \neq p_{L}, p_{\sigma(i)}=p_{L}, i \in\{2,3, \ldots, n\}$. Let $\boldsymbol{\sigma}^{\prime}$ be the sequence obtained from $\boldsymbol{\sigma}$ by exchanging the position of the first job and the longest job (which is in position $e$ ). We have the followings:

$$
\begin{align*}
\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})= & 0 \times p_{\sigma(1)}+(e-1)(n-e+1) p_{L}+\sum_{h \neq e}^{n}(h-1)(n-h+1) p_{\sigma(h)}  \tag{4.15}\\
\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)= & 0 \times p_{L}+(e-1)(n-e+1) p_{\sigma(1)}+\sum_{h \neq e}^{n}(h-1)(n-h+1) p_{\sigma(h)}  \tag{4.16}\\
& \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)=\underbrace{(e-1)(n-e+1)}_{>0} \underbrace{\left(p_{L}-p_{\sigma(1)}\right)}_{>0}>0 \tag{4.17}
\end{align*}
$$

Equation 4.17) clearly states that $\boldsymbol{\sigma}^{\prime}$ is strictly better than $\boldsymbol{\sigma}$, therefore, the longest job must be scheduled first in $\stackrel{+}{\boldsymbol{\sigma}}_{2}(\mathbf{P})$.

We show that $\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ the optimal sequence to $\max _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ is $\Lambda$-shaped. In a $\Lambda$-shaped sequence, jobs before and after the job with the longest processing time are ordered by $S P T$ and $L P T$ rules, respectively. Figure 4.2 shows an example of a $\Lambda$-shaped sequence. We propose Algorithm 3 to find $\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})$, which has the same complexity as that of Algorithm 2 i.e. $O(n \log n)$.


Figure 4.2: $\Lambda$-shaped processing sequence

## Lemma 2.

$\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ must have the following properties:

1. $\bar{\sigma}_{2}(\mathbf{P})$ is $\Lambda$-shaped.
2. The shortest job must be scheduled first

Proof of Lemma 2 1
If $\boldsymbol{\sigma}$ is the optimal sequence to $\max _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ and is NOT $\Lambda$-shaped, then there are three consecutive jobs $(i, j, k)$ such that $p_{j}<p_{i}$ and $p_{j}<p_{k}$, as depicted in Figure 4.3 .

We prove that $\boldsymbol{\sigma}$ cannot be the optimal sequence to $\max _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ because exchanging either $i$ and $j$ or $j$ and $k$ will result in a greater $\gamma_{2}(\mathbf{P})$. Let $\boldsymbol{\sigma}^{\prime}$ be the sequence obtained form $\boldsymbol{\sigma}$ by exchanging jobs $i$ and $j$. Similarly, let $\boldsymbol{\sigma}^{\prime \prime}$ be the sequence generated by exchanging jobs $j$ and $k$. We need to prove the following:


Figure 4.3: A non- $\Lambda$-shaped processing sequence

$$
\begin{aligned}
& \text { I. } \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)>0 \Rightarrow \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)<0 \\
& \text { II. } \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)>0 \Rightarrow \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)<0
\end{aligned}
$$

Proof for II. is similar to the proof for T. therefore, we only prove I. Let $e$ be the position of job $i$ in $\boldsymbol{\sigma}$, then:

$$
\begin{gather*}
\text { By assumption } \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)=\underbrace{\left(p_{i}-p_{j}\right)}_{>0}(2 e-n-1)>0 \\
\Rightarrow(2 e-n-1)>0  \tag{4.18}\\
\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)=\underbrace{\left(p_{j}-p_{k}\right)}_{<0} \underbrace{((2 e-n-1)}_{>0}+2) \\
\Rightarrow \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)<0 \tag{4.19}
\end{gather*}
$$

Equation(4.19) clearly states that $\sigma^{\prime \prime}$ is strictly better than $\boldsymbol{\sigma}$, therefore, $\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ must be $\Lambda$-shaped.

Proof of Lemma 2 2
Assume $\boldsymbol{\sigma}$ is the optimal sequence to $\max _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ and the shortest job with processing time $p_{S}$ is NOT scheduled first i.e. $p_{\sigma(1)} \neq p_{S}, p_{\sigma(i)}=p_{S}, i \in\{2,3, \ldots, n\}$. Let $\boldsymbol{\sigma}^{\prime}$ be the sequence obtained from $\boldsymbol{\sigma}$ by exchanging the position of the first job and the shortest job (which is in position $e$ ). We have the following:

$$
\begin{align*}
\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})= & 0 \times p_{\sigma(1)}+(e-1)(n-e+1) p_{S}+\sum_{h \neq e}^{n}(h-1)(n-h+1) p_{\sigma(h)}  \tag{4.20}\\
\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)= & 0 \times p_{S}+(e-1)(n-e+1) p_{\sigma(1)}+\sum_{h \neq e}^{n}(h-1)(n-h+1) p_{\sigma(h)}  \tag{4.21}\\
& \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-\gamma_{2}\left(\boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)=\underbrace{(e-1)(n-e+1)}_{>0} \underbrace{\left(p_{S}-p_{\sigma(1)}\right)}_{<0}<0 \tag{4.22}
\end{align*}
$$

Equation (4.22) clearly states that $\boldsymbol{\sigma}^{\prime}$ is strictly better than $\boldsymbol{\sigma}$, therefore, the shortest job must be scheduled first in $\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})$.

By utilizing Algorithm 2 and Algorithm 3, we are able to calculate $R_{2}(\mathbf{P})=U B_{2}(\mathbf{P})-$ $L B_{2}(\mathbf{P})=\gamma_{2}\left(\overline{\boldsymbol{\sigma}}_{2}, \mathbf{P}\right)-\gamma_{2}\left(\stackrel{\rightharpoonup}{\boldsymbol{\sigma}}_{2}, \mathbf{P}\right)$ in $O(n \log n)$.

## Algorithm $3 \overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})$, the optimal sequence to $\max _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$

Let:
$U$ be the of unscheduled jobs, $B, A$ be empty sets; For $j=1: n$

Remove the shortest jobs from $U$ and label it $J_{j}$ if $j$ is odd place $J_{j}$ in the first empty position of $B$ else
place $J_{j}$ in the last empty position of $A$ End
$\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})=\{B, A\}$

### 4.3 Inconsistency between $\min (T C T)$ and $\min (T A D C)$

From the shape of $\stackrel{+}{\boldsymbol{\sigma}}_{1}(\mathbf{P})$ and $\stackrel{+}{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ (i.e. $S P T$ and $V$-shaped, respectively), we can intuitively conclude that the objective of $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})\right)$ and $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ are inconsistent. We also provide a proof of this inconsistency in Appendix A.1.

Lemma 3. $\min (T C T)$ and $\min (T A D C)$ are inconsistent.
A system of linear equations is called inconsistent if it has no solutions [39]. We show that $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})\right)$ and $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ are inconsistent, consequently, there is no sequence that simultaneously minimizes both. Given inconsistency between objectives, any sequence generates some levels of trade-offs in the system. Therefore, balancing trade-offs is an interesting objective.

## Trade-off balancing

We use compromised programming $(C P)$, which was first proposed by Zeleny and Cochrane (1973) [9] to the define the objective function of trade-off balancing. The first step in $C P$ is to establish an 'ideal point', the coordinates of the ideal point are given by the optimum values of all objectives. It is obvious that with inconsistent objectives the ideal point is not feasible, therefore, the ideal point is only a point of reference for $C P$. The second step in $C P$ is to establish an 'anti-ideal' point. The coordinates of the anti-ideal point are given by the worst values of all objectives. The objective of $C P$ is to find the closest efficient solution to the ideal point [40]. In the minimization sense, the coordinates of ideal point and anti-ideal point are $\left(L B_{1}, L B_{2}, \ldots, L B_{k}\right)$ and $\left(U B_{1}, U B_{2}, \ldots, U B_{k}\right)$, respectively. Therefore, the degree of closeness between solution $\boldsymbol{\sigma}$ to the coordinate $k$ of the ideal point is defined by $y_{k}(\boldsymbol{\sigma})=\frac{\gamma_{k}(\boldsymbol{\sigma})-L B_{k}}{U B_{k}-L B_{k}}$.

To measure the distances between outcomes of solution $\boldsymbol{\sigma}$ and the ideal point, a family of distance functions are introduced by $L_{g}\left(\alpha_{k}, \boldsymbol{\sigma}\right)=\left(\sum_{k=1}^{K}\left(\alpha_{k} y_{k}(\boldsymbol{\sigma})\right)^{g}\right)^{1 / g}$, where $\alpha_{k}$ is the weight of attribute $k$ in the objective function, without loss of generality we assume $\sum_{k=1}^{K} \alpha_{k}=$ 1. When $g=1, L_{1}$ measures the longest distance (geometrically speaking) between the solution and the ideal point. The best compromise or the closest solution to ideal point is obtained by solving the linear programming (LP) model presented by Equation 4.23). Where,
$F$ is the set of feasible solutions.

$$
\begin{equation*}
\min \quad L_{1}=\sum_{k=1}^{K} \alpha_{k} y_{k}, \quad \text { s.t. } \quad x \in F \tag{4.23}
\end{equation*}
$$

In balancing trade-offs between $\min (T C T)$ and $\min (T A D C)$, we can reformulate Equation (4.23) into Equation 4.24 , where $\alpha$ is the weight of $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})\right)$.

$$
\begin{gather*}
\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\sum_{k=1}^{2} \alpha_{k} y_{k} \Rightarrow \\
\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\alpha \frac{\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})-L B_{1}(\mathbf{P})}{R_{1}(\mathbf{P})}+(1-\alpha) \frac{\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})-L B_{2}(\mathbf{P})}{R_{2}(\mathbf{P})} \tag{4.24}
\end{gather*}
$$

Since it has been proven that $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$. is $N P$-hard, it can be speculated that $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$ is also $N P$-hard with $|\Omega|=n!$. Therefore, seeking the optimal solution to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})$ is not computationally efficient.

### 4.4 Optimal sequence to $\min _{\boldsymbol{\sigma} \in \Omega} Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})$

We prove that $\stackrel{+}{\boldsymbol{\sigma}}_{Z}(\mathbf{P})$ the optimal solution to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$ is $V$-shaped. Given $n$ jobs, there exist $2^{n-1} V$-shaped sequences, therefore, $\left|\Omega_{\min (Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))}\right|=2^{n-1}$.

Lemma 4. $\stackrel{+}{\boldsymbol{\sigma}}_{Z}(\mathbf{P})$; the optimal solution to $\min (Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$ is $V$-shaped.
Proof for Lemma 4 is similar to that of Lemma 1.1. and is also provided in Appendix A.2.
Equation(4.24) can be rewritten by Equation(4.25). The second term on the right hand side (RHS) of Equation 4.25) denoted by $\pi(\alpha, \mathbf{P})$ is a constant given $\alpha$ and $\mathbf{P}$, therefore, $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})$ is a surrogate function for $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})$.

$$
\begin{gather*}
\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\underbrace{\frac{\alpha}{R_{1}(\mathbf{P})} \gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})+\frac{(1-\alpha)}{R_{2}(\mathbf{P})} \gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})}_{Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})}-\underbrace{\left(\frac{\alpha L B_{1}(\mathbf{P})}{R_{1}(\mathbf{P})}+\frac{(1-\alpha) L B_{2}(\mathbf{P})}{R_{2}(\mathbf{P})}\right)}_{\pi(\alpha, \mathbf{P})}  \tag{4.25}\\
Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\frac{\alpha \sum_{i=1}^{n}(n-i+1) p_{\sigma(i)}}{R_{1}(\mathbf{P})}+\frac{(1-\alpha) \sum_{i=1}^{n}(i-1)(n-i+1) p_{\sigma(i)}}{R_{2}(\mathbf{P})} \Rightarrow \\
Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\frac{\alpha \sum_{i=1}^{n} w_{1}(i) p_{\sigma(i)}}{R_{1}(\mathbf{P})}+\frac{(1-\alpha) \sum_{i=1}^{n} w_{2}(i) p_{\sigma(i)}}{R_{2}(\mathbf{P})} \\
Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\sum_{i=1}^{n}\left(\frac{\alpha w_{1}(i)}{R_{1}(\mathbf{P})}+\frac{(1-\alpha) w_{2}(i)}{R_{2}(\mathbf{P})}\right) p_{\sigma(i)}=\sum_{i=1}^{n} w_{Z^{\prime}}^{[\alpha, \mathbf{P}]}(i) p_{\sigma(i)} \tag{4.26}
\end{gather*}
$$

We can rewrite $Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})$ into Equation 4.26). $w_{1}(i)$ and $w_{2}(i)$ are independent of $\mathbf{P}$, moreover, $R_{1}(\mathbf{P})$ and $R_{2}(\mathbf{P})$ are obtained as discussed in 4.2 and 4.2 , respectively. Thus, given $\alpha$ and $\mathbf{P}$, we are able to obtain $\mathbf{W}_{Z^{\prime}}^{[\alpha, \mathbf{P}]}=\left[w_{Z^{\prime}}^{[\alpha, \mathbf{P}]}(1), \ldots, w_{Z^{\prime}}^{[\alpha, \mathbf{P}]}(n)\right]$ that results in $Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})=$ $\mathbf{W}_{Z^{\prime}}^{[\alpha, \mathbf{P}]}\left[p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(n)}\right]^{T}$.

Note that $Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})$ is the sum of pair-wise products of two vectors of $\mathbf{W}_{Z^{\prime}}^{[\alpha, \mathbf{P}]}$ and $\left[p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(n)}\right]^{T}$, therefore, it can be minimized by arranging $\mathbf{W}_{Z^{\prime}}^{[\alpha, \mathbf{P}]}$ in the non-decreasing order and $\left[p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(n)}\right]^{T}$ in the non-increasing order [93]. We propose a fast heuristic with complexity of $O(n \log n)$ to find $\stackrel{+}{\boldsymbol{\sigma}}_{Z^{\prime}}(\alpha, \mathbf{P})$ the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})\right)$ as presented by Algorithm 4. It is worth reminding that $Z^{\prime}(\alpha, \boldsymbol{\sigma}, \mathbf{P})$ is the surrogate function for $Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})$, therefore, $\stackrel{+}{\boldsymbol{\sigma}}_{Z^{\prime}}(\alpha, \mathbf{P})$ is also the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$.

Algorithm $\mathbf{4}{\underset{\boldsymbol{\sigma}}{Z}}_{+[\alpha]}^{(\mathbf{P}), \text { the optimal sequence to } \min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})), ~(, ~}$
Given $\alpha$ and $\mathbf{P}$
Step. $1 \quad$ Obtain $R_{1}(\mathbf{P})=U B_{1}(\mathbf{P})-L B_{1}(\mathbf{P})=\gamma_{1}\left(\overline{\boldsymbol{\sigma}}_{1}(\mathbf{P}), \mathbf{P}\right)-\gamma_{1}\left(\stackrel{+}{\boldsymbol{\sigma}}_{1}(\mathbf{P}), \mathbf{P}\right)$, where $\overline{\boldsymbol{\sigma}}_{1}(\mathbf{P})$ and $\stackrel{+}{\boldsymbol{\sigma}}_{1}(\mathbf{P})$ are obtained by $L P T$ and $S P T$ rules, respectively.
Step. $2 \quad$ Obtain $R_{2}(\mathbf{P})=U B_{2}(\mathbf{P})-L B_{2}(\mathbf{P})=\gamma_{2}\left(\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P}), \mathbf{P}\right)-\gamma_{2}\left(\stackrel{\rightharpoonup}{\boldsymbol{\sigma}}_{2}(\mathbf{P}), \mathbf{P}\right)$, where $\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ and $\stackrel{+}{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ are obtained by Algorithm 3 and Algorithm 2 , respectively.
Step. 3 Obtain the positional weights $w_{Z^{\prime}}^{[\alpha, \mathbf{P}]}(i)=\frac{\alpha(n-i+1)}{R_{1}(\mathbf{P})}+\frac{(1-\alpha)(i-1)(n-i+1)}{R_{2}(\mathbf{P})}$, and form $\mathbf{W}_{Z^{\prime}}^{[\alpha, \mathbf{P}]}=\left[w_{Z}^{[\alpha, \mathbf{P}]}(1), \ldots, w_{Z^{\prime}}^{[\alpha, \mathbf{P}]}(n)\right]$.
Step. 4 Order $\mathbf{W}_{Z^{\prime}}^{[\alpha, \mathbf{P}]}$ in the non-decreasing order, arbitrarily break the ties, and call it $\mathbf{W}^{\prime}$.
Step. 5 Sort $\mathbf{P}$ by $S P T$ rule and call it $\boldsymbol{\sigma}$.
Step. 6 Obtain $\stackrel{+}{\boldsymbol{\sigma}}_{Z}(\alpha, \mathbf{P})$ by replacing each job in $\boldsymbol{\sigma}$ by the rank of its $w^{\prime}(i)$ in $\mathbf{W}$.

We demonstrate the determination of $\stackrel{+}{\boldsymbol{\sigma}}_{Z}(\alpha, \mathbf{P})$ by Algorithm 4 for a 7 -job example with $\alpha=0.4$ i.e. $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))=0.4 y_{1}+0.6 y_{2}$ as shown by Table 4.1.

Table 4.1: An example for demonstrating Algorithm 4

|  | P | 2 | 21 | 9 | 65 | 82 | 3 | 6 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Step. 1 | $\overline{\boldsymbol{\sigma}}_{1}(\mathbf{P})$ | 82 | 65 | 21 | 9 | 6 | 3 | 2 | $U B_{1}=1131$ | $R_{1}=785$ |
|  | $\stackrel{+}{\boldsymbol{\sigma}}_{1}(\mathbf{P})$ | 2 | 3 | 6 | 9 | 21 | 65 | 82 | $L B_{1}=373$ |  |
| Step. 2 | $\overline{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ | 2 | 6 | 21 | 82 | 65 | 9 | 3 | $U B_{2}=2118$ | $R_{2}=1392$ |
|  | $\stackrel{\rightharpoonup}{\boldsymbol{\sigma}}_{2}(\mathbf{P})$ | 82 | 21 | 6 | 2 | 3 | 9 | 65 | $L B_{2}=726$ |  |
| Step. 3 | $w_{Z}^{[\alpha, \mathbf{P}]}(i)$ | 0.0037 | 0.0058 | 0.0069 | 0.0073 | 0.0068 | 0.0054 | 0.0031 |  |  |
| Step. 4 | $\mathbf{w}^{\prime}$ | 0.0073 | 0.0069 | 0.0068 | 0.0058 | 0.0054 | 0.0037 | 0.0031 |  |  |
|  | Rank | 4 | 3 | 5 | 2 | 6 | 1 | 7 |  |  |
| Step. 5 | $\sigma$ | 2 | 3 | 6 | 9 | 21 | 65 | 82 |  |  |
| Step. 6 | $\stackrel{+}{\sigma}_{Z}(\alpha, \mathbf{P})$ | 65 | 9 | 3 | 2 | 6 | 21 | 82 |  |  |

## Bi-criteria models

Multiple works 92, 94 have been reported on optimizing the bi-criteria problem of $T C T$ $T A D C$ in the form $Z_{B C}(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\alpha T C T+(1-\alpha) T A D C$ as presented by Equation 4.27).

$$
\begin{gather*}
\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z_{B C}(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\alpha \gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})+(1-\alpha) \gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})= \\
\alpha \sum_{i=1}^{n} w_{1}(i) p_{\sigma(i)}+(1-\alpha) \sum_{i=1}^{n} w_{2}(i) p_{\sigma(i)} \Rightarrow \\
\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z_{B C}(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\sum_{i=1}^{n}\left(\alpha\left(w_{1}(i)-w_{2}(i)\right)+w_{2}(i)\right) p_{\sigma(i)}=\sum_{i=1}^{n} w_{B C}^{[\alpha]}(i) p_{\sigma(i)} \Rightarrow \\
\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z_{B C}(\alpha, \boldsymbol{\sigma}, \mathbf{P})=\mathbf{W}_{B C}^{[\alpha]}\left[p_{\sigma(i)}, \ldots, p_{\sigma(n)}\right]^{T} \tag{4.27}
\end{gather*}
$$

$\stackrel{+}{\boldsymbol{\sigma}}_{B C}^{[\alpha]}(\mathbf{P})$ the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z_{B C}(\alpha, \boldsymbol{\sigma}, \mathbf{P})$ is obtained by ordering $\mathbf{W}_{B C}^{[\alpha]}$ in nonincreasing order and $\left[p_{\sigma(i)}, \ldots, p_{\sigma(n)}\right]^{T}$ in the non-decreasing order. In order to demonstrate the effectiveness and efficiency of trade-off balancing, we compare the performance of $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$ with that of $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}} Z_{B C}(\alpha, \boldsymbol{\sigma}, \mathbf{P})$.

### 4.5 Case studies

By varying $\alpha$ the weight of $\min (T C T)$ in the objective functions with $\alpha \in\{0.0,0.1, \ldots, 1.0\}$, we examine the performance of sequencing methods with different objective functions. Therefore, we have 11 methods of $Z(\alpha)$ for the trade-off balancing problem, and 11 methods of $Z_{B C}(\alpha)$ for the bi-criteria problem.

Let $h$ be the index of methods where $h \in H=\{1,2, \ldots, 11,12,13, \ldots, 22\}$. Where $H \equiv$ $\left\{Z(0.0), Z(0.1), \ldots, Z(1.0), Z_{B C}(0.0), Z_{B C}(0.1), \ldots, Z_{B C}(1.0)\right\}$ denote the set of objective functions. OR Historical data of 260 days per year, for 5 years from the University of Kentucky HealthCare (UKHC) were used in this case study $(Q=1300$ samples in total). Let $q=1,2, \ldots, Q$ be the index of samples, and $\mathbf{P}_{q}$ be the surgical case times vector of surgeries in sample $q$.

Given $\mathbf{P}_{q}$, let $\boldsymbol{\sigma}_{h}\left(\mathbf{P}_{q}\right), h \in H$ denote the optimal sequence generated by method $h$. Thus, we have the normalized deviation of attribute $k$ generated by method $h$ from its possible best value as $y_{h, k, q}\left(\boldsymbol{\sigma}_{h}, P_{q}\right)=\frac{\gamma_{k}\left(\boldsymbol{\sigma}_{h}, \mathbf{P}_{q}\right)-L B_{k, q}}{U B_{k, q}-L B_{k, q}}$. Let $D_{h, k}=\frac{\sum_{q=1}^{Q} y_{h, k, q}}{Q}$ be the average normalized deviation of attribute $k$ generated by method $h$ over all $Q$ samples.

## Evaluation schemes

In order to systematically evaluate the performance of sequencing methods, we discuss the results from two perspectives (i) inconsistency between objectives of $\min (T C T)$ and $\min (T A D C)$, (ii) inconsistency between the first order (expected value) and the second order (variance) of trade-off. To address (i), we utilize the Pareto dominance conditions, and to address (ii), we use modern portfolio theory (MPT).

## Pareto dominance

For minimization problems, if $\mathbf{x}_{A}^{[k]}$ and $\mathbf{x}_{B}^{[k]} \in \mathbb{R}^{K}$ are two vectors that measure a positive attribute $k$ such as the utility of decision $A$ and $B$, respectively, decision $A$ dominates decision
$B$ if the following conditions are satisfied:

$$
\begin{array}{lll}
x_{A}^{[k]} \leq x_{B}^{[k]}, & \forall k \in\{1,2, . ., K\} \\
x_{A}^{[k]}<x_{B}^{[k]}, & \exists & k \in\{1,2, . ., K\} \tag{4.29}
\end{array}
$$

Equation (4.28) states that decision $A$ is not worse than decision $B$ in any dimension, while Equation (4.29) states that decision $A$ is better than decision $B$ at least in one dimension. Pareto optimal outcomes cannot be improved without the sacrificing of at least one objective.

## Modern portfolio theory

Modern portfolio theory (MPT) was first proposed by Markowitz (1952) [95]. MPT is a mathematical framework to balance the trade-offs between the expected value of return/loss and the associated risk. We formulate an MPT model for balancing the trade-off between the expected value of trade-off and its variance as the measure of risk. Assume that $y_{k}$ has a density function $p\left(y_{k}\right)$ with mean $\mu_{k}$ and variance $\sigma_{k}^{2}$. Now, we consider the case where the decision vector $\mathbf{x}$ represents a portfolio of preferences on $y_{k}$ such that $\mathbf{x}=\left(x_{1}, x_{2}\right)$ with the conditions that $x_{k} \geq 0$ and $x_{1}+x_{2}=1$ Let $\mathbf{Y}=\left(y_{1}, y_{2}\right)$ denote a random vector representing the normalized deviations. The distribution of $\mathbf{Y}$ is a joint distribution of $y_{1}$ and $y_{2}$ and is independent of $\mathbf{x}$ with density of $p(\mathbf{Y})$.

Let us define the loss of portfolio $\mathbf{x}$ as the sum of normalized deviations on individual $y_{k}$ scaled by proportion $x_{k}$. Therefore, the loss is given by Equation 4.30).

$$
\begin{equation*}
f(\mathbf{x}, \mathbf{Y})=\left[x_{1} y_{1}+x_{2} y_{2}\right]=\mathbf{x}^{T} \mathbf{Y} \tag{4.30}
\end{equation*}
$$

Given mean and variance of $y_{k}$, we have the vector of means $\mathbf{m}=\left(\mu_{1}, \mu_{2}\right)$ and the covariance matrix $\operatorname{COV}\left(y_{k}\right)$. Therefore, we are able to calculate the mean and variance of the loss associated with portfolio x as follows:

$$
\begin{gather*}
\mu_{f(\mathbf{x}, \mathbf{Y})}=\mathbf{x}^{T} \mathbf{m}  \tag{4.31}\\
\sigma_{f(\mathbf{x}, \mathbf{Y})}^{2}=\mathbf{x}^{T} \operatorname{COV} \mathbf{x} \tag{4.32}
\end{gather*}
$$

For each $\mathbf{x}$, the loss $f(\mathbf{x}, \mathbf{Y})$ is random variable with a distribution in $\mathbb{R}$ induced by $p(\mathbf{Y})$. In this work, efficient portfolio frontiers are utilized to assess the performance of the sequencing methods, in terms of $\mu_{f(\mathbf{x}, \mathbf{Y})}$ and $\sigma_{f(\mathbf{x}, \mathbf{Y})}^{2}$.

### 4.6 Results and Discussions

The performance of 22 sequencing methods including 11 trade-off balancing methods and 11 bi-criteria methods are evaluated with the UKHC OR historical data.

## Pareto dominance

Table 4.2 shows the statistics of the average normalized deviations for the sequencing methods. It is observed that the bi-criteria method i.e. $Z_{B C}(\alpha)$ is almost insensitive to $\alpha$ and it is strictly to the favor of $\min (T A D C)$ by generating large deviations from the best value of $T C T$ i.e. $D_{1}$. This behavior makes the bi-criteria model unsuitable for simultaneously optimizing $T C T$ and $T A D C$. On the other hand as it is shown by Figure 4.4, trade-off balancing methods i.e. $Z(\alpha)$ are sensitive to $\alpha$, thus, able to reflect the weights of objectives. Based on the average deviations none of the $Z(\alpha)$ methods are dominated.

Table 4.2: Statistics of average normalized deviations

| Method | Metric | $\alpha$ |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 |
| $Z(\alpha)$ | $D_{1}$ | 0.5075 | 0.4779 | 0.4220 | 0.3707 | 0.3016 | 0.2150 | 0.1279 | 0.0384 | 0.0000 | 0.0000 | 0.0000 |
|  | $D_{2}$ | 0.0000 | 0.0018 | 0.0118 | 0.0291 | 0.0663 | 0.1384 | 0.2470 | 0.4138 | 0.5200 | 0.5200 | 0.5200 |
|  | $\bar{D}$ | 0.2537 | 0.2398 | 0.2169 | 0.1999 | 0.1839 | $\mathbf{0 . 1 7 6 7}$ | 0.1874 | 0.2261 | 0.2600 | 0.2600 | 0.2600 |
| $Z_{B C}(\alpha)$ | $D_{1}$ | 0.5075 | 0.5075 | 0.5075 | 0.5075 | 0.5075 | 0.4925 | 0.4925 | 0.4779 | 0.4497 | 0.3831 | 0.0000 |
|  | $D_{2}$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0006 | 0.0006 | 0.0018 | 0.0057 | 0.0242 | 0.5200 |
|  | $\bar{D}$ | 0.2537 | 0.2537 | 0.2537 | 0.2537 | 0.2537 | 0.2466 | 0.2466 | 0.2398 | 0.2277 | 0.2036 | 0.2600 |



Figure 4.4: Pareto frontiers based on $D_{1}$ and $D_{2}$

Let $\bar{D}_{h}=(D 1+D 2) / 2$ denote the grand average normalized deviations generated by method $h$. As it is shown by Table 4.2 and Figure 4.5, trade-off balancing methods i.e. $Z(\alpha)$ dominate the bi-criteria methods in terms of $\bar{D}$ except for $\alpha=0.8$ and $\alpha=0.9$. Although at $\alpha=0.8$ and $\alpha=0.9$, respectively $80 \%$ and $90 \%$ of the weight in the objective function is assigned to $\min \left(\gamma_{1}\right)$, it is observed that $Z_{B C}(0.8)$ and $Z_{B C}(0.90)$ generate very large deviations of $44.97 \%$ and $38.31 \%$ from $L B_{1}$, respectively. The cause of this behavior of $Z_{B C}(\alpha)$ is the difference between the magnitudes of $T C T$ and $T A D C$ where $T A D C \gg T C T$ that forces the solutions towards $\min \left(\gamma_{2}\right)$. Therefore, we conclude that the bi-criteria methods are not solid and appropriate methods for simultaneously optimizing $T C T$ and $T A D C$, whereas, trade-off balancing methods by using the normalized deviations i.e. $y_{k}$ fairly treat both attributes.

It is also observed that $Z(0.5)$ generate minimum $\bar{D}$ with the value of 0.1767 . Another interesting observation from Table 4.2, Figure 4.4, and Figure 4.5 is that for $\alpha>0.7, \min \left(\gamma_{1}\right)$ dominates the solutions of $Z(\alpha)$ returning the $S P T$ sequence.

Since the performance of the bi-criteria methods are inferior to those of the trade-off balancing methods, we exclude the bi-criteria methods from our further discussions.

## Modern portfolio theory

Using Equation (4.31) and Equation 4.32, we calculate the loss generated by trade-off balancing functions i.e. $Z(\alpha)$ for a wide range of portfolios $\mathbf{x}_{m}=\left(x_{1}, x_{2}\right)$ such that $\mathbf{x}_{m} \in \mathbf{M}=$ $\{(0.0,1.0),(0.01,0.99), \ldots,(1.0,0.00)\}$ as shown by Figure 4.6. Our purpose by loss analysis is to examine the ability of sequencing methods to satisfy the needs of decision makers with a


Figure 4.5: Grand average normalized deviations of sequencing methods
different portfolio of preferences. Figure 4.6 shows that $Z(0.5)$ is the method with the least sensitivity to $\mathbf{x}$ with a small range of loss average and standard deviation.


Figure 4.6: Portfolio frontiers of trade-off balancing functions

In order to comprehensively evaluate the performance of the sequencing methods, let $\bar{\mu}_{h}=\frac{\sum_{m=1}^{|\mathbf{M}|} \mu_{f\left(\mathbf{x}_{m}, \mathbf{Y}_{h}\right)}}{\bar{M}^{\mid} \mid}$denote the average loss generated by method $h$ over all portfolios, similarly, $\bar{\sigma}_{h}$ denote the standard deviation of loss generated by method $h$ over all portfolios. Also let $\mu_{h}^{\max }=\max _{m \in \mathbf{M}}\left(\mu_{f\left(\mathbf{x}_{m}, \mathbf{Y}_{h}\right)}\right), \mu_{h}^{\min }=\min _{m \in \mathbf{M}}\left(\mu_{f\left(\mathbf{x}_{m}, \mathbf{Y}_{h}\right)}\right), \sigma_{h}^{\max }=\max _{m \in \mathbf{M}}\left(\sqrt{\sigma_{f\left(\mathbf{x}_{m}, \mathbf{Y}_{h}\right)}^{2}}\right), \sigma_{h}^{\min }=$ $\min _{m \in \mathbf{M}}\left(\sqrt{\sigma_{f\left(\mathbf{x}_{m}, \mathbf{Y}_{h}\right)}^{2}}\right)$ denote maximum loss, minimum loss, maximum loss standard deviation,
and minimum loss standard deviation by method $h$ over all portfolios, respectively, as shown by Table 4.3 .

Table 4.3: Statistics of loss generated by sequencing methods over all portfolios

| Method $(h)$ | Loss Average |  |  |  |  | Loss Standard Deviation |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\bar{\mu}_{h}$ | $\mu_{h}^{\max }$ | $\mu_{h}^{\min }$ | $R_{h}^{[\mu]}$ | $\bar{\sigma}_{h}$ | $\sigma_{h}^{\max }$ | $\sigma_{h}^{\min }$ | $\sigma_{h}^{\max }-\sigma_{h}^{\max }$ |  |
| $Z(0.0)$ | 0.2537 | 0.5075 | 0.0000 | 0.5075 | 0.1480 | 0.0012 | 0.0000 | 0.0012 |  |
| $Z(0.1)$ | 0.2398 | 0.4779 | 0.0018 | 0.4761 | 0.1388 | 0.0019 | 0.0000 | 0.0019 |  |
| $Z(0.2)$ | 0.2169 | 0.4220 | 0.0118 | 0.4102 | 0.1196 | 0.0052 | 0.0001 | 0.0051 |  |
| $Z(0.3)$ | 0.1999 | 0.3707 | 0.0291 | 0.3415 | 0.0997 | 0.0078 | 0.0003 | 0.0075 |  |
| $Z(0.4)$ | 0.1839 | 0.3016 | 0.0663 | 0.2353 | 0.0688 | 0.0104 | 0.0007 | 0.0096 |  |
| $Z(0.5)$ | $\mathbf{0 . 1 7 6 7}$ | 0.2150 | 0.1384 | $\mathbf{0 . 0 7 6 7}$ | $\mathbf{0 . 0 2 3 2}$ | 0.0120 | 0.0018 | 0.0102 |  |
| $Z(0.6)$ | 0.1874 | 0.2470 | 0.1279 | 0.1191 | 0.0353 | 0.0108 | 0.0036 | 0.0072 |  |
| $Z(0.7)$ | 0.2261 | 0.4138 | 0.0384 | 0.3753 | 0.1097 | 0.0129 | 0.0051 | 0.0078 |  |
| $Z(0.8)$ | 0.2600 | 0.5200 | 0.0000 | 0.5200 | 0.1521 | 0.0203 | 0.0000 | 0.0203 |  |
| $Z(0.9)$ | 0.2600 | 0.5200 | 0.0000 | 0.5200 | 0.1521 | 0.0203 | 0.0000 | 0.0203 |  |
| $Z(1.0)$ | 0.2600 | 0.5200 | 0.0000 | 0.5200 | 0.1521 | 0.0203 | 0.0000 | 0.0203 |  |

It is observed that $Z(0.5)$ has the minimum average loss of $\bar{\mu}_{Z(0.5)}=0.1767$ among all sequencing methods, it also has the minimum standard deviations of $\bar{\sigma}_{Z(0.5)}=0.0232$. We use $R_{h}^{[\mu]}=\mu_{h}^{\max }-\mu_{h}^{\min }$ to assess the range of loss generated by method $h$ over all portfolios. $Z(0.5)$ exhibits a uniform performance over all portfolios with $R_{Z(0.5)}^{[\mu]}=0.0767$ which implies that $Z(0.5)$ is fairly insensitive to the decision maker's portfolio of preferences. Figure 4.7 shows the Pareto frontier of loss average and standard deviation for the trade-off balancing functions. $Z(0.5)$ dominates all other sequencing methods in terms loss average and standard deviation. Single-attribute models of $Z(1.0)$ and $Z(0.0)$ show the two worst performances, respectively.


Figure 4.7: Trade-off balancing functions Loss average and Standard deviations over all portfolios

Balancing trade-offs between flow time mean and flow time variance provides a powerful yet flexible tool to shift the performance of the single-machine system towards a desired state where the outcomes are more predictable and also close to the infeasible ideal point. By translating this conclusion into the operating room scheduling, we can conclude that balancing trade-offs between patient flow time mean and patient flow time variance provides a smooth
patient flow across the OR peri-op process. Smoothing patient flow results in reduced waiting times and overcrowdings that consequently improves the quality of care. Balancing tradeoffs enables the OR managers to efficiently use resources by reflecting different stakeholders interests into the objective function.

### 4.7 Conclusion

Flow time mean and variance are two important measures in any production system, specifically, in operating room (OR) scheduling, mean patient flow time (PtF) and patient flow time variance ( $P t F V$ ) are two indicators for service throughput and uniformity, respectively. Minimizing patient flow time i.e. $\min (P t F)$ reduces waiting times and overcrowdings between upstream resources and OR but it may negatively affect the downstream resources. Minimizing patient flow variance i.e. $\min (P t F V)$ provides the patients with an approximately the same waiting time that could improve patient's satisfaction but at cost of increased waiting time. However, we showed that the objectives of $\min (P t F)$ and $\min (P t F V)$ are inconsistent objectives and optimizing one is on the cost of the other. By modeling OR as a single-machine production system, we utilized $y_{k}$ the normalized deviations from the best value of attribute $k$ to formulate the objective function in the form of $Z(\alpha)=\alpha y_{1}+(1-\alpha) y_{2}$ to balance the trade-offs between $\min (P t F)$ and $\min (P t F V)$. We also proposed a fast heuristic to find the optimal sequence to $\min Z(\alpha)$ in $O(n \log n)$. Through extensive case studies on the University of Kentucky Healthcare (UKHC) OR historical data, we showed that balancing trade-offs outperforms single-attribute models of $\min (P t F)$ and $\min (P t F V)$, and the bi-criteria models in the form of $Z_{B C}(\alpha)=\alpha P t F+(1-\alpha) P t F V$ as well.

The results showed that trade-off balancing models are able of reflecting the the stakeholders interests and the decision maker's preferences while shifting the performance of the system towards a predictable state with minimum deviation from an 'ideal point' at which all attributes are at their optima. Balancing trade-offs between $\min (P t F)$ and $\min (P t F V)$ can smooth the patient flow across the OR peri-operative process resulting in an improved quality of care.

## Chapter 5

## An Optimization Model for Operating Room Scheduling to Reduce Blocking Across the Perioperative Process

Operating room scheduling is important. Because of increasing demand for surgical services, hospitals must provide high quality care more efficiently with limited resources. When constructing the OR schedule, it is necessary to consider the availability of downstream resources, such as ICU and PACU. The unavailability of downstream resources causes blockings between every two consecutive stages. In this chapter, we address the master surgical schedule (MSS) problem in order to minimize blockings between two consecutive stages. First, we present a blocking minimization ( $B M$ ) model to construct the MSS using linear integer programming, based on deterministic data. The $B M$ model determines the OR block schedule for the next day by considering the current occupancy (number of patients) of the downstream resources with the objective of minimizing the number of blockings between intra-opertaive and postoperative stages. Second, we test the effectiveness of our model under variations in case times and patient arrivals, using discrete event simulation. The simulation results show that the $B M$ model can significantly reduce the number of blockings by $94 \%$ improvement over the hospital base model. Scheduling patient flow across the 3 -stage peri-operative process can be applied to work flow scheduling for the $s$-stage flow shop production in manufacturing, and also smoothing patient flow in peri-operative processes can be applied to no-wait flow shop production. The results of this chapter appears in [96].

### 5.1 Introduction

In most hospitals when OR blocks are assigned to surgery groups, there is no specific mechanism to ensure the availability of downstream resources, such as the beds in ICU or PACU [97]. Because of the unavailability of downstream resources, patients cannot be sent to the next stage, but are held in the current stage, causing blockings between every two consecutive stages. Blockings negatively impact OR performance across the peri-op process, such as increased waiting time in each stage, increased length of stay (LoS) across the peri-op process, excessive overtime and overnight shifts, etc. In this chapter, we develop a blocking minimization $(B M)$ model to reduce the number of blockings between two consecutive stages. To avoid blocking, the number of patients in each stage should not exceed the number of beds in that stage. The number of patients in each stage is affected by three events:

- Patients arriving on each day from the upstream stage.
- Patients from previous days who still need to stay in the current stage.
- Patients from previous days who have spent enough time in the current stage and are ready to leave.

Our $B M$ model provides an optimal OR block schedule by taking these three events into consideration. The $B M$ model determines the OR block schedule for the next day by considering the current stage occupancy (number of patients in the stage) in order to minimize the number of blockings between the intra-op and the post-op stages. Therefore, the $B M$ model balances the admission and departure of the post-op stage, while avoiding the accumulated number of patients exceeding the number of resources. Using discrete event simulation, we show that our $B M$ model effectively dampens the variations in the case times and patient arrivals.

Surgeries are categorized into two major classes: elective cases, and emergency cases [6, 49. Elective cases are scheduled several days prior to the intervention date, but emergency cases should be scheduled as soon as possible [50]. OR scheduling consists of three major phases, which challenges the OR block scheduling. Three major phases in OR scheduling are described as follows:

- Strategic: The main objective of the strategic phase is to provide a 'case mix plan', which allocates OR blocks to surgery groups. The strategic phase is based on historical data and/or forecasts, and typically has a time horizon of one year 49, 24, 98, 61, 60].
- Tactical: The main objective of the tactical phase is to provide a master surgical schedule. The MSS determines the number, type and opening hours of ORs for each surgery group. In the MSS, surgery types are clustered to surgery groups based on similar characteristics of specialties and requirement on resources, such as facilities in ORs, ICUs and PACUs. The time horizon of tactical phase usually is one to three months. The MSS is mainly based on elective cases, the number and case time of which do not vary remarkably in three months. Therefore, the MSS is cyclic and repeated in the tactical phase. Hospital administrators prefer to assign OR blocks to surgery groups instead of individual surgery types. This allows them to swap OR blocks among surgery types within the same group in case of necessity. These small swaps will not change the optimal schedule, thus, they do not have to develop a new schedule for any small change. Figure 5.1 shows an example of MSS, where the number of available OR blocks (an OR block means daily working hours of an OR or an OR-day) is 6 on each day, there are three surgery groups G1, G2, and G3. The MSS should be revised whenever the total amount of available OR blocks changes [49].

|  | Mon | Tue | Wed | Thu | Fri |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 2 | G1 | G2 | G3 | G1 | G3 |
| 4 |  |  | G1 | G2 |  |
| 5 | G2 |  | G2 |  | G1 |
| 6 |  | G3 |  |  | G2 |

Figure 5.1: An example of MSS

- Operational: After development of the MSS, the assignment of cases to ORs and start/end time of each case are determined on a daily basis.

High OR utilization is one of the main objectives of OR scheduling and allocates OR blocks to surgery groups. However, high variations in case times and patient arrivals may fail an OR schedule with high OR utilization at the tactical phase [97, 6, 99, 100, 101. According to the literature, the OR utilization should be maximized to avoid under-utilization costs, however, due to the high variations in case times and patients arrivals, highly utilized ORs are unstable [6, 45, 102]. Highly utilized ORs cannot dampen the variations because there is no time buffer, thus, a slight variation in case or arrival time may cause high overtime cost or surgery cancellation [6]. It is also important to develop an OR schedule that leads to a leveled occupancy of downstream stages 6].

OR utilization has been intensively studied [103, 68, 51, 104, 105, 106, 67], but few papers addressed the patient flow across the peri-op process. OR schedule directly affects PACUs and ICUs in the post-op stage [97, 107]. After performing a surgery in an OR, the patient is recovered in PACU, then moved to an ICU bed to receive the required care. The amount of time that a patient stays in ICU before being moved to the ward is referred as length of stay $(\mathrm{LoS})$ in ICU. After spending the required time in ICU, the patient is moved to a NonICU bed in the ward. Sometimes based on the acuity level, the patient is directly moved from PACU to the ward. For those patients who do not need the ICU care, we can consider their ICU LoS as zero.

Figure 5.2 shows the patients' path in the peri-operative process. If there is no available


Figure 5.2: The patients' path in the peri-operative process
bed in PACU, patients have to stay in the OR [107], which is considered as blocking. OR blocking means no surgery is performed until a bed in PACU or ICU is available. OR blocking decreases the OR utilization that leads to the waste of costly OR time. An efficient OR schedule not only maximizes the OR utilization, but also smooths the patient flow across the peri-op process, i.e., reducing the number of blockings.

Price et al. (2011) 107 proposed a deterministic PACU Boarding model (hereafter we refer to this model as $P B$ ) to develop an optimal MSS that balances the admission and discharge rate of ICU. The $P B$ model used deterministic data for patient arrivals and ICU LoS. They used simulation to evaluate $P B$ model under the presence of variations. According to their results, the $P B$ can reduce the number of blockings between OR and ICU compared to the base model of the studied hospital. Their focus was on minimizing the discrepancy between admission and discharge rate of ICU. Because in their model the current number of patients in the ICU was neglected, the impact of patient accumulation in ICU has not been taken into consideration. Using simulation, Marcon and Dexter (2006) 35 examined the effect of different sequencing rules in ORs, on the PACU workload. The results showed that the case sequencing in ORs has a minor effect on the PACU workload. They stated that the best practice is to use optimization in order to match the PACU workload and its capacity.

The remainder of this chapter is organized as follows: Section 5.2 describes the problem settings and mathematical formulation of the $B M$ model. Section 5.3 presents the details of case studies for evaluating the effectiveness of OR block schedules. Section 5.4 draws conclusion and proposes future work.

### 5.2 Problem Formulation

We are interested in the performance of $B M$ model in terms of weekly number of blockings and daily post-op occupancy. Our $B M$ model recursively considers the number of patients in the post-op stage in order to minimize the likelihood of blocking between ORs in the intraop stage and the post-op stage. $B M$ model takes the current post-op stage occupancy into consideration to determine the OR assignment for the next day such that the accumulated number of patients in post-op does not exceed the number of available beds. Using discrete event simulation, we show that our $B M$ model outperforms the model proposed by Price et al.(2011) [107]. For the sake of simplicity, hereafter we refer to the Price et al.(2011) model as $P B$ model.

## Notations

$g \quad$ Index of surgery groups, $g \in G=\{1,2, \ldots, G\}$.
$j, k \quad$ Index of days, $j \in D=\{1,2, \ldots, D\}$.
$b_{g j} \quad$ The number of OR blocks assigned to surgery group $g$ on day $j$.
$n_{g} \quad$ The total number of OR blocks that group $g$ requires.
$O$ The number of available ORs in the intra-op stage.
$\lambda_{g} \quad$ The expected number of patients per block for group $g$.
$\mu_{g} \quad L o S$ in the post-op stage for group $g$.
$L_{g} \quad$ The minimum number of OR blocks that should be assigned to group $g$.
$U_{g} \quad$ The maximum number of OR blocks that can be assigned to group $g$.
$B \quad$ The number of available beds in the post-op stage.
$P_{j} \quad$ The number of patients in the post-op stage on day $j$.
$\lambda_{g}$ is the expected number of patients per block for surgery group $g . \lambda_{g}$ is estimated by dividing the length of an OR block by the mean case times plus mean setup time plus the mean cleanup time. The setup time is the time required to adjust equipment before the surgical intervention, and the cleanup time is the required time to clean and disinfect the equipment after completion of a surgery. The required maximum and minimum of OR blocks for each surgery group (i.e. $L_{g}$ and $U_{g}$ are determined from the historical data.

The $B M$ model is formulated as follow:

$$
\begin{gather*}
\min \quad z=\sum_{j \in D}\left|P_{j}-B\right|  \tag{5.1}\\
\text { s.t. } \\
\sum_{j \in D} b_{g, j}=n_{g}  \tag{5.2}\\
\sum_{g \in G} b_{g, j} \leq 0  \tag{5.3}\\
P_{j}=\sum_{g \in G} \lambda_{g} b_{g, j} X_{g, j}+\sum_{k} \sum_{g \in G} \lambda_{g} b_{g, j} Y_{g, j}  \tag{5.4}\\
X_{g, j}= \begin{cases}1, & \text { if } \mu_{g}>0 \\
0, & \text { otherwise }\end{cases}  \tag{5.5}\\
Y_{g, j}=\left\{\begin{array}{cc}
1, & \text { if } \mu_{g}-(j-k)>0 \\
0, & \text { otherwise } \\
L_{g} \leq b_{g, j} \leq U_{g}, \forall g, j \\
b_{g, j} \geq 0, b_{g, j} \in \mathbb{Z}
\end{array}, \forall g, k<j\right. \tag{5.6}
\end{gather*}
$$

Objective function(5.1) minimizes the difference between the post-op stage capacity and its occupancy (number of patients in the ICU). This function not only minimizes the overoccupancy in the post-op stage, but also balances the daily post-op occupancy. Constrain$\mathrm{t}(5.2)$ imposes that the number of assigned OR blocks is equal to each group required OR blocks. Constraint 5.3 imposes that the number of assigned OR blocks does not exceed the number of available ORs. Constraint (5.4) recursively determines the occupancy of the postop stage on day $j$. The term $\sum_{g \in G} \lambda_{g} b_{g, j} X_{g, j}$ determines the number of arrivals from OR to the post-op stage on day $j$. The $B M$ model distinguishes between different LoSs, i.e. $X_{g j}=1$ if the $\mu_{g}>0$, which means that the patient needs to be sent to the post-op stage, otherwise he/she is sent directly to the ward (i.e. if $\mu_{g}=0$ ).

The term $\sum_{k<j} \sum_{g \in G} \lambda_{g} b_{g, j} Y_{g, j}$ determines the number of patients in the post-op stage transferred to day $j$ from the previous days, $Y_{g, k}=1$ if the patient still needs to stay in the post-op stage. Constraint 5.5 describes $X_{g, j}$, as a binary decision variable, which is equal to 1 , if $\mu_{g}>0$. A patient is sent to the post-op stage if his/her post-op LoS is greater than zero, otherwise he/she will be directly sent to a bed in the ward. Constraint 5.6) describes $Y_{g, k}$, as a binary decision variable, which is equal to 1 , if $\mu_{g}-(j-k)>0$. If a patient in the post-op stage still needs to spend more time in the post-op, $Y_{g, k}$ is equal to 1 , otherwise he must be sent to a bed in the ward (discharge from post-op stage). Constraint 5.7 ) imposes that the daily number of OR blocks assigned to each group must fall between specified minimum and maximum values. Constraint 5.8) imposes that the number of OR blocks assigned to each group must be a positive integer. The $B M$ model uses the deterministic average LoS of the post-op stage to determine the MSS. The $B M$ model explicitly deals with the post-op occupancy by considering arrivals, discharges, and current occupancy of the post-op stage. Here an assumption is that the ORs and post-op resources are universal, which means they can be assigned to any surgery group.

### 5.3 Case Study

To validate the performance of the $B M$ model for OR block scheduling, we carry out three types of case studies. First, we solve the $B M$ model to obtain the optimal MSS. Second, we construct a discrete event simulation model to assess the robustness of $M S S$ (s) generated by $B M, P B$, and the current $M S S$ implementing in the studied hospital by Price et al.(2011) [107]. For the sake of simplicity, we name the hospital MSS in Price et al.(2011) as 'Base model'. We introduce variations in $\mu_{g}$ and $\lambda_{g}$ in the simulation model, in order to observe the number of blockings and daily occupancy of the post-op stage generated by the three models. Third, we evaluate the process capabilities for the three models, using statistical process control (SPC).

## Optimal MSS

We use the data provided by Price et al.(2011) [107] to construct the simulation model. There are 16 ORs, 31 beds in the ICU, and 3 surgery groups. Patients are clustered into groups based on similarity in their case times and LoS. Table 5.1 shows the historical data of surgery groups. Patients in group 1 have short LoS and case times. Patients of group 2 have relatively longer LoS than group 1, and the case time is also longer. Finally, group 3 has the longest case time and also the longest LoS. Studies show that the prolonged case times in OR are associated with increased LoS in the post-op stages [108]. The required, maximum and minimum OR blocks for each surgery group are determined from the historical data.

Table 5.1: Historical data of the surgery groups

| Surgery <br> group | Minimum <br> blocks | Maximum <br> blocks | Required <br> blocks | ICU <br> LoS | Total <br> LoS | Patient <br> per block |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 1.0 | 5.0 | 11.1 | 0 | 1 | 2.5 |
| 2 | 2.0 | 13.0 | 30.6 | 1 | 2 | 1.5 |
| 3 | 3.0 | 15.0 | 33.2 | 2 | 3 | 1.25 |

Table 5.2 shows the optimal $M S S$ generated by the $B M$ model. Figure 5.3 shows the graphical representation of the OR block schedule, each entry is the number of OR blocks assigned to each surgery group.. It is possible to assign an OR block to a surgery group only for the morning or the afternoon, which is called 'half-block'. The MSS is constructed for a 5 -week horizon, therefore the base unit of OR block is a tenth of a block over one week [107]. Since the base unit of OR block is 0.1 , the $b_{g j}$ is scaled by a factor of 10 to maintain the integer nature of the problem, for example, the value 21 obtained from the $B M$ model is equal to 2.1 blocks in the MSS. As

Table 5.2: The optimal MSS obtained from the $B M$ model

| Surgery group | Mon | Tue | Wed | Thu | Fri | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 2.1 | 1.0 | 3.0 | 2.0 | 3.0 | 11.1 |
| 2 | 5.9 | 5.8 | 7.1 | 6.6 | 5.2 | 30.6 |
| 3 | 7.9 | 8.2 | 5.9 | 6.2 | 3.0 | 31.2 |
| Total | 15.9 | 15.0 | 16.0 | 14.8 | 11.2 | $\mathbf{7 2 . 9}$ |



Figure 5.3: A graphical representation of the $B M$ model optimal $M S S$

## Simulation

The $B M$ model uses deterministic values for $\mu_{g}$ and $\lambda_{g}$, but in practice these parameters are random variables following known distributions. We construct a simulation model to assess the robustness of the $B M$ model under the presence of variations in ICU LoS and number of patients per block. Variation in number of patients per block is associated with variations in the case times and also the arrival of emergency cases. We compare the robustness of the $B M$ model with the $P B$ model, and the Base model. A robust $M S S$ can absorb the variations, and leads to a fewer blockings.

First, we simulate the model for a long period of time to observe and collect the number of weekly blockings and daily occupancy of ICU. Second, we compare the performance of the $B M$ model $M S S$ with that of the $P B$ model, and the Base model in Price et al.(2011). In each OR block, a random number of cases was generated. In order to involve the disturbances induced by variations in case times and emergency arrivals, we considered $550 \%$ variation in the number of patients per block (i.e. $\lambda_{g}$ ). For example, for group 2, we have $1.5[0.50,1.50]=$ $[0.75,2.75] \approx[1,3]$.

The ICU LoS for each case was randomly generated from a uniform distribution between $1 / 6$ and $1 / 2$ of the total LoS (see table 5.1). All random numbers were rounded to the nearest greater integer. The simulation was warmed up for 20 weeks in order to reach a steady state behavior; the model was then run for 52 weeks (one year) and results were averaged.

Table 5.3 represents the average and standard deviation of daily occupancy of the post-op stage. Figure 5.4 shows the average ICU occupancy generated by the three models. $B M$ model has a maximum average occupancy of 25.30 which is far from the maximum capacity of ICU ( 31 beds), therefore, $B M$ model can dampen the variations in number of patients and ICU LoS better than the $P B$ and the Base models with average maximum occupancy of 30.01 and 30.49 , respectively. The maximum ICU occupancy of $P B$ and Base models is 30.1 and 30.49 , respectively, which occur on Tuesday. Their maximum ICU occupancy is remarkably close to the maximum capacity ( 31 beds), therefore a slight variation causes blockings in the peri-op process. Although the $P B$ model has a lower occupancy on Wednesday (21.64) and Thursday (18.72) compared to those of the $B M$ model ( 24.41 and 22.68 , respectively), on Tuesday and Thursday its ICU occupancy is close to the maximum capacity (30.49 and 29.92 ), respectively. This makes the $P B$ model generate more blockings on Tuesday and Thursday, under the presence of variations. Besides, average ICU occupancy on Saturday is only 2.03 in the $B M$ model, which is desirable for OR managers because less weekend shifts will be required. Figure 5.5 shows the individual plot of blockings generated by the three models. Since the $B M$ model takes the current occupancy of the ICUs into consideration it
exhibits more slack capacity by efficiently assigning the surgery cases to ORs. The $B M$ model could generate fewer blockings compared to the $P B$ and the Base models. Therefore, we can conclude that the $B M$ model is more robust under the presence of variations in patients' arrivald and ICU LoS.

Table 5.3: Daily occupancy of the post-op stage

| Model | Mon |  | Tue |  | Wed |  | Thu |  | Fri |  | Sat |  | Sun |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ |  |

$\mu$ : Average, $\sigma$ : Standard Deviation


Figure 5.4: The average daily ICU occupancy


Figure 5.5: Dotplot of blockings for the studied models

In these models the total number of required OR blocks is fixed and is equal to 72.9 (obtained from the historical data), but the number of available OR blocks is 80 ( 5 day and 16 OR blocks on each). Using the $B M$ model, we can provide post-op resources for more OR blocks, because $B M$ evenly distributes the post-op workload over the weekdays. Therefore, we can schedule for more patients using the unutilized OR blocks ( $80-72.9=7.1$ OR blocks).

## Statistical Process Control (SPC)

To compare the robustness and stability of the block scheduling models, we use SPC method to generate the process capability results for the discussed models. Figure 5.6 shows the process capability diagrams.

(c) Base model

Figure 5.6: Capabilities of the studied model for the number of blockings

Process capability $c_{p}$ indicates if the outcomes of a process are within the control limits. With the fixed range of specification limits, which is $[L S L, U S L]$, the greater the $c_{p}$, the less variations in the outcomes. Process capability index $c_{p k}$ indicates if the outcomes are centred around the average performance. The greater the $c_{p k}$, the less likely that the outcomes fall out of $[L S L, U S L]$ range. As shown in Figure $5.6, c_{p}=1.66$ for $B M, c_{p}=0.57$ for $P B$, and $c_{p}=0.34$ for the Base model. Therefore, the $B M$ model generates the least variations in the outcomes (number of blockings). $c_{p k}=1.56$ for $B M, c_{p k}=0.33$ for $P B$, and $c_{p k}=-0.01$ for the Base model. Therefore, the blockings generated by the $B M$ model are the most centred within limits. Obviously, the average blockings generated by $B M$ is more centred within the specification limits and with less variation, compared with those of the $P B$ and the Base model. Table 5.3 shows the results of SPC and process capability indices. SPC $\bar{x}-R$ charts of average weekly blockings are shown shown by Figure 5.7.

Table 5.4: SPC and capability results of weekly number of blockings

| Model | capability <br> index |  | Number of <br> Blockings |  |  | Improvement <br> vs |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $c_{p}$ | $c_{p k}$ | Max | Improvement <br> vs |  |  |  |
| $B M$ | 1.66 | 1.56 | 8 | Min | Base model | $P B$ |  |
| $P B$ | 0.54 | 0.33 | 12 | 7.90 | 0.30 | $94 \%$ | $85 \%$ |
| Base | 0.34 | -0.01 | 31 | 13.36 | 5.12 | - | - |

We simulated the models over a 52 weeks time period. As shown by Figure 5.7 and Table 5.3, the average weekly number of blockings are 0.30 for $B M, 2.07$ for $P B$, and 5.12 for the Base model. Besides, there are occasions for the $B M$ model generating almost zero


Figure 5.7: SPC charts of average weekly number of blockings
blockings. Compared with the Base model, the improvement on average number of blockings can be calculated as $(5.12-0.30) / 5.12100=94 \%$. Moreover, compared with the $P B$ model, the improvement of $(2.07-0.30) / 2.07100=85 \%$ can be achieved. The average range of variation of the number of blockings for the $B M$ model is 2.71 , less than 7.9 and 13.36 those of $P B$ and the Base model, respectively. Therefore, the $B M$ model generates the least average variations in the number of blockings. The results from $\bar{x}-R$ charts support those of $c_{p}$ and $c_{p k} .94 \%$ improvement on the number of blockings indicates that potentially on average 250 additional cases could be served in one year, if the $B M$ model is used for the OR block scheduling. There would be an increase in the hospital net revenue from the additional cases. An estimate of this value can be calculated by multiplying the number of additional cases by the net revenue per case, therefore, there would be an annual increase of $\$ 250 \times 15000=3.75$ million dollars in the hospital revenue. However, in practice, the availability of ORs, surgeons and staff are other factors that may limit the estimated increase in the hospital revenue.

### 5.4 Conclusion

Operating room (OR) scheduling is important, because ORs have the largest cost and revenue within a hospital, and the demand for surgical services is increasing. Therefore, hospitals must provide high quality care more effectively with limited resources by developing efficient OR schedules. In most hospitals, when OR blocks are assigned to surgery groups, there is no specific mechanism to ensure the availability of downstream resources such as the beds in ICU and PACU. Because of the unavailability of downstream resources, patients cannot be sent from ORs to ICU or PACU, causing blockings between every two consecutive stages.

This leads to numerous negative impacts on the OR peri-operative process performance, such as increased waiting times, length of stays (LoS), excessive overtime, and overnight shifts, etc. Therefore, when the MSS is constructed, it is necessary to consider the availability of downstream resources.

In this chapter, we developed a model called $B M$ to reduce the number of blockings between two consecutive stages. To minimize blocking, the arrivals, departures, and current occupancy of the post-op stage were taken into account. Our objective is to assign OR blocks to surgery groups such that the post-op occupancy does not exceed the number of available beds in the post-op. Simulation results showed that the $B M$ model outperformed the $P B$ model proposed by Price et al. (2011), and the studied hospital Base model as well. Moreover, using $B M$ model, an improvement of $94 \%$ in reducing the number of blockings (over the Base model) can be achieved, meaning that by using $B M$ model, potentially we can serve more patients. The SPC results showed that the $B M$ model can dampen the variations in the case times and ICU LoSs. Our work showed that considering downstream resources in OR department is important and the proposed $B M$ model can effectively improve the overall performance of the OR department. The $B M$ model can be generalized to any two consecutive stages across the peri-op process.

The managerial implications of this work is that the interaction of different stages should be taken into consideration. While OR block scheduling, the intake and discharge flow of the stages should be considered in order to guarantee that upstream stages can feed downstream stages on time without overflowing, because overflowing generates blockings between stages, that lowers the utilization of the whole peri-op process and slows down the patient flow across the process. Scheduling patient flow across the 3 -stage peri-op process can be applied to work flow scheduling for the s-stage flow shop production in manufacturing, and also smoothing patient flow in the peri-op process can be applied to no-wait flow shop production.

## Chapter 6

## Risk Management in Surgical Case Sequencing by Balancing Trade-offs Among Inconsistent Objectives

The aim of this chapter is to mark the boundaries of risk management in an operating room peri-operative process. To this aim, we elaborate risk definition, risk sources, and propose risk management strategies. We identify uncertain demand, uncertain surgical case time, and the inconsistencies among objectives as risk sources. We adopt the risk definition as "the variations in the distributions of OR peri-op process outcomes, their likelihood, and their subjective values". Finally, we propose balancing trade-offs among inconsistent objectives as a risk management strategy. OR peri-op process stages are tightly coupled, therefore, risk sources affect the value flow and patient flow across the OR peri-op process borders. We identify process utilization (Util) and patient length of stay ( $L o S$ ) as two major key performance indicators (KPIs) driving value flow and patient flow, respectively. By demonstrating inconsistency between $\max (U t i l)$ and $\min (L o S)$, we show that single attribute optimizations are not suitable to manage risks. Through extensive case studies on OR historical data-set from the University of Kentucky HealthCare (UKHC), we demonstrate balancing trade-offs between Util and $L o S$ as a risk mitigating strategy.

### 6.1 Introduction

The implementation of the mandatory alternative payment model (APM) guarantees savings for Medicare regardless of participant hospitals ability for reducing spending [109] that shifts the cost minimization burden from insurers onto the hospital administrators 81. Surgical interventions are the primary cause of nearly $70 \%$ of hospitals admissions [4] that account for more than $30 \%$ and $40 \%$ of a hospital total cost and total revenue, respectively [110]. Operating rooms are also the most expensive cost center in a hospital with an estimated cost of $\$ 36$ to $\$ 37$ per minute in 2014 81. Childers and Maggard-Gibbons (2018) point out that the OR cost structure consists of nearly $56 \%$ direct cost, with about $60 \%$ being attributed to labor [81]. Therefore, large cost reductions are possible by reducing overtime hours [111] which may be achieved by increasing utilization. Thus, operating room utilization (Util) is of importance in reducing costs and improving hospital financial solvency [112, 111].

An operating room peri-operative process consists of three main stages including preoperative, intra-operative, and post-operative where patient preparation is performed in the pre-op stage, anesthesia and the surgical intervention happens in the intra-op stage, and post-anesthesia care occurs in the post-up stage. Any disturbance in the patient flow across the OR peri-op process may result in an extended $L o S$ that consequently increases the cost
of care. Stey et al. (2015) point out that room and board costs accounted for nearly half of all costs and were highly correlated with length of stay ( $\operatorname{LoS}$ ) [112]. Therefore, studying the ORs as an isolated unit may be inefficient, instead modeling surgical interventions as a peri-operative process has the potential to simultaneously address the ORs and their upstream and downstream resources [113]. Therefore, Util and LoS are two important key performance indicators (KPIs) which must be addressed simultaneously. However, studies show that $\max (U t i l)$ and $\min (L o S)$ are two inconsistent objectives meaning improving one may occur at the cost of worsening the other [114].

From a queueing point of view, the expected waiting time between each pair of sequential stages is formulated by Equation (6.1) which is known as the Kingman's formula [115], where $\mathbb{E}(W)$ is the expected waiting time, $U$ is the utilization of the stage, $C_{a}$ and $C_{p}$ are the coefficient of variation of arrivals to the stage and service times at the stage, respectively, and $\mu_{p}$ is the mean service time at the stage.

$$
\begin{equation*}
\mathbb{E}(W) \approx\left(\frac{U}{1-U}\right)\left(\frac{C_{a}^{2}+C_{p}^{2}}{2}\right) \mu_{p} \tag{6.1}
\end{equation*}
$$

A clear conclusion from the VUT model is the inconsistency between waiting time and the utilization, because if utilization of the stage tends to $100 \%$ (i.e. $U=1$ ) the waiting time will tend to infinity. The second conclusion from the VUT model is that the variations in arrivals and service times at the stage directly affect the distributions of the process outcomes. $L o S$ is the sum of waiting times and service times across the OR peri-op process, therefore, we can conclude that the objectives of $\min (L o S)$ and $\max (U t i l)$ are inconsistent objectives.

As it is schematically shown by Figure 6.1. OR peri-op process stages are tightly coupled, therefore, any disturbance in any stage may result in blocking or starvation at other stages that negatively affect Util and $L o S$ across the entire peri-op process. Disturbances across the peri-op process result in uncertainties in the peri-op process outcomes. Any process with uncertainties in its outcomes carries some elements of risk [116], and internal or external variables that reduce the predictability of the outcomes are considered as risk sources [117].


Figure 6.1: Schematic representation of an OR peri-op process

In an OR peri-op process risk can be defined as "the variations in the distributions of $O R$ peri-op process outcomes, their likelihood, and their subjective values", therefore, we can consider the variations in arrivals and service times, and the inconsistencies among the objectives as risk sources. In response to the complex nature of the OR peri-op process and its risks, risk management would be a value-added function. Risk management in the OR peri-op process
not only mitigates the negative impacts of risks, but also provides insights into the OR cost structure which are of importance in capacity planning and futuristic developments.

Risk taking is inherent to decision making [118] which means any decision made under uncertainty carries some elements of risks. The common practice in OR scheduling is to optimize the expected value $(\mathbb{E}(\cdot))$ of a loss (reward) function [119, 63, 4], however, two fundamental shortcomings are associated with optimizing $\mathbb{E}(\cdot)$ as follows (i) optimizing $\mathbb{E}(\cdot)$ is considered as a risk-neutral approach which is not capable of addressing the concerns of risk-averse/risk-seeking decision makers [120], (ii) optimizing $\mathbb{E}(\cdot)$ equivalently treats two outcome distributions with the same $\mathbb{E}(\cdot)$ and completely disregards the variablity/symmetry of the distributions that may result in poor realization. An approach to deal with these shortcomings is minimizing the variance of the outcomes, however, minimizing variance also has several mathematical and statistical drawbacks including (i) quadratic nature of variance that leads to nonlinear optimization models which are computationally difficult to solve for large instances, and (ii) minimizing variance performs poorly when the underlying random parameters are non-symmetric [121, 122 . Therefore, managing the risk would be an efficient alternative approach in order to shift the outcomes of the OR peri-op process toward a less risky situation.

The first step in risk management is defining the risk measures. A risk measure maps a set of random variables to a real number [123]. Multiple risk measures have been introduced over the past decades [124] among which value at risk ( $V a R$ ) [125] and conditional value at risk $(C V a R)$ [126, 122] have attracted great attention.

Hereafter, in order to distinguish between vectors and scalars, we use bold font for the vectors. Given $K$ KPIs, let $\mathbf{x} \in \mathbb{R}^{K}$ be the decision vector representing a portfolio of preferences for for $K$ KPIs where $x_{k} \geq 0$, and $\sum_{k=1}^{K} x_{k}=1$. Let $\mathbf{y} \in \mathbb{R}^{m}$ denote a random vector representing the underlying random variables (uncertainties) in the systems with density function $p(\mathbf{y})$. The loss function $f(\mathbf{x}, \mathbf{y})$ is also a random variable with a distribution in $\mathbb{R}$ induced by $\mathbf{y}$. The probability that loss does not exceed a threshold $\xi$ is $\Psi(\mathbf{x}, \xi)=\int_{f(\mathbf{x}, \mathbf{y}) \leq \xi} p(\mathbf{y}) d \mathbf{y}$. For a fixed $\mathbf{x}, \Psi(\mathbf{x}, \xi)$ is a function of $\xi$ representing the cumulative function of the loss associated with $\mathbf{x}[122,127]$. The assumption is that $\Psi(\mathbf{x}, \xi)$ is everywhere continuous with respect to $\xi$ 122.

Given a specified probability level $\alpha \in(0,1)$ (also known as confidence level), value at risk $(\alpha-V a R)$ is described by $\xi_{\alpha}(\mathbf{x})=\min \{\xi \in \mathbb{R} ; \quad \Psi(\mathbf{x}, \xi) \geq \alpha\}$ which is the lowest value of $\xi$ that with probability $\alpha$ the loss does not exceed $\xi$ [128. Despite the popularity of $\alpha$-VaR, it has multiple shortcomings that makes it difficult to implement in practice. For instance $\alpha-V a R$ is not coherent for non-symmetric distributions, and $\alpha-V a R$ optimization is very difficult for scenario-based problems [123], $\alpha-V a R$ also lacks convexity. On the other hand, conditional value at risk $(\alpha-C V a R)$ defined by $\phi_{\alpha}(\mathbf{x})=\frac{1}{1-\alpha} \int_{f(\mathbf{x}, \mathbf{y}) \geq \xi_{\alpha}(\mathbf{x})} f(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d \mathbf{y}$ is a convex function with monotonic first order and second order stochastic dominance [129]. $\alpha-C V a R$ is the expected value of extreme losses in the tail of the loss distribution if the worst case threshold $\xi$ is crossed. Studies show that portfolios with low $\alpha-C V a R$ also have low $\alpha-V a R$, therefore, in the present study we accept $\alpha-C V a R$ as the risk measure.

Surgical case sequencing can affect the distribution of the OR peri-op process outcomes [36], therefore, efficient sequencing methods can shift the distribution of the outcomes towards a less risky state in which the outcomes are more predictable. With inconsistencies between the objectives of $\max (U t i l)$ and $\min (L o S)$, the probability of finding a sequence that simultaneously optimize both objectives is zero, therefore, seeking optimal solutions is a risky approach with unpredictable outcomes for either objectives. Instead, we propose balancing the trade-offs between $\max (U t i l)$ and $\min (L o S)$ can serve as a risk mitigating strategy in the

OR peri-op process. We use the amount of trade-off generated in the system to define the loss function as $f(\mathbf{x}, U t i l, L o S)$ where $\mathbf{x}$ is the vector of decision maker's preferences portfolio.

We examine the performance of different sequencing methods on the University of Kentucky Healthcare (UKHC) OR historical data in terms of loss expected value $\mathbb{E}(f)$, loss variance $\sigma^{2}(f)$, and loss conditional value at risk $\alpha-C V a R$. To this aim, we model the OR peri-op process as a 3 -stage flow shop scheduling problem. In a flow shop scheduling problem Util and LoS are assessed with maximum completion time ( $M C T$ ) and total completion time ( $T C T$ ), respectively [114, 43].

We compare the performance of three simple rules of first come first serve (FCFS), longest processing time first ( $L P T$ ), shortest processing time first ( $S P T$ ), and three sequencing methods including the NEH [130], LR [131, and current and future deviation (CFD) [114] which are the best known constructive heuristics to minimize $M C T, T C T$, and the trade-offs between $T C T$ and $M C T$, respectively.

The CFD heuristic proposed by Li et al. (2019) 114 is a constructive heuristic to balance the trade-offs between $\min (M C T)$ and $\min (T C T)$. CFD minimizes the weighted of deviations i.e. $\min \left(\beta d_{1}+(1-\beta) d_{2}\right)$ where $\beta \in[0,1]$ is the wight for $\min (M C T)$, and $d_{1}$ and $d_{2}$ are the normalized deviations for $\min (M C T)$ and $\min (T C T)$, respectively. By varying $\beta$ in a stepwise manner i.e. $\beta=0.0,0.1, \ldots, 1.0$, we can have a series of functions to serve different objective, specifically, if $\beta=0.0$, CFD00 is to $\min (T C T)$, if $\beta=1.0$, CFD10 is to $\min (M C T)$, and finally for $0.1 \leq \beta \leq 0.9$, CFD01 to CFD09 are functions minimizing the trade-offs between $\min (M C T)$ and $\min (T C T)$.

Since objective setting is the most important step in an optimization problem, we prefer to demonstrate the significance of trade-off balancing as a risk mitigating strategy in the OR peri-op process and leave addressing the variations in arrivals and service times for a future study.

## Summary

Operating rooms (ORs) are the most important cost and revenue center in a hospital. Studies show that efficient OR scheduling is able to decrease the cost of care and improve the hospital financial solvency. Modeling ORs as an isolated unit may lead to poor performance that increases the cost of care. Therefore, modeling the ORs as a peri-operative (peri-op) process is necessary. Utilization (Util) and length of stay (LoS) are two key performance indicators in the OR peri-op process deriving many areas of the systems such as cost and revenue. However, studies show that the objectives of $\max (U t i l)$ and $\min (L o S)$ are inconsistent.

With inconsistencies among objectives, the outcomes of the OR peri-op process are not predictable which leads the system into a risky state. We accept $\alpha-C V a R$ as the risk measure, and identify the inconsistencies among objectives as the major risk source, and poor performance as the risk consequence. We propose trade-off balancing as a risk mitigating strategy in the OR peri-op process.

Surgical case sequencing significantly affects the distribution of the outcomes and has the potential to move the performance of the system towards a less risky situation in which the outcomes are more predictable. We model the OR peri-op process as a 3 -stage flow shop scheduling problem. Through extensive case studies on the historical data of UKHC, we show the efficiency and effectiveness of trade-off balancing in mitigating the risks.

The remainder of this chapter is organized as follows: in Section 6.2, we present the mathematical models of the study. In Section 6.3 , we examine the performance of different sequencing methods on the UKHC ORs historical data. In Section 6.4, we discuss the results
of the case studies. Finally, in Section 6.5, we draw conclusions and present future research directions.

### 6.2 Problem description

In this section, we present the mathematical models to calculating $\alpha-C V a R$ in a 3 -stage permutation flow shop representing the OR peri-op process. We use normalized deviations to formulate the trade-off in the OR peri-op process. We utilize the amount of trade-off generated in the system as the indicator of the risk source. We formulate $\alpha-C V a R$ as a function of the sequence of surgical cases, the portfolio of preferences, and the amount trade-off generated by the sequence. In an $N$-job serial process, there are $N$ ! possible sequences i.e. the cardinality of the decision space is $|\Omega|=N!$. As it was mentioned earlier, in a serial process, the utilization (Util) and length of stay ( $L o S$ ) are assessed with maximum completion time ( $M C T$ ) and total completion time (TCT), respectively. Therefore, hereafter, we use $M C T$, and $T C T$ as the equivalents for $U t i l$ and $L o S$, respectively.

## Notations

| $N$ | Number of surgical cases |
| :--- | :--- |
| $S$ | Number of stages |
| $p_{n, s}$ | Service time of case $n$ at stage $s, n=1,2,, N, s=1,2,, S$ |
| $\mathbf{P}$ | Matrix of serive times, $P(n, s)=p_{n, s}$ |
| $\Omega$ | Decision space, $\|\Omega\|=N!$ |
| $\pi$ | $\pi \in \Omega$, a permutation of $N$ cases |
| $i$ | Index of positions in sequence $\pi, i=1,2, \ldots, N$ |
| $\pi(i)$ | Case that occupies position $i$ in sequence $\pi=\{\pi(1), \pi(2), \ldots, \pi(N)\}$ |
| $k$ | Index of attributes, $k \in\{1,2\}=\{M C T, T C T\}$ |
| $\gamma_{k}(\pi, \mathbf{P})$ | A function describing attribute $k$ given $\pi$ and $\mathbf{P}$ |
| $L B_{k}(\mathbf{P})$ | The minimum possible value for attribute $k$ given $\mathbf{P}$ |
| $U B_{k}(\mathbf{P})$ | The maximum possible value for attribute $k$ given $\mathbf{P}$ |
| $y_{k}(\pi, \mathbf{P})$ | The normalized deviation of attribute $k$ from its best value given $\pi$ and $\mathbf{P}$ |

Given $\mathbf{P}$ and $\pi$, we can formulate $(T C T)$ and $(M C T)$ as follows:

$$
\begin{gather*}
C_{i, 1}=\sum_{l=1}^{i} p_{\pi(l), 1} \\
C_{i, s}=\max \left(C_{i-1, s}, C_{i, s-1}\right)+p_{\pi(i), s} \\
T C T=\gamma_{1}(\pi, \mathbf{P})=\sum_{n=1}^{N} C_{i, S}  \tag{6.2}\\
M C T=\gamma_{2}(\pi, \mathbf{P})=C_{N, S} \tag{6.3}
\end{gather*}
$$

As we mentioned in Section 6.1 the objectives of $\min \left(\gamma_{1}\right)$ and $\min \left(\gamma_{2}\right)$ are inconsistent [114, 132], therefore, finding a sequence $\pi$ that simultaneously optimizes both objectives is impossible. Instead, we formulate the trade-off in the system by utilizing the normalized deviations from the best possible values. Let $y_{k}(\pi, \mathbf{P})=\frac{\gamma_{k}(\pi, \mathbf{P})-L B_{k}}{U B_{k}-L B_{k}}$ be the normalized deviation of attribute $k$ from its best value, where $L B_{k}=\min _{\pi \in \Omega}\left(\gamma_{k}(\pi, \mathbf{P})\right)$, and $L B_{k}=\max _{\pi \in \Omega}\left(\gamma_{k}(\pi, \mathbf{P})\right)$.

Assume that $y_{k}$ has a density function $p\left(y_{k}\right)$ with mean $\mu_{k}$ and variance $\sigma_{k}^{2}$. Now, we consider the case where the decision vector $\mathbf{x}$ represents a portfolio of preferences on $y_{k}$ such that $\mathbf{x}=\left(x_{1}, x_{2}\right)$ with the following conditions: $x_{k} \geq 0$ and $x_{1}+x_{2}=1$.

Let $\mathbf{Y}=\left(y_{1}, y_{2}\right)$ denote a random vector representing the normalized deviations. The distribution of $\mathbf{Y}$ is a joint distribution of $y_{1}$ and $y_{2}$ and is independent of $\mathbf{x}$ with density of $p(\mathbf{Y})$. Let us define the loss of portfolio $\mathbf{x}$ as the sum of normalized deviations on individual $y_{k}$ scaled by proportion $x_{k}$. Therefore, the loss is given by equation 6.2).

$$
\begin{gather*}
f(\mathbf{x}, \mathbf{Y})=x_{1} \frac{\gamma_{1}(\pi, \mathbf{P})-L B_{1}(\mathbf{P})}{U B_{1}(\mathbf{P})-L B_{1}(\mathbf{P})}+x_{2} \frac{\gamma_{2}(\pi, \mathbf{P})-L B_{2}(\mathbf{P})}{U B_{2}(\mathbf{P})-L B_{2}(\mathbf{P})} \Rightarrow \\
f(\mathbf{x}, \mathbf{Y})=\left[x_{1} y_{1}+x_{2} y_{2}\right]=\mathbf{x}^{T} \mathbf{Y} \tag{6.4}
\end{gather*}
$$

Given mean and variance of $y_{k}$, we have the vector of means $\mathbf{m}=\left(\mu_{1}, \mu_{2}\right)$ and the variance matrix $\mathbf{V}\left(y_{k}\right)$. Therefore, we are able to calculate the mean and variance of the loss associated with portfolio x as follows:

$$
\begin{array}{r}
\mu_{f(\mathbf{x}, \mathbf{Y})}=\mathbf{x}^{T} \mathbf{m} \\
\sigma_{f(\mathbf{x}, \mathbf{Y})}^{2}=\mathbf{x}^{T} \mathbf{V} \mathbf{x} \tag{6.6}
\end{array}
$$

For each $\mathbf{x}$, the loss $f(\mathbf{x}, \mathbf{Y})$ is random variable with a distribution in $\mathbb{R}$ induced by the $p(\mathbf{Y})$. The probability of $f(\mathbf{x}, \mathbf{Y})$ not exceeding a given threshold $\xi$ is given by equation 6.7)

$$
\begin{equation*}
\Psi(\mathbf{x}, \xi)=\int_{f(\mathbf{x}, \mathbf{Y}) \leq \xi} p(\mathbf{Y}) d \mathbf{Y} \tag{6.7}
\end{equation*}
$$

As a function of $\xi$ for a fixed $\mathbf{x}, \Psi(\mathbf{x}, \xi)$ is the cumulative distribution function for the loss associated with $\mathbf{x}$. This random variable is fundamental in defining $\alpha-\operatorname{VaR}$ and $\alpha$-CVaR in the OR peri-op process. The $\alpha-\mathrm{VaR}$ and $\alpha-\mathrm{CVaR}$ values for the loss random variable associated with $\mathbf{x}$ and any specified probability level $\alpha \in(0,1)$ will be denoted by $\xi_{\alpha}(\mathbf{x})$ and $\phi_{\alpha}(\mathbf{x})$ as follows:

$$
\begin{array}{r}
\xi_{\alpha}(\mathbf{x})=\min \{\xi \in \mathbb{R} ; \quad \Psi(\mathbf{x}, \xi) \geq \alpha\} \\
\phi_{\alpha}(\mathbf{x})=\frac{1}{1-\alpha} \int_{f(\mathbf{x}, \mathbf{Y}) \geq \xi_{\alpha}(\mathbf{x})} f(\mathbf{x}, \mathbf{Y}) p(\mathbf{Y}) d \mathbf{Y} \tag{6.9}
\end{array}
$$

Rockafellar and Uryasev 122 characterized $\xi_{\alpha}(\mathbf{x})$ and $\phi_{\alpha}(\mathbf{x})$ in terms of a function $F_{\alpha}(\mathbf{x}, \xi)$ given by equation (6.10) and showed that $F_{\alpha}(\mathbf{x}, \xi)$ is convex and continuously differentiable.

$$
\begin{equation*}
F_{\alpha}(\mathbf{x}, \xi)=\xi+\frac{1}{1-\alpha} \int_{\mathbf{Y} \in \mathbb{R}^{m}}[f(\mathbf{x}, \mathbf{Y})-\xi]^{+} p(\mathbf{Y}) d \mathbf{Y} \tag{6.10}
\end{equation*}
$$

where $[t]^{+}=\max (t, 0)$.
The $\alpha$-CVaR of the loss associated with any $\mathbf{x}$ can be determined by $\phi_{\alpha}(\mathbf{x})=\min _{\xi \in \mathbb{R}}\left(F_{\alpha}(\mathbf{x}, \xi)\right)$. The integral in equation 6.10 can be approximated in by sampling the probability distribution of $\mathbf{Y}$ according to its density $p(\mathbf{Y})$. If the sampling generates a collection of vectors $\mathbf{Y}_{\mathbf{1}}, \mathbf{Y}_{\mathbf{2}}, \ldots, \mathbf{Y}_{\mathbf{Q}}$, then the corresponding approximation can be written by equation 6.11.
$\widetilde{F}_{\alpha}(\mathbf{x}, \xi)$ in convex and piecewise linear with respect to $\xi$ [122], and it can be minimized by linear programming.

$$
\begin{equation*}
\widetilde{F}_{\alpha}(\mathbf{x}, \xi)=\xi+\frac{1}{Q(1-\alpha)} \sum_{q=1}^{Q}\left[f\left(\mathbf{x}, \mathbf{Y}_{\mathbf{q}}\right)-\xi\right]^{+} \tag{6.11}
\end{equation*}
$$

In order to systematically evaluate the performance of sequencing methods in the OR peri-op process, we define three objective functions covering average loss (i.e. $\min \left(\mu_{f(\mathbf{x}, \mathbf{Y})}\right)$, variance of loss (i.e. $\min \left(\sigma_{f(\mathbf{x}, \mathbf{Y})}^{2}\right)$, and $\alpha$-CVaR of loss $\left(\right.$ i.e. $\min \left(\tilde{F}_{\alpha}(\mathbf{x}, \xi)\right)$ ).

Problem (P.1) is a linear function of $\mathbf{x}$ that minimizes the average value of loss associated with portfolio $\mathbf{x}$. Constraints 6.13 imposes that the proportion of $y_{k}$ in portfolio $\mathbf{x}$ is nonnegative. Constraints 6.14 imposes that the sum of proportions in portfolio x is equal 1 .

$$
\begin{align*}
& \text { (P.1) } \\
& \begin{aligned}
\min \left(\mu_{f(\mathbf{x}, \mathbf{Y})}\right) & =\mathbf{x}^{T} \mathbf{m} \\
\text { s.t. } \quad x_{1}, x_{2} & \geq 0 \\
x_{1}+x_{2} & =1
\end{aligned} \tag{6.12}
\end{align*}
$$

Problem (P.2) is a quadratic function of $\mathbf{x}$ that minimizes the variance of loss associated with portfolio x subject to the linear constraints (6.13) and (6.14).

$$
\begin{align*}
& (P .2) \\
& \quad \min \left(\sigma_{f(\mathbf{x}, \mathbf{Y})}^{2}\right)=\mathbf{x}^{T} \mathbf{V} \mathbf{x}  \tag{6.15}\\
& \quad \text { s.t. } 6.13), 6.14 \tag{6.16}
\end{align*}
$$

Problem (P.3) is a linear function of $\mathbf{x}$ and $\xi$ that minimizes the $\alpha$-CVaR of loss associated with portfolio $\mathbf{x}$ for a given probability level $\alpha \in(0,1)$ subject to the linear constraints (6.13) and (6.14)..

$$
\begin{equation*}
\min \left(\tilde{F}_{\alpha}(\mathbf{x}, \xi)\right)=\xi+\frac{1}{Q(1-\alpha)} \sum_{q=1}^{Q}\left[\mathbf{x}^{T} \mathbf{Y}_{\mathbf{q}}-\xi\right]^{+} \tag{P.3}
\end{equation*}
$$

$$
\begin{equation*}
\text { s.t. } 6.13, \text { 6.14 } \tag{6.17}
\end{equation*}
$$

Given the $N P$-completeness of the problem [133, 134] it is not possible to optimally solve $\min (M C T)$ and $\min (T C T)$ in a reasonable computation time. Instead, we review the performance of 16 scheduling methods including $F C F S, L P T, S P T, N E H, L R$, and $C F D(\beta)$ with ( $\beta=0.00: 0.10, \ldots, 1.0$ ) in terms of P.1, P.2, and P. 3 on the historical data. In section 6.3, we provide the details of the case studies.

### 6.3 Case studies

We examine the performance several sequencing methods including FCFS, LPT, SPT, $N E H, L R$, and $C F D(\beta)$. Historical data of 260 days per year, for 5 years are used in this case study ( $Q=1300$ samples in total). Let $q=1,2, \ldots, Q$ be the index of samples, and $\mathbf{P}_{q}$ be the service times matrix of surgical cases in sample $q$. Let $\pi_{h}$ be the sequence generated
by method $h \in H=\{1,2, \ldots, 16\}=\{F C F S, S P T, L P T, N E H, L R, C F D 00, \ldots, C F D 10\}$, therefore, we have the normalized deviation of attribute $k$ generated by method $h$ from its possible best value as $y_{h, k, q}\left(\pi_{h}, P_{q}\right)=\frac{\gamma_{k}\left(\pi_{h}, P_{q}\right)-L B_{k, q}}{U B_{k, q}-L B_{k, q}}$. Where $U B_{k, q}=\max _{h \in H}\left(\gamma_{k}\left(\pi_{h}, P_{q}\right)\right)$, and $L B_{k, q}=\min _{h \in H}\left(\gamma_{k}\left(\pi_{h}, P_{q}\right)\right)$.

Let $A N D_{k, h}=\frac{\sum_{q=1}^{Q} y_{h, k, q}}{Q}$ be the average normalized deviation of attribute $k$ generated by method $h$ over all $Q$ samples. After calculating $A N D_{k, h}$ for all sequencing methods, we use the condition of Pareto dominance to assess if some methods are dominated by others. For minimization problems, if $x_{A}^{[k]}$ and $x_{B}^{[k]} \in \mathbb{R}^{K}$ are two vectors that measure a positive attribute $k$ such as the utility of decisions $A$ and $B$, respectively, decision $A$ dominates decision $B$ if the following conditions are satisfied:

$$
\begin{array}{lll}
x_{A}^{[k]} \leq x_{B}^{[k]}, & \forall k \in\{1,2, . ., K\} \\
x_{A}^{[k]}<x_{B}^{[k]}, & \exists & k \in\{1,2, . ., K\} \tag{6.20}
\end{array}
$$

Equation 7.14 states that decision $A$ is not worse than decision $B$ in any dimension, while equation 7.15 states that decision $A$ is better than decision $B$ at least in one dimension. Pareto optimal outcomes cannot be improved without sacrificing of at least one objective.

### 6.4 Results and Discussions

Figure 6.2 shows $A N D_{k, h}$ for all 16 methods. It can be clearly observed that the simples rules of $F C F S, S P T$, and $L P T$ are dominated by the other methods. Therefore, we eliminate the dominated methods from $H$, thus, we recalculate $U B_{k, q}=\max _{h \in H}\left(\gamma_{k}\left(\pi_{h}, P_{q}\right)\right)$, $L B_{k, q}=\min _{h \in H}\left(\gamma_{k}\left(\pi_{h}, P_{q}\right)\right)$, and $y_{h, k, q}\left(\pi_{h}, P_{q}\right)=\frac{\gamma_{k}\left(\pi_{h}, P_{q}\right)-L B_{k, q}}{U B_{k, q}-L B_{k, q}}$ with $h \in H=\{1,2, \ldots, 13\}=$ $\{N E H, L R, C F D 00, \ldots, C F D 10\}$. The reason for eliminating the dominated methods is that the dominated methods enlarge the value of $U B_{k}-L B_{k}$ which may be misleading in calculating the normalized deviations.


Figure 6.2: Average Normalized Deviations of sequencing methods

## Loss expected value(P.1)

Using recalculated $y_{h, k, q}$ for non-dominated methods, we calculate the expected value of loss associated with method $h$ for different portfolios $\mathbf{x}=\left(x_{1}, x_{2}\right)$ as shown by Table 6.1. CFD02 generates the minimum amount of expected loss of 0.0512 , and as it is shown by figure 6.3 it is less sensitive to the portfolio $\mathbf{x}$. NEH and $L R$ generated the expected loss of 0.5000 and 0.4984 , respectively, and it is observed that they are highly sensitive to the portfolio $\mathbf{x}$ with associated expected loss ranging from 0 to 1 for $N E H, 0.0032$ to 0.9935 for $L R$. It is also observed that CFD10 and CFD00 perform poorly comparable to those of $N E H$ and $L R$. Therefore, we can conclude that the single attribute optimizations are not able to address different decision maker's preference portfolios even for the expected value of loss.

Table 6.1: Expected loss of 13 scheduling methods with different portfolios

| Portfolio | Method |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\left(x_{1}, x_{2}\right)$ | NEH | LR | CFD00 CFD01 CFD02 | CFD03 | CFD04 | CFD05 | CFD06 | CFD07 CFD08 CFD09 CFD10 |  |  |  |  |  |
| $(0.0,1.0)$ | 1.0000 | 0.0032 | 0.0012 | 0.0180 | 0.0325 | 0.0499 | 0.0820 | 0.0999 | 0.1465 | 0.1573 | 0.1733 | 0.1941 | 0.6144 |
| $(0.1,0.9)$ | 0.9000 | 0.1022 | 0.0654 | 0.0249 | 0.0362 | 0.0511 | 0.0797 | 0.0952 | 0.1356 | 0.1446 | 0.1582 | 0.1760 | 0.5530 |
| $(0.2,0.8)$ | 0.8000 | 0.2013 | 0.1297 | 0.0319 | 0.0400 | 0.0523 | 0.0773 | 0.0905 | 0.1246 | 0.1319 | 0.1431 | 0.1579 | 0.4915 |
| $(0.3,07)$ | 0.7000 | 0.3003 | 0.1939 | 0.0388 | 0.0437 | 0.0535 | 0.0750 | 0.0858 | 0.1137 | 0.1192 | 0.1280 | 0.1398 | 0.4301 |
| $(0.4,0.6)$ | 0.6000 | 0.3993 | 0.2581 | 0.0457 | 0.0474 | 0.0547 | 0.0727 | 0.0811 | 0.1028 | 0.1065 | 0.1129 | 0.1217 | 0.3687 |
| $(0.5,0.5)$ | 0.5000 | 0.4984 | 0.3224 | 0.0527 | 0.0512 | 0.0559 | 0.0703 | 0.0764 | 0.0919 | 0.0938 | 0.0979 | 0.1035 | 0.3072 |
| $(0.6,0.4)$ | 0.4000 | 0.5974 | 0.3866 | 0.0596 | 0.0549 | 0.0571 | 0.0680 | 0.0717 | 0.0810 | 0.0811 | 0.0828 | 0.0854 | 0.2458 |
| $(0.7,0.3)$ | 0.3000 | 0.6964 | 0.4508 | 0.0666 | 0.0587 | 0.0583 | 0.0656 | 0.0669 | 0.0701 | 0.0684 | 0.0677 | 0.0673 | 0.1843 |
| $(0.8,0.2)$ | 0.2000 | 0.7955 | 0.5151 | 0.0735 | 0.0624 | 0.0595 | 0.0633 | 0.0622 | 0.0591 | 0.0557 | 0.0526 | 0.0492 | 0.1229 |
| $(0.9,0.1)$ | 0.1000 | 0.8945 | 0.5793 | 0.0804 | 0.0661 | 0.0607 | 0.0609 | 0.0575 | 0.0482 | 0.0430 | 0.0375 | 0.0311 | 0.0615 |
| $(1.0,0.0)$ | 0.0000 | 0.9935 | 0.6436 | 0.0874 | 0.0699 | 0.0619 | 0.0586 | 0.0528 | 0.0373 | 0.0303 | 0.0225 | 0.0130 | 0.0000 |
| Overall | 0.5000 | 0.4984 | 0.3224 | 0.0527 | 0.0512 | 0.0559 | 0.0703 | 0.0764 | 0.0919 | 0.0938 | 0.0979 | 0.1035 | 0.3072 |



Figure 6.3: Loss expected value of the sequencing methods for different portfolios

## Loss variance(P.2)

Table 6.2 shows the loss standard deviations generated by the studied sequencing methods for different portfolios $\mathbf{x}$. It is observed that CFD03 generates the minimum loss standard deviation over all portfolios with the value of 0.0355 . As shown by figure 6.4, although $N E H, L R$ show lower loss variance in each portfolio their variance is the two highest over all portfolios (within groups).

Table 6.2: Loss standard deviation of the scheduling methods with different portfolios

| Portfolio | Method |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\left(x_{1}, x_{2}\right)$ | NEH | LR | CFD00 CFD01 CFD02 CFD03 | CFD04 CFD05 CFD06 CFD07 CFD08 CFD09 CFD10 |  |  |  |  |  |  |  |  |  |
| $(0.0,1.0)$ | 0.0000 | 0.0026 | 0.0040 | 0.0104 | 0.0158 | 0.0200 | 0.0251 | 0.0342 | 0.0389 | 0.0356 | 0.0381 | 0.0421 | 0.0429 |
| $(0.1,0.9)$ | 0.0000 | 0.0043 | 0.0341 | 0.0123 | 0.0158 | 0.0192 | 0.0232 | 0.0306 | 0.0346 | 0.0318 | 0.0341 | 0.0378 | 0.0386 |
| $(0.2,0.8)$ | 0.0000 | 0.0073 | 0.0677 | 0.0173 | 0.0180 | 0.0202 | 0.0227 | 0.0281 | 0.0310 | 0.0285 | 0.0303 | 0.0336 | 0.0343 |
| $(0.3,07)$ | 0.0000 | 0.0104 | 0.1014 | 0.0234 | 0.0216 | 0.0226 | 0.0238 | 0.0270 | 0.0282 | 0.0259 | 0.0269 | 0.0295 | 0.0300 |
| $(0.4,0.6)$ | 0.0000 | 0.0136 | 0.1350 | 0.0300 | 0.0262 | 0.0261 | 0.0262 | 0.0273 | 0.0265 | 0.0242 | 0.0241 | 0.0257 | 0.0257 |
| $(0.5,0.5)$ | 0.0000 | 0.0169 | 0.1687 | 0.0369 | 0.0312 | 0.0303 | 0.0297 | 0.0291 | 0.0261 | 0.0236 | 0.0221 | 0.0223 | 0.0215 |
| $(0.6,0.4)$ | 0.0000 | 0.0202 | 0.2024 | 0.0438 | 0.0366 | 0.0349 | 0.0339 | 0.0320 | 0.0271 | 0.0242 | 0.0211 | 0.0194 | 0.0172 |
| $(0.7,0.3)$ | 0.0000 | 0.0235 | 0.2361 | 0.0508 | 0.0421 | 0.0398 | 0.0386 | 0.0359 | 0.0292 | 0.0259 | 0.0212 | 0.0173 | 0.0129 |
| $(0.8,0.2)$ | 0.0000 | 0.0268 | 0.2698 | 0.0578 | 0.0477 | 0.0449 | 0.0437 | 0.0405 | 0.0324 | 0.0286 | 0.0225 | 0.0164 | 0.0086 |
| $(0.9,0.1)$ | 0.0000 | 0.0301 | 0.3034 | 0.0649 | 0.0534 | 0.0502 | 0.0489 | 0.0455 | 0.0363 | 0.0319 | 0.0248 | 0.0167 | 0.0044 |
| $(1.0,0.0)$ | 0.0000 | 0.0334 | 0.3371 | 0.0720 | 0.0592 | 0.0556 | 0.0543 | 0.0508 | 0.0407 | 0.0357 | 0.0277 | 0.0184 | 0.0006 |
| Overall | 0.3162 | 0.3138 | 0.2847 | 0.0485 | 0.0384 | 0.0355 | 0.0361 | 0.0384 | 0.0473 | 0.0496 | 0.0549 | 0.0632 | 0.1959 |



Figure 6.4: Loss average and standard deviation frontiers, each point shows a portfolio

As it is shown by Table 6.1, Table 6.2, and Figure 6.4, there are trade-offs among the results of P. 1 and P. 2 (i.e. $\min \left(\mu_{f(x, Y)}\right)$ and $\min \left(\sigma_{f(x, Y)}^{2}\right)$, which means minimizing loss expected value does not necessarily lead to minimizing loss variance. Therefore, there is trade-offs between the first order and the second order effects of loss in the OR peri-op process.

Figure 6.5 shows the expected value and standard deviation of the sequencing methods over all portfolios. Single attribute optimization methods including $N E H, L R, C F D 00$,


Figure 6.5: Scheduling methods loss average and standard deviation over all portfolio
and $C F D 10$, are clearly dominated by trade-off balancing methods. It is observed that CFD02 and CFD03 are the only two non-dominated methods in terms of both loss expected value and standard deviation. The Pareto frontiers of CFD02 and CFD03 are shown by figure 6.6. CFD03 shows a fairly stable performance on loss expected value for different


Figure 6.6: Loss expected value and standard deviation frontiers of non-dominated methods, each point shows a portfolio
portfolios (from left to right $\mathbf{x}=(0.0,1.0),(0.1,0.9), \ldots,(1.0,0.0))$, whereas its performance on loss standard deviation degrades. CFD02 seems to be more sensitive to $\mathbf{x}$ in terms of both loss expected value and standard deviations. However, CFD02 dominates CFD03 for the first seven portfolios (i.e. $\mathbf{x}=(0.0,1.0),(0.1,0.9), \ldots,(0.7,0.3)$, whereas CFD03 dominates CFD02 for the last three portfolios (i.e. $\mathbf{x}=(0.8,0.2),(0.9,0.1),(1.0,0.0)$. Therefore, finding a method that is dominant in the first order and second order of loss is necessary. As it was mentioned earlier, if a method has a lower CVaR, it dominates other methods in the stochastic first order and the second order. In subsection 6.4, we compare the performance of the sequencing methods in terms of $C V a R$.

## Conditional value at risk (P.3)

The results of our study in subsections 6.4 and 6.4 showed that there are two sources of tradeoffs in an OR peri-op process including (i) inconsistency between the objectives of max (Util) and $\min (L o S)$, (ii) inconsistency between the first order (expected value) and the second order (variance) of loss. To address (i), we propose to utilize trade-off balancing as the objective function, and to address (ii) as mentioned earlier, we utilize minimizing conditional value at risk i.e. $\min (\alpha-C V a R)$ described by P. 3 in section 6.2 . Pflug 129 proved that $C V a R$ is a coherent risk measure with monotonic first and second order stochastic dominance properties.

As shown in Table 6.3, we calculate $\alpha-C V a R$ at three common probability levels of $\alpha \in$ $\{0.90,0.95,0.99\}$. We also calculate the range of $\alpha-C V a R$ at each probability level denoted by Max - Min in Table 6.3. Max - Min measures the sensitivity of methods to portfolios thus the smaller Max - Min the less sensitive the method. Methods with small values of Max - Min are able to address a wider range of managerial preferences. It is observed that single attribute scheduling methods including $N E H, L R, C F D 00, C F D 10$ have the highest $\alpha-C V a R$, and are highly sensitive to both the probability level $\alpha$ and the portfolio $\mathbf{x}$.

At $\alpha=0.90$ and $\alpha=0.95$, CFD03 dominates other sequencing methods, whereas, at $\alpha=0.99$, CFD02 dominates other sequencing methods. CFD03 has $0.90-C V a R=0.0603$ and $0.95-C V a R=0.0683$ which means in the $10 \%$ and $5 \%$ of the worst outcomes the expected value of loss is 0.0603 and 0.0683 , respectively. In the $1 \%$ of the worst outcomes, CFD02 has the expected value of loss of 0.0850 . As it is shown by figure 6.7d and figure 6.7a to figure 6.7k, CFD03 and CFD02 are sensitive to neither the probability level $\alpha$ nor the portfolio of preferences $\mathbf{x}$ in terms of $\alpha-C V a R$.

It is also seen that all the bi-objective trade-off functions (i.e. CFD01, .., CFD09) intersect at portfolio $\mathbf{x}=(0.5,0.5)$ with fairly similar $\alpha-C V a R$, therefore, equal preference for Util and $L o S$ provides the decision maker with a desirable flexibility on the sequencing method in terms of $\alpha-C V a R$.

## Overall evaluation

In subsection 6.4, we demonstrated $\alpha-C V a R$ and trade-off balancing as the proper approach for surgical case sequencing in an OR peri-op process. According to the adopted risk definition "the variations in the distributions of OR peri-op process outcomes, their likelihood, and their subjective values", a risk mitigating strategy not only must move the outcomes of the system towards a more predictable region but also must result in acceptable subjective values for the objectives. Table 6.4 shows the performance of CFD03 (with minimum $0.95-C V a R$ ), NEH (the best method max $(U t i l)$ ), and $L R$ (the best method to $\min (L o S)$ ) in terms of average loss, value at risk $(V a R)$, conditional value at risk $(C V a R)$, and average normalized deviations for $M C T$ and $T C T$ (i.e. $A N D_{k}$ ). The arguments for other probability level is similar, therefore, are omitted for the sake of brevity.

It is observed that CFD03 (i.e. the best method for $\min (C V a R)$ at $\alpha=0.95)$ not only has the minimum value of $C V a R$ with small variations for different portfolios, but also provides small and uniform variations for $A N D_{k}$. CFD03 generates only $3.2 \%$ and $1.9 \%$ deviations from the best possible values of $M C T$ and $T C T$, respectively. Therefore, we conclude that trade-off balancing can provide the decision maker with a powerful tool to move the outcomes of the OR peri-op process towards a less risky point with acceptable subjective values.

Table 6.3: $\alpha-C V a R$ of sequencing methods at $\alpha \in\{0.90,0.95,0.99\}$

| $\alpha$ | $\begin{gathered} \mathbf{x} \\ \left(x_{1}, x_{2}\right) \end{gathered}$ | Sequencing method |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | NEH LR | CFD00 |  |  |  |  | 05 |  |  |  |  |  |
| 0.90 | (0.0,1.0) | 0.54810 .003 | 0.0043 | 0.0147 | 0.0244 | 0.0343 | 0.0496 | 0.0667 | 0.0967 | 0.0970 | 0.1042 | 0.117 | 0.3 |
|  | (0.1,0.9) | 0. | 0.0705 | 0.0219 | 0.0281 | 0.0359 | 0.0489 | 0.0630 | 0.0889 | 0.0888 | 0.0950 | 0.1063 | 0.3108 |
|  | $(0.2,0.8)$ | 0. | 0.1399 | 0.0318 | 0.034 | 0.0399 | 0.050 | 0.0605 | 0.0816 | 0.0811 | 0.0860 | 0.0953 | 0.2763 |
|  | $(0.3,0.7)$ | 0. | 0.2093 | 0.0425 | 0.0427 | 0.0456 | 0.0540 | 0.0601 | 0.0750 | 0.0741 | 0.0776 | 0.0844 | 0.2417 |
|  | $(0.4,0.6)$ | 0.3 | 0.2787 | 0.0534 | 0.0510 | 0.0520 | 0.0583 | 0.0613 | 0.0700 | 0.0683 | 0.0699 | 0.073 | 0.2072 |
|  | $(0.5,0.5)$ | 0.27410 .37 | 0.3481 | 0.0644 | 0.0595 | 0.0586 | 0.0631 | 0.0635 | 0.0665 | 0.0639 | 0.0632 | 0.0636 | 0.17 |
|  | $(0.6,0.4)$ | 0. | 0.4175 | 0.0754 | 0.0682 | 0.0655 | 0.0683 | 0.0667 | 0.0647 | 0.0610 | 0.0579 | 0.0541 |  |
|  | $(0.7,0.3)$ | 0.16440 .52 | 0.4869 | 0.0864 | 0.0769 | 0.0724 | 0.0737 | 0.0704 | 0.0641 | 0.0591 | 0.0537 | 0.0457 |  |
|  | $(0.8,0.2)$ | 0. | 0.55 | 0.097 | 0.0857 | 0.0795 | 0.0792 | 0.0746 | 0.0642 | 0.0578 | 0.0502 | 0.038 | 0.0691 |
|  | (0.9,0.1 | 0.0 | 0.6 | 0.1085 | 0.0946 | 0.0865 | 0.0847 | 0.0790 | 0.0647 | 0.0570 | 0.0474 | 0.033 | 0.0346 |
|  | (1.0,0.0) | 0.00000 .746 | 0.6951 | 0.1195 | 0.1034 | 0.0936 | 0.0903 | 0.0836 | 0.0655 | 0.0564 | 0.0449 | 0.0286 | 0.0002 |
|  | Overall | 0.27410 .374 | 0.3484 | 0.0651 | 0.0609 | 0.0603 | 0.0655 | 0.0681 | 0.0729 | 0.0695 | 0.0682 | 0.0674 | . 17 |
|  | Max- | 0.54810 .743 | 0.6908 | 0.1049 | 0.0790 | 0.0593 | 0.0414 | 0.0235 | 0.0326 | 0.0407 | 0.0593 | 0.0887 | 0.3452 |
| 0.95 |  |  | 0.0061 | 0.0167 | 0.0 | 0.0 | 0.0 | 0.0725 | 0.1055 | 0.1068 | 0.1127 | 0.1271 | 0.3664 |
|  | $(0.1,0.9)$ | 0.5 | 0.0750 | 0.0248 | 0.0315 | 0.0387 | 0.0527 | 0.0684 | 0.0970 | 0.0976 | 0.1028 | 0.1150 | . 32 |
|  | $(0.2,0.8)$ | 0.4 | 0.1483 | 0.036 | 0.0391 | 0.04 | 0.05 | 0.065 | 0.0890 | 0.08 | 0.09 | 0.1031 | 0.2931 |
|  | $(0.3,0.7)$ | 0.4 | 0.2 | . 048 | 0.0481 | 0.0505 | 0.0590 | 0.0648 | 0.0817 | 0.0804 | 0.0836 | 0.09 | 0.2564 |
|  | $(0.4,0.6)$ | 0.34770 .3128 | 0.2950 | 0.0608 | 0.0576 | 0.0581 | 0.0647 | 0.0669 | 0.0759 | 0.0739 | 0.0752 | 0.0799 | . 21 |
|  | $(0.5,0.5)$ | 0.2 | 0.3684 | 0.0734 | 0.0674 | 0.0662 | 0.0710 | 0.0705 | 0.0728 | 0.0699 | 0.0686 | 0.0689 | 0.18 |
|  | $(0.6,0.4)$ | 0.23180 .468 | 0.441 | 0.0861 | 0.0774 | 0.0744 | 0.0775 | 0.0752 | 0.0717 | 0.0676 | 0.0638 | 0.0589 | 0.1465 |
|  | $(0.7,0.3)$ | 0. | 0.5 | 0.098 | 0.087 | 0.0828 | 0.0842 | 0.0806 | 0.072 | 0.0666 | 0.0600 | . 0 | 0.1099 |
|  | $(0.8,0.2)$ |  | 0.5887 | 11 | 0.097 | 0.0912 | 0.090 | 0.0864 | 0.073 | 0.0665 | 0.05 | . 0 | 0.0733 |
|  | (0.9,0.1 | 0. | 0.6 | 0.1243 | 0.1078 | 0.0997 | 0.0978 | 0.0922 | 0.0747 | 0.0665 | 0.0554 | 0.0385 | 0.03 |
|  | (1.0,0.0) | 0.00000 .7801 | 0.7356 | 0.1370 | 0.1180 | 0.1082 | 0.1047 | 0.0982 | 0.0765 | 0.0670 | 0.0538 | 0.0342 | 0.000 |
|  | Overall | 0.28970 .3910 | 0.368 | 0.0744 | 0.0690 | 0.0683 | 0.0737 | 0.0765 | 0.0809 | 0.0774 | 0.075 | 0.0737 | 0.1832 |
|  | Max-Mi | 0.5 | 0.7295 | 0.1203 | 0.0907 | 0.0708 | 0.0520 | 0.0334 | 0.0338 | 0.0404 | 0.0589 | 0.0929 | 0.366 |
| 0.99 | (0.0,1.0) |  | 0.0099 | 0.0224 | 0.0 | 0.0 | 0.0 | 0.0 | 0.123 | 0.1287 | 0.1300 | 0.1516 | 0.4146 |
|  | $(0.1,0.9)$ | 0.5 | 0.0842 | 0.0312 | 0.0386 | 0.0446 | 0.0604 | 0.0796 | 0.1126 | 0.1183 | 0.1185 | 0.1372 | 0.373 |
|  | $(0.2,0.8)$ | 0.52270 .17 | 0.1658 | 0.0448 | 0.0474 | 0.0519 | 0.0606 | 0.0748 | 0.1027 | 0.1088 | 0.107 | 0.1231 | . 33 |
|  | $(0.3,0.7)$ | 0. | 0.2 | 0.0604 | 0.058 | 0.0619 | 0.0679 | 0.0724 | 0.093 | 0.0996 | 0.09 | 0.10 | 0.2902 |
|  | $(0.4,0.6)$ | 0.3 | 0.3302 | 0.0763 | 0.0710 | 0.0725 | 0.0771 | 0.0763 | 0.0882 | 0.0910 | 0.0882 | 0.0953 | 0.2487 |
|  | $(0.5,0.5)$ | 0.32670 .43 | 0.4123 | 0.0926 | 0.0830 | 0.0832 | 0.0869 | 0.0846 | 0.0852 | 0.0838 | 0.0798 | 0.0821 | 0.20 |
|  | $(0.6,0.4)$ | 0.26140 .517 | 4945 | 0.1091 | 0.0952 | 0.0939 | 0.0973 | 0.0933 | 0.0854 | 0.0801 | 0.0737 | 0.0702 | 0.165 |
|  | $(0.7,0.3)$ | 0.1 | 0.5767 | 0.1255 | 0.1078 | 0.1050 | 0.1079 | 0.1023 | 0.0880 | 0.0790 | 0.07 | 0.0607 | . 12 |
|  | $(0.8,0.2)$ | 0. | . 658 | 0.1419 | 0.1204 | 0.1163 | 0.118 | 0.111 | 0.091 | 0.080 | 0.07 | 0.05 | . 08 |
|  | $(0.9,0.1)$ | 0.0 | 10 | 0.1584 | 0.1331 | 0.1277 | 0.1296 | 0.1204 | 0.0964 | 0.0824 | 0.0724 | 0.0502 | 0.04 |
|  | (1.0,0.0) | 0.00000 .8612 | 0.8231 | 0.1748 | 0.1459 | 0.1391 | 0.1404 | 0.1295 | 0.1014 | 0.0853 | 0.0738 | 0.0463 | 0.0015 |
|  | Overall | 0.32670 .4318 | 0.4131 | 0.0943 | 0.0850 | 0.0853 | 0.0919 | 0.0937 | 0.0971 | 0.0943 | 0.0896 | 0.0892 | 0.2074 |
|  | Max-Min | 0.65340 .855 | 0.8132 | 0.1524 | 0.1121 | 0.0969 | 0.0800 | 0.0571 | 0.0379 | 0.0496 | 0.0585 | 0.105 | . 41 |

### 6.5 Conclusion

In this chapter, we addressed two fundamental issues associated with surgical case sequencing in an operating room peri-operative process including (i) inconsistency between two key performance indicators of utilization (Util) and length of stay (LoS), (ii) inconsistency between the first order and the second order of loss (i.e. expected value and variance, respectively). To address (i) we demonstrated that balancing trade-offs as the objective function is superior to the single attribute optimizations of $\max (U t i l)$ and $\min (L o S)$. To address (ii), we showed that optimizing conditional value at risk $(C V a R)$ can provide the first order and the second order dominance.

By modeling the OR peri-op process as a 3 -stage serial process, we elaborated the risk


Figure 6.7: $\alpha-C V a R$ of sequencing methods

Table 6.4: Performance of selected sequencing methods at $\alpha=0.95$

| Portfolio x | Mehtod |  |  |
| :---: | :---: | :---: | :---: |
|  | CFD03 | NEH | LR |
| $\left(x_{1}, x_{2}\right)$ | $f(\mathbf{x}, \mathbf{Y}) \mathrm{VaR} \mathrm{CVaR} A N D_{1} A N D_{2}$ | $f(\mathbf{x}, \mathbf{Y})$ VaR CVaR $A N D_{1} A N D_{2}$ | $f(\mathbf{x}, \mathbf{Y}) \mathrm{VaR} \mathrm{CVaR} A N D_{1} A N D_{2}$ |
| $(0.0,1.0)$ | 0.01940 .03330 .03740 .03270 .0194 | 0.39830 .53720 .57940 .00000 .3983 | 0.00130 .00340 .00440 .53790 .0013 |
| $(0.1,0.9)$ | 0.02070 .03470 .03870 .03270 .0194 | 0.35850 .48350 .52150 .00000 .3983 | 0.05490 .07430 .07930 .53790 .0013 |
| $(0.2,0.8)$ | 0.02200 .03810 .04380 .03270 .0194 | 0.31870 .42970 .46350 .00000 .3983 | 0.10860 .14720 .15710 .53790 .0013 |
| (0.3,0.7) | 0.02340 .04400 .05050 .03270 .0194 | 0.27880 .37600 .40560 .00000 .3983 | 0.16220 .22010 .23500 .53790 .0013 |
| (0.4,0.6) | 0.02470 .04960 .05810 .03270 .0194 | 0.23900 .32230 .34770 .00000 .3983 | 0.21590 .29320 .31280 .53790 .0013 |
| $(0.5,0.5)$ | 0.02600 .05600 .06620 .03270 .0194 | 0.19920 .26860 .28970 .00000 .3983 | 0.26960 .36640 .39070 .53790 .0013 |
| $(0.6,0.4)$ | 0.02740 .06240 .07440 .03270 .0194 | 0.15930 .21490 .23180 .00000 .3983 | 0.32320 .43960 .46860 .53790 .0013 |
| $(0.7,0.3)$ | 0.02870 .06810 .08280 .03270 .0194 | 0.11950 .16120 .17380 .00000 .3983 | 0.37690 .51280 .54650 .53790 .0013 |
| $(0.8,0.2)$ | 0.03000 .07480 .09120 .03270 .0194 | 0.07970 .10740 .11590 .00000 .3983 | 0.43050 .58590 .62440 .53790 .0013 |
| (0.9,0.1) | 0.03140 .08120 .09970 .03270 .0194 | 0.03980 .05370 .05790 .00000 .3983 | 0.48420 .65910 .70230 .53790 .0013 |
| (1.0,0.0) | 0.03270 .08800 .10820 .03270 .0194 | 0.00000 .00000 .00000 .00000 .3983 | 0.53790 .73230 .78010 .53790 .0013 |
| Overall | 0.02600 .05730 .06830 .03270 .0194 | 0.19920 .26860 .28970 .00000 .3983 | 0.26960 .36680 .39100 .53790 .0013 |
| max- min | $0.01340 .05470 .0708-$ | 0.39830 .53720 .5794 | 0.53660 .72890 .7758 - |

sources, risk consequences, and proposed risk mitigating strategies. Inconsistency between objectives and inconsistency between the first order and the second order of outcomes were identified as the risk sources. High fluctuations in the outcomes values are the risk sequences. Finally, trade-off balancing was proposed as the risk mitigating strategy that leads the out-
come of the OR peri-op process towards a less risky state. Trade-off balancing is capable of reflecting the decision maker's preferences into the objective function by which the decision maker has a powerful yet flexible tool to shift the outcomes of the OR peri-op process towards a more predictable state with average ( $0.95-C V a R$ ) of $6.8 \%$, and acceptable subjective values for key performance indicators with only $3.27 \%$ and $1.94 \%$ deviations from the optimal solutions to $\max (U t i l)$ and $\min (L o S)$, respectively.

## Chapter 7

## Trade-off Balancing for Sustainable Production Scheduling: A Metric-Based Approach

Production scheduling involves operational level decision making at the shop floor that covers not only the manufacturing stage of the product life-cycle but also the use stage of the processes. Despite the substantial research in sustainable manufacturing, a holistic model for sustainable production scheduling is virtually absent. To address this gap, this chapter presents a metric-based model to systematically and holistically evaluate the sustainability of the production schedules. To this aim, we first perform an extensive literature review to identify the fundamental performance metrics in production scheduling. Second, we asses those metrics with respect to the triple bottom lines $(T B L)$ including economic, environmental, and social pillars. Third, we show the inconsistencies among the fundamental performance metrics and consequently among the objectives defined in the $T B L$. Finally, we propose a generic model for production scheduling for sustainability based on balancing the trade-offs among the inconsistent objectives. The efficiency and effectiveness of the proposed model is demonstrated using a comprehensive case study. Balancing trade-offs among the fundamental performance metrics not only provides a sustainable schedule but also results in a better control over the production scheduling fundamental performance metrics.

### 7.1 Introduction

Trade-off balancing in production scheduling is important for sustainable manufacturing, because sustainable manufacturing can be regarded as an optimization problem with many inconsistent objectives. Elkington (1998) [135] introduced the triple bottom line (TBL) including economic, environmental, and social pillars to holistically evaluate the performance of a production firm. Considering all three pillars of the $T B L$, an obvious observation is that the objectives defined in the pillars are not consistent with each other, for example, a production plant may demonstrate excellent monetary profit (i.e. economic pillar) but at the cost of water/air pollution (i.e. environmental pillar). That is, sustainable production scheduling can be seen as an approach to balancing the trade-offs among the inconsistent objectives defined in the $T B L$.

United State Environmental Agency [136] defines sustainable manufacturing as follows: "Sustainable manufacturing is the creation of manufactured products through economicallysound processes that minimize negative environmental impacts while conserving energy and natural resources. Sustainable manufacturing also enhances employee, community and product
safety". We can conclude that sustainable manufacturing is able to address all three pillars of $T B L$.

Jawahir et al. (2015) 137] expressed that sustainable manufacturing must address the $T B L$ at product, process, and system levels. Sustainable manufacturing must also cover all four stages of the product life-cycle including: pre-manufacturing, manufacturing, use, and post-use [138]. Therefore, sustainable manufacturing can serve as a technological tool for the transition from the Linear Economy to the Circular Economy [139].

Production scheduling is an operational level decision made at the shop floor that falls into the manufacturing stage of the product life-cycle; it also covers the use stage of the processes [140]. Production scheduling sequences a set of jobs on one or multiple machines in order to optimize a given objective [141, 142]. The common objectives for production scheduling are minimization of maximum completion time ( $M C T$ ), total completion time $(T C T)$, flow time mean and variance, and tardiness/earliness [143]. Although production scheduling with regard to single pillars of $T B L$ such as greenhouse gas emission [144], energy consumption [145, 146, 147, and labor cost 148 have been reported, there is no report on production scheduling with regard to all three pillars of $T B L$ [149. Therefore, developing an approach to holistically evaluate production scheduling sustainability is necessary.

Currently, we do not have a holistic evaluation scheme that can balance trade-offs not only among inconsistent performance metrics, but also among different pillars of the $T B L$. The closest work to the sustainable production scheduling evaluation is the work proposed by Badurdeen et al. (2015) 150 for evaluating process sustainability. They proposed a four-level hierarchical structure called ProcSI (Process Sustainability Index) that segregates the overall process sustainability into individual process-level quantifiable metrics. ProcSI includes four levels of ProcSI, clusters, sub-clusters, and individual metrics. Their work provided detailed structures for each level. Since different metrics are measured with different units (e.g. cost (\$), energy (Kw), etc.), a normalization scheme is adopted to normalize individual metrics into a 0 to 10 scale. After normalization, a weighting scheme is developed to balance the normalized values according to their relative importance (preference) or level of impact. Once the weights are assigned, the normalized values are aggregated to calculate the scores for subclusters, clusters, and ProcSI. The aggregation follows a bottom-up approach. However, ProcSI is merely focused on evaluating the sustainability of the process and it does not have any consideration of the production schedule.

In this chapter, we propose a generic model for balancing trade-offs among inconsistent objectives defined in the $T B L$ for production scheduling problem. Analogous to inconsistent objectives of the $T B L$, inconsistencies exist among objectives in production scheduling[151]. We hypothesize that balancing trade-offs among inconsistent objectives of production scheduling can result in a sustainable schedule. We model trade-off balancing as a function of $z=f(\cdot)$, which is generally applicable to $\min (\cdot)$ or $\max (\cdot)$ problems at the low (metrics) level, but also consistent at the high (sub-cluster, cluster) levels. We show that balancing trade-offs among production scheduling objectives (e.g. MCT, TCT, etc.) can indirectly balance the trade-offs among the $T B L$ inconsistent objectives resulting in a sustainable schedule. Contributions of this chapter are as follows:

- We provide a comprehensive list of fundamental production scheduling performance metrics that drive multiple areas of the system performance.
- We propose a generic model that attributes the fundamental production scheduling performance metrics with the three pillars of $T B L$. Production scheduling sustainability index (PSSI) is proposed to systematically evaluate the sustainability of the production schedules. PSSI offers a flexible optimization model that is able to address: (i)
the decision maker preferences, (ii) inconsistencies among objectives, (iii) sustainability at the higher level of decision making, and (iv) better control over the production at the lower level.
- Through extensive case studies we show that decision making must be with regard to both sustainability and the process control.

The remainder of this chapter is organized as follows: in Section 7.2, we provide a comprehensive literature review for practices in production scheduling for sustainability, in Section 7.3, we develop the proposed mathematical model for production scheduling for sustainability, in Section 7.4, we present extensive case studies, followed by results and discussions. Finally, in Section 7.5, we draw conclusions and future research directions.

### 7.2 Literature review

In this section, we provide a comprehensive literature review on sustainability practices in production scheduling. For detailed literature on production scheduling, and sustainable manufacturing readers are referred to [143, 149], and [152], respectively. The objective of this section is to systematically review the production scheduling objectives, and establish a bridge between them and the objectives defined in the $T B L$. To this aim, at the end of this section, we categorize the production scheduling objectives into three clusters of (i) Economic-oriented scheduling, (ii) Environmental-oriented scheduling, and (iii) Social-oriented scheduling. Each cluster is then divided into several sub-clusters, and subsequently, each sub-cluster is divided into multiple measurable performance metrics.

## Economic-oriented scheduling

In this subsection, we review recent papers with economic-oriented objectives. We identify four types of economic performance indicators for production scheduling including: (i) production cost, (ii) energy cost, (iii) labor cost, and (iv) inventory cost.

## Production cost

Let $C_{n}$ be the completion time of job $n \in\{1,2, \ldots, N\}$, maximum completion time; $M C T=$ $\max \left(C_{1}, C_{2}, \ldots, C_{N}\right)$ is the time elapsed from the start-time of the first job to the finishtime of the last job. Thus, minimizing $M C T$ i.e. $\min (M C T)$ is equivalent to maximizing the production system utilization i.e. $\max (U t i l)$ [143]. Utilization is the ratio between the actual output of a given resource and the potential output if that resource was fully utilized. A production system with a utilization less than $\% 100$, theoretically has the potential to increase its production output without any recurring and/or capital cost. Production scheduling with the objective of $\min (M C T)$ has been extensively studied, and numerous methods have been proposed for different production environments such as flow shop [153, 154], job shop [155], and continuous manufacturing [156].

Tardiness and earliness costs are other metrics affecting the production cost [157]. Earliness costs may occur if job $n$ is completed earlier than its due date $\left(d_{n}\right)$, the earliness of a job is defined as $E_{n}=\max \left(d_{n}-C_{n}, 0\right)$ [143]. Earliness costs could result from deterioration of the final products or from the need for an extended time/capacity for holding the inventory [158]. Tardiness costs occur if job $n$ is finished after its due date, and includes expedited delivery costs, lost customer, contract penalties, etc. [159, 157]. Tardiness is defined
as $T_{n}=\max \left(C_{n}-d_{n}, 0\right)[143$. If both tardiness and earliness are considered, the objective function can be defined as by equation 7.1 .

$$
\begin{equation*}
\min (z)=\sum_{n=1}^{N}\left(E_{n}+T_{n}\right)=\sum_{n=1}^{N}\left|C_{n}-d_{n}\right| \tag{7.1}
\end{equation*}
$$

In many production systems products are produced in batches with a common due date $d$ [160, 161, 162, therefore, the objective of minimizing earliness and tardiness can be rewritten by equation 7.2 . It is worth mentionimg that in this way earliness and tardiness are equally penalized.

$$
\begin{equation*}
d_{n}=d \Rightarrow \min (z)=\sum_{n=1}^{N}\left|C_{n}-d\right| \tag{7.2}
\end{equation*}
$$

For a special cease where the common due date $d$ is equal to the average completion time, i.e. $d=A C T=\frac{T C T}{N}$, it has been shown than $\min (z)=\sum_{n=1}^{N}\left|C_{n}-A C T\right|$ is equivalent to minimizing completion time variance $(C T V)$ where $C T V=\frac{\sum_{n=1}^{N}\left(C_{n}-A C T\right)^{2}}{N}$. Therefore, $\min (C T V)$ can serve as an alternative objective function to minimize tardiness and earliness around a common due date [163, 90, 91].

## Energy cost

\%31 of primary energy consumption is consumed by the manufacturing industry [159], and utility companies have step-wise pricing policies in which the power costs are significantly higher if the demand exceeds a base load. Therefore, considering a system-wide approach for energy efficiency in production scheduling is necessary to attain sustainable manufacturing [164. Production power must be managed with respect to three main components including: (i) idle energy consumption, (ii) power peak [165], and (iii) inventory handling power consumption.

Minimizing idle energy consumption is equivalent to minimizing machines idle time which is addressed by $\min (M C T)$. Some papers also proposed an "ON-OFF" strategy for shutting the non-bottleneck machines down while idle [166, 167, 168, however, the latter seems to be impractical in many production systems [169]. Power peak management has been addressed by varying the machines speeds in order to shift production load to the off-peak periods [170], however, varying machines speed is not practical in many situations that the quality of products is a function of the machine speed (e.g. roughness of machined parts).

The power consumed for inventory handling is directly related to the inventory level incurred during the period of schedule. Total completion time $T C T=\sum C_{n}$ gives an indication of the inventory level in the system during the period of schedule [143], therefore, $\min (T C T)$ can minimize the inventory level, and consequently, the power required for inventory handling.

Another factor affecting the energy costs is the energy-efficiency of the machines, therefore, developing energy-efficient machines has been the subjects of multiple studies [171, 172]. However, machine selection and layout design falls into a higher level of production management known as production planing, and requires a significant capital [144]. Thus, energy-efficient production scheduling is more practical and fits the purpose of this chapter.

## Labor cost

Labor wages are higher in the overtime hours [148, therefore, finishing the production in the regular hours results in a lower labor cost. $\min (M C T)$ can finish the production in a shorter time period, thus resultimg in a lower labor overtime cost.

Inventory handling requires staffed hours e.g. for transportation, security, etc., therefore, reducing the inventory level can result in a lower inventory handling cost. As it was mentioned earlier $\min (T C T)$ can reduce the inventory level, thus, resulting in a lower inventory handling labor cost.

## Inventory cost

Inventory costs refers to all costs of holding an inventory including opportunity cost of the inventory monetary value, infrastructure cost, insurance, depreciation cost, taxes, etc. [173]. As mentioned earlier, $\min (T C T)$ is directly associated with minimizing the inventory level, and consequently the inventory cost.

## Environment-oriented scheduling

Energy consumption and greenhouse gas emissions are the most studied environmental indicators in production scheduling [149]. As it discussed in 7.2 , energy consumption is a function of machines state, power peak, and processing speed [174]. Greenhouse gas emission has been studied by multiple works [175, 176, 144]. Since fossil fuels are the primary source of energy generation, managing energy consumption significantly reduces greenhouse gas emissions [144. Most studies calculated the greenhouse gas emissions as a function of energy consumption [149]. Therefore, all the discussion in 7.2 applies here as well.

## Social-oriented scheduling

Average completion time; $A C T=\frac{T C T}{N}$ also known as flow time is the average time that a job spends in the system. For a fixed number of jobs $(N)$, minimizing average completion time; $\min (A C T)$ is equivalent to $\min (T C T)$, and as discussed earlier has been subjected to an extensive study. Attractiveness of $\min (A C T)$ is that it is equivalent to minimizing waiting time, which is of importance for customers [92. Minimizing waiting time can help the society to have access to products/services with the minimum waiting time. Therefore, we conclude that $\min (T C T)$ can also serve social-oriented scheduling objectives.

Completion time variance; $C T V=\frac{\sum_{n=1}^{N}\left(C_{n}-A C T\right)^{2}}{N}$ is another performance metric which has been the subject of numerous studies in production scheduling [177, 89, 163, 94]. In service-oriented production systems such as hospital, call centers, etc. it is important to provide customers a uniform service (in terms of waiting time or completion time) [92, therefore, $\min (C T V)$ also can serve social-oriented scheduling objectives. $\min (C T V)$ and $\min (T C T)$ are possibly inconsistent objectives, and balancing trade-off between them is required.

## Summary

The metrics used for production scheduling to achieve different goals are summarized in Table 7.1 based on their impacts on different pillars of $T B L$. We deliberately kept these metrics generic, so that, practitioners may use/modify them according to different production environment such as flow shop, job shop, single/multiple machines, etc. MCT, TCT, and
$C T V$ were identified as the most fundamental metrics for production scheduling. As it is shown by Table 7.1 some of these metrics are simultaneously evaluated in multiple $T B L$ pillars.

Table 7.1: Metric-based hierarchical decomposition for sustainable production scheduling

| Cluster | Sub-cluster | Metric |
| :---: | :---: | :---: |
| Economic-oriented scheduling | Production cost | MCT |
|  |  | CTV |
|  | Energy cost | MCT |
|  |  | TCT |
|  | Labor cost | MCT |
|  |  | TCT |
|  | Inventory cost | TCT |
| Environmental-oriented scheduling | Energy consumption | MCT |
|  |  | $T C T$ |
|  | Greenhouse gas emissions | MCT |
|  |  | TCT |
| Social-oriented scheduling | Waiting time | TCT |
|  | Waiting time variance | CTV |

$T C T$ : Total completion time, $M C T$ : Maximum completion time, $C T V$ : Completion time variance

As mentioned earlier $\min (M C T), \min (T C T)$, and $\min (C T V)$ are inconsistent objectives which means optimizing one may result in worsening the others. Therefore, a production schedule that simultaneously optimize all objectives is infeasible.

### 7.3 Problem formulation

This section presents the mathematical formulations of the proposed models. In order to develop a comprehensive Production Scheduling Sustainability Index ( $P S S I$ ), we propose a top to bottom decomposition followed by a bottom to top aggregation scheme. At the decomposition phase, we divide $P S S I$ into three clusters covering the three pillars of the $T B L$ including economy, environment, and society. Each cluster is then divided into sub-clusters. Each sub-cluster covers a specific area of impact of its cluster. Accordingly, each sub-cluster is then divided to individual metrics that specifically measure a single performance indicator. Once the top-bottom structure is developed and all the individual metrics are measured, a bottom-up aggregation approach including normalization and weighting is utilized to calculate PSSI.

## Notations

$i \quad$ Index of clusters, $i \in\{1,2,3\}$
$J_{i} \quad$ Number of sub-clusters in cluster $i$
$j_{i} \quad$ Index of sub-clusters in cluster $i, j_{i} \in\left\{1,2, \ldots, J_{i}\right\}$
$k \quad$ Index of the individual metrics, $k \in\{1,2,3\}$ with $k=1$ for $T C T, k=2$ for $M C T$, and $k=3$ for $C T V$.
$h \quad$ Index of production scheduling alternatives, $h \in H=\{1,2, \ldots, H\}$
$x_{h}^{[k]} \quad$ The value of metric $k$ for production schedule $h$
$D_{h}^{[k]} \quad$ Normalized deviation of metric $k$ for production schedule $h$
$M_{h, i, j}^{[k]} \quad$ Sustainability score of $h$ for metric $k$ in sub-cluster $j$ of cluster $i$.
Let $i \in\{1,2,3\}$ denote the index of clusters; with 1 for economic, 2 for environmental, and 3 for social pillars of the the TBL. $J_{i}$ is the number of sub-clusters in each cluster, therefore, $j_{i} \in\left\{1,2, \ldots, J_{i}\right\}$ denotes sub-cluster $j$ in cluster $i$. Let $k \in\{1,2,3\}$ denote the index of the individual metrics, with $k=1$ for $T C T, k=2$ for $M C T$, and $k=3$ for $C T V$.

Given production schedule $h \in H=\{1,2, \ldots, H\}$, where $H$ is a set of alternative production schedules, $x_{h}^{[k]}$ is the value of metric $k$ for production schedule $h$. Because individual metric could be in different scales/units, we define the normalized deviation from the best possible value as $D_{h}^{[k]}=\frac{x_{h}^{[k]}-L B\left(x_{h}^{[k]}\right)}{U B\left(x_{h}^{k]}\right)-L B\left(x_{h}^{k]}\right)}$, where $U B\left(x_{h}^{[k]}\right)=\max _{h \in H}\left(x_{h}^{[k]}\right)$, and $L B\left(x_{h}^{[k]}\right)=\min _{h \in H}\left(x_{h}^{[k]}\right)$. In the minimization sense, $D_{h}^{[k]}$ is the degree of closeness between $x_{h}^{[k]}$ and its best value $L B\left(x_{h}^{[k]}\right)$.

At the metric level, we use $M_{h, i, j}^{[k]}=10\left(1-D_{h}^{[k]}\right)$ to calculate the sustainability score of production schedule $h$ for metric $k . M_{h, i, j}^{[k]} \in[0,10]$ normalizes each metric to a scale of 0 to 10 , where 0 is the worst performance and 10 is the best performance in terms of sustainability. $M_{h, i, j}^{[k]}$ attributes the sustainability of an individual metric to its normalized deviation. For example, as shown in Table 7.1 waiting time variance is directly affected by $C T V$; therefore, a production schedule with $L B\left(x_{h}^{[3]}\right)$ generates the highest sustainability score for the waiting time variance.

Once the top-bottom structure is developed and all the individual metrics are measured, a bottom-up aggregation approach by equations 7.3 to 7.8 is utilized to calculate PSSI. Equation 7.3 is the aggregation of individual metrics sustainability score to the Sub-cluster Sustainability Score $(S S S)$, where $\omega_{k, i, j} \in[0,1]$ is the weight assigned to the metric $k$ of sub-cluster $j$ in cluster $i$. Equation 7.4 imposes that the sum of all weights must be equal 1 . Equation 7.5 is the aggregation of sub-cluster sustainability scores to the Cluster Sustainability Score (CSS), where $\omega_{i, j}$ is the weight of sub-cluster $j$ in cluster $i$. Equation 7.6 indicates that the sum of sub-cluster weights must be 1. Equation 7.7 aggregates cluster sustainability scores into the Production Scheduling Sustainability Index PSSI, where, $\omega_{i}$ is the weight of cluster $i$. Equation 7.8 indicates that the sum of cluster weights is equal to 1 .

$$
\begin{gather*}
S S S_{h, i, j}=\sum_{k=1}^{K} \omega_{i, j, k} M_{h, i, j}^{[k]}  \tag{7.3}\\
\sum_{k=1}^{K} \omega_{i, j, k}=1  \tag{7.4}\\
C S S_{h, i}=\sum_{j, 1}^{J_{i}} \omega_{i, j} S S S_{h, i, j}  \tag{7.5}\\
\sum_{j=1}^{J_{i}} \omega_{i, j}=1  \tag{7.6}\\
P S S I_{h}=\sum_{i=1}^{3} \omega_{i} C S S_{h, i}  \tag{7.7}\\
\sum_{i=1}^{3} \omega_{i}=1 \tag{7.8}
\end{gather*}
$$

### 7.4 Case study

The PSSI model is demonstrated here by evaluating the sustainability performance of different scheduling methods in a permutation flow shop. Flow shop scheduling problem arises where a set of jobs on one or multiple machines must be sequenced in order to optimize a
given objective function. Permutation flow shop is a special type of flow shop in which the processing order of jobs is identical on all machines. Permutation flow shop has been the subject of a massive body of literature [143].

## Notations

| $N$ | Number of jobs |
| :--- | :--- |
| $n$ | Job index, $n=1,2, \ldots, N$ |
| $M$ | Number of machines |
| $m$ | Machine index, $m=1,2, \ldots, M$ |
| $\Psi$ | Decision space, $\|\Psi\|=N!$ |
| $\pi$ | $\pi \in \Psi$, a permutation of $N$ jobs |
| $M C T$ | Maximum completion time |
| $T C T$ | Total completion time |
| $A C T$ | Average completion time |
| $C T V$ | Completion time variance |

The following formulations provide the mathematical descriptions of $M C T, T C T$ and $C T V$ in an $N$-job, $M$-machine permutation flow shop. $C_{n, m}$ is the completion time of job $n$ on machine $m . M C T$ is the completion time of the last job on the last machine and is shown by equation 7.9. $T C T$ is the sum of completion times of all jobs on the last machine and is presented by equation 7.10. Completion times variance is calculated by equation 7.12 ,

$$
\begin{gather*}
M C T=C_{N, M}  \tag{7.9}\\
T C T=\sum_{n=1}^{N} C_{n, M}  \tag{7.10}\\
A C T=T C T / N  \tag{7.11}\\
C T V=\frac{1}{N} \sum_{n=1}^{N}\left(C_{n, M}-A C T\right)^{2} \tag{7.12}
\end{gather*}
$$

Given an instance with $N$ jobs, there are $N$ ! different possible sequences. Let $\pi \in$ $\{1,2, \ldots, N!\}$ denote a sequence of $N$ jobs, as mentioned earlier, the normalized deviation of metric $k$ generated by sequence $\pi$ can be defined as $D_{\pi}^{[k]}=\frac{x_{\pi}^{[k]}-\mathrm{LB}\left(x^{[k]}\right)}{\mathrm{UB}\left(x^{[k]}\right)-\mathrm{LB}\left(x^{[k]}\right)}$, with $k=1$ for $\min (T C T), k=2$ for $\min (M C T), k=3$ for $\min (C T V)$, and $x_{\pi}^{[k]}$ is the performance of sequence $\pi$ on metric $k$. $\mathrm{UB}\left(x^{[k]}\right)$ and $\mathrm{LB}\left(x^{[k]}\right)$ are the worst (maximum) and the best (minimum) values for metric $k$ in the instance, respectively. Let $\omega_{k}$ be as the decision maker's preference to metric $k$ and $\Omega=\left[\omega_{1}, \omega_{2}, \omega_{3}\right]$, we propose a trade-off function represented by equation 7.13 .

$$
\begin{gather*}
z_{\pi}^{[\Omega]}=\sum_{k=1}^{3} \omega_{k} D_{\pi}^{[k]}  \tag{7.13}\\
\sum_{k=1}^{3} \omega_{k}=1,
\end{gather*}
$$

Equation 7.13 explicitly integrates the decision maker's preference into the trade-off function by assigning a weight $\left(\omega_{k}\right)$ to each metric. With dynamics in production, decision makers'
preferences might change as the process reveals its performance over the time. $\min (z)$ is precisely equivalent to minimizing the deviations from an ideal but infeasible point at which all the objectives are at their optimum values.

To show the inconsistencies among objectives of $\min (T C T), \min (M C T)$, and $\min (C T V)$, and to verify the effectiveness of balancing trade-offs among inconsistent objectives, we carry out a series of case studies. The number of jobs $N$ changes from $N=5, \ldots, 10$, resulting in six choices, number of machines $M$ changes from 3 to $19(M=2 l+1, l=1,2, \ldots, 9)$, yielding nine choices. This configuration results in 54 combinations. For each combination, 100 instances are randomly generated. The processing times are randomly drawn from a uniform distribution in $[1,99]$. Therefore, in total we have 5400 instances. Given $\omega_{k}$ changing from [0.0: $0.1: 1.0]$, we have 66 combinations of three weights with $\sum_{k=1}^{3} \omega_{k}=1$, i.e. $\Omega \in$ $\left\{\Omega_{1}, \Omega_{2}, \ldots, \Omega_{66}\right\}$. The 66 minimization functions of $\min \left(z_{\pi}^{[\Omega]}\right)$ cover the three single-objective minimizations of $\min (T C T), \min (M C T)$ and $\min (C T V)$ with a weight equal to $[1,0,0],[0$, $1,0]$, and $[0,0,1]$, respectively. Since the number of jobs is relatively small $(N \leq 10)$, we are able to use enumeration/optimization to find $U B\left(x^{[k]}\right)$ and $L B\left(x^{[k]}\right)$ for all $k$ as well as $z_{\pi}^{[\Omega]}$ for each weight.

## Inconsistencies among objectives

Figure 7.1 clearly shows that single optimization of $\min (T C T), \min (M C T)$, and $\min (C T V)$ are inconsistent with each other, since there is no single point that yields the best value for all three objectives. In order to statistically confirm the inconsistency among $\min (T C T)$, $\min (M C T)$, and $\min (C T V)$, we perform Spearman's rank correlation analyses for the sequences generating $\min (T C T), \min (M C T)$, and $\min (C T V)$. Table 7.2 shows the Spearman's rank correlation coefficient ( $\rho$ ) between sequences. Small values of ( $\rho$ ) confirms that the sequences generating minimum values for single objectives are not significantly correlated.

Table 7.2: Spearman's $\rho$ among sequence for scheduling objectives

|  | $T C T$ vs. $M C T$ | $T C T$ vs. $C T V$ | $M C T$ vs. $C T V$ |
| :--- | :---: | :---: | :---: |
| Spearman's $\rho$ | 0.0823 | 0.2336 | 0.0528 |

## Pareto dominance

In the case of multi-objective optimization with inconsistent objectives, Pareto dominance is useful for decision making [178, 114]. For minimization problems, if $x_{A}^{[k]}$ and $x_{B}^{[k]} \in \mathbb{R}^{K}$ are two vectors that measure a positive attribute $k$ such as the utility of decision $A$ and $B$, respectively, decision $A$ dominates decision $B$ if the following conditions are satisfied:

$$
\begin{array}{lll}
x_{A}^{[k]} \leq x_{B}^{[k]}, & \forall & k \in\{1,2, . ., K\} \\
x_{A}^{[k]}<x_{B}^{[k]}, & \exists & k \in\{1,2, . ., K\} \tag{7.15}
\end{array}
$$

Equation 7.14 states that decision $A$ is not worse than decision $B$ in any dimension, while equation 7.15 states that decision $A$ is better than decision $B$ at least in one dimension. Pareto optimal outcome cannot be improved without sacrificing of at least one objective. Pareto dominant solutions are shown in figures 7.1 by red markers, each cross shows the data point obtained from one weight. It is observed that there is no Pareto-optimal solution when all three objectives are taken into consideration.


Figure 7.1: Scatter plots of normalized deviations of 66 weights, Pareto-dominant solution are shown by red markers

## Production scheduling sustainability

In order to demonstrate the efficiency and effectiveness of trade-off balancing to achieving a sustainable production schedule, we perform the scheduling with respect to 66 trade-off functions as described earlier. As shown by table 7.3, we consider a special case for calculating $P S S I$ in which equal weights are assigned to all metrics in each sub-cluster, and equal weights for clusters as well. The equal weights assignment means that all aspects of $T B L$ have the same importance for the decision maker.

Table 7.3: PSSI weight assignment

|  | Weight $\left(\omega_{i}\right)$ | Cluster | Weight $\left(\omega_{i, j}\right)$ | Sub-cluster | Weight $\left(\omega_{i, j, k}\right)$ | Metric |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| PSSI | 0.333 | Economic-oriented scheduling | 0.25 | Production cost | 0.5 | MCT |
|  |  |  |  |  | 0.5 | CTV |
|  |  |  | 0.25 | Energy cost | 0.5 | MCT |
|  |  |  |  |  | 0.5 | TCT |
|  |  |  | 0.25 | Labor cost | 0.5 | MCT |
|  |  |  |  |  | 0.5 | TCT |
|  |  |  | 0.25 | Inventory cost | 1 | TCT |
|  | 0.333 | Environmental-oriented scheduling | 0.5 | Energy consumption | 0.5 | MCT |
|  |  |  |  |  | 0.5 | TCT |
|  |  |  | 0.5 | Greenhouse gas | 0.5 | MCT |
|  |  |  |  | emission | 0.5 | TCT |
|  | 0.333 | Social-oriented scheduling | 0.5 | Waiting time | 1 | TCT |
|  |  |  | 0.5 | Waiting time variance | 1 | CTV |

## Results and discussion

In this subsection, we statistically present the results of our case study in terms of production sustainability indices, sustainability scores at sub-cluster level, and fundamental performance metrics of $T C T, M C T$, and $C T V$. Using equal weights presented by Table 7.3, we can rewrite equations 7.3 to 7.8 as follows:

$$
\begin{gather*}
P S S I_{\pi}=10\left[0.5\left(1-D_{\pi}^{[1]}\right)+0.292\left(1-D_{\pi}^{[2]}\right)+0.208\left(1-D_{\pi}^{[3]}\right)\right] \Rightarrow \\
P S S I_{\pi}=10[1-\underbrace{\left(0.5 D_{\pi}^{[1]}+0.292 D_{\pi}^{[2]}+0.208 D_{\pi}^{[3]}\right)}_{=z_{\pi}^{[0.5,0.292,0.208]} \approx z_{\pi}^{[0.5,0.3,0.2]}}] \tag{7.16}
\end{gather*}
$$

As it is shown by equation 7.16 production scheduling with maximizing $P S S I$ i.e. $\max _{\pi \in \Psi}(P S S I)$ as the objective function is equivalent to $\min _{\pi \in \Psi}\left(z_{\pi}^{[0.5,0.3,0.2]}\right)$. Therefore, depending on the weight assignment for $P S S I$, scheduling for sustainable production always leads to a trade-off function among the fundamental performance metrics of $T C T, M C T$, and $C T V$. It is worth mentioning that the weight assignments may change over time as the production system reveals its nature or due to production/market needs, therefore, a different weight for the trade-off function will be used. The contribution of the $P S S I$ model is twofold: (i) At a higher level, $P S S I$ evaluates the sustainability of the production schedules in terms a quantifiable value (ii) At the lower level, $P S S I$ provides insight to the system performance in terms of the fundamental performance metrics. These two contributions together provide a comprehensive control over the production performance.

Table 7.4 shows the statistics of sustainability index for the 66 studied trade-off functions (i.e. $\min \left(z^{[\Omega]}\right)$ ). It is observed that single objective objective functions of $\min (T C T)$, $\min (M C T)$, and $\min (C T V)$ (i.e. $\min \left(z^{[1,0,0]}\right), \min \left(z^{[0,1,0]}\right)$, and $\min \left(z^{[0,0,1]}\right)$, respectively) are outperformed by $\min \left(z^{[0.5,0.3,0.2]}\right)$ in terms of sustainability index. $\min \left(z^{[0.5,0.3,0.2]}\right)$ has not only the highest average sustainability index of 9.06 but also the minimum standard deviation of 0.48 . In terms of the fundamental performance metrics as it is shown by Table 7.5 single objective functions generate no deviation on their intended metric but very large deviations on the others. On the other hand $\min \left(z^{[0.5,0.3,0.2]}\right)$ generate small and uniform deviations on all three metrics which provides a better control over the system performance, which means a stable compromise among the inconsistent objectives is achieved.

In order to further discuss the performance of production scheduling objective functions, we narrow our discussion to four production scheduling alternatives of $\min \left(z^{[1,0,0]}\right)$, $\min \left(z^{[0,1,0]}\right), \min \left(z^{[0,0,1]}\right), \min \left(z^{[0.5,0.3,0.2]}\right)$ which hereafter we name $\min (T C T), \min (M C T)$, $\min (C T V)$, and $\min (T O)$, respectively. It is worth reminding that $\min \left(z^{[0.5,0.3,0.2]}\right)$ is equivalent to maximizing $P S S I$ i.e. $\max (P S S I)$.

## Production scheduling Sustainability

In this subsection, we compare the performance of the aforementioned production scheduling alternatives including $\min (T C T), \min (M C T), \min (C T V)$, and $\min (T O)$ in terms of sustainability index $(P S S I)$ and the fundamental performance metrics of $T C T, M C T$, and $C T V$.

Table 7.6 shows the $P S S I$ for the studied production scheduling alternatives. min $(z)$ has the greatest average and also has the smallest standard deviation. $t$-tests revealed that the performance of $\min (z)$ is significantly different compared to the other production scheduling alternatives.

Table 7.4: PSSI statistics of trade-off functions $\left(\min \left(z^{[\Omega]}\right)\right)$

| $\overline{P S S I}^{\Omega}$ | $(1,0,0)$ | (0,1,0) | (0,0,1) | (0.9,0.1,0) | (0.8,0.2,0) | (0.7,0.3,0) | (0.6,0.4,0) | (0.5,0.5,0) | (0.4,0.6,0) | (0.3,0.7,0) | (0.2,0.8,0) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Ave | 8.52 | 8.69 | 7.04 | 8.72 | 8.87 | 8.96 | 9.00 | 9.01 | 9.00 | 8.97 | 8.93 |
| Std | 0.74 | 0.67 | 1.26 | 0.68 | 0.62 | 0.57 | 0.52 | 0.51 | 0.52 | 0.54 | 0.57 |
| Max | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Min | 5.36 | 5.44 | 2.61 | 5.58 | 5.67 | 5.86 | 6.21 | 6.48 | 6.23 | 5.70 | 5.64 |
| $\underbrace{\Omega S S I}$ | (0.1,0.9,0) | (0.9,0,0.1) | (0.8,0,0.2) | (0,7, $0,0.3$ ) | (0.6,0,0.4) | (0.5,0,0.5) | (0.4,0,0.6) | (0.3,0,0.7) | (0.2,0,0.8) | (0.1,0,0.9) | $(0,0.1,0.9)$ |
| Ave | 8.87 | 8.63 | 8.74 | 8.83 | 8.87 | 8.80 | 8.61 | 8.29 | 7.88 | 7.46 | 7.30 |
| Std | 0.61 | 0.71 | 0.68 | 0.64 | 0.62 | 0.68 | 0.81 | 1.00 | 1.18 | 1.26 | 1.24 |
| Max | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Min | 5.64 | 5.57 | 5.57 | 5.70 | 5.50 | 5.03 | 4.78 | 3.86 | 3.10 | 2.67 | 2.79 |
| $\operatorname{PSSI}^{\Omega}$ | (0,0.2,0.8) | (0,0.3,0.7) | (0,0.4, 0.6 ) | (0,0.5,0.5) | (0,0.6,0.4) | (0,0.7, 0.3 ) | (0,0.8,0.2) | (0,0.9,0.1) | (0.1,0.1,0.8) | (0.1,0.2,0.7) | (0.1,0.3,0.6) |
| Ave | 7.55 | 7.79 | 8.00 | 8.18 | 8.33 | 8.46 | 8.57 | 8.64 | 7.73 | 7.97 | 8.18 |
| Std | 1.21 | 1.15 | 1.07 | 0.98 | 0.91 | 0.83 | 0.76 | 0.71 | 1.21 | 1.12 | 1.03 |
| Max | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Min | 3.10 | 3.57 | 3.86 | 4.46 | 4.66 | 4.84 | 5.21 | 5.25 | 3.10 | 3.86 | 3.86 |
| $\operatorname{PSSI}^{\Omega}$ | (0.1,0.4,0.5) | (0.1,0.5,0.4) | (0.1,0.6,0.3) | (0.1,0.7,0.2) | (0.1,0.8,0.1) | (0.2,0.1,0.7) | (0.2,0.2,0.6) | (0.2,0.3,0.5) | (0.2,0.4,0.4) | (0.2,0.5,0.3) | (0.2,0.6,0.2) |
| Ave | 8.36 | 8.51 | 8.63 | 8.73 | 8.82 | 8.15 | 8.35 | 8.53 | 8.67 | 8.78 | 8.78 |
| Std | 0.93 | 0.85 | 0.76 | 0.70 | 0.64 | 1.07 | 0.96 | 0.87 | 0.78 | 0.69 | 0.63 |
| Max | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Min | 4.66 | 4.69 | 5.27 | 5.49 | 5.64 | 3.86 | 3.86 | 4.66 | 4.86 | 5.46 | 5.49 |
| $\underbrace{\text { PSSI }}$ | (0.2,0.7,0.1) | (0.3,0.1,0.6) | (0.3,0.2,0.5) | (0.3,0.3,0.4) | (0.3,0.4,0.3) | (0.4,0.3,0.3) | (0.3,0.5,0.2) | (0.3,0.6,0.1) | (0.4,0.1,0.5) | (0.4,0.2,0.4) | (0.4,0.4,0.2) |
| Ave | 8.92 | 8.52 | 8.67 | 8.81 | 8.91 | 9.00 | 8.97 | 8.98 | 8.80 | 8.92 | 9.03 |
| Std | 0.59 | 0.88 | 0.79 | 0.69 | 0.62 | 0.54 | 0.56 | 0.54 | 0.70 | 0.61 | 0.51 |
| Max | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Min | 5.64 | 4.58 | 4.69 | 5.32 | 5.49 | 5.87 | 5.64 | 5.64 | 4.91 | 5.55 | 6.04 |
| $\underbrace{\operatorname{PSSI}}{ }^{\Omega}$ | (0.4,0.5,0.1) | (0.5,0.1,0,4) | (0.5,0.2,0.3) | (0.5,0.3,0.2) | (0.5,0.4,0.1) | (0.6,0.1,0.3) | (0.6,0.2,0.2) | (0.6,0.3,0.1) | (0.7,0.1,0.2) | (0.7,0.2,0.1) | (0.8,0.1, 0.1$)$ |
| Ave | 9.02 | 8.96 | 9.04 | 9.06 | 9.04 | 8.98 | 9.02 | 9.02 | 8.92 | 8.95 | 8.83 |
| Std | 0.51 | 0.56 | 0.50 | 0.48 | 0.49 | 0.55 | 0.51 | 0.51 | 0.60 | 0.57 | 0.65 |
| Max | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Min | 6.23 | 5.95 | 6.46 | 6.65 | 6.52 | 5.70 | 6.04 | 5.86 | 5.67 | 5.67 | 5.67 |

Figur 7.2 shows the Pareto frontiers of sustainability index mean and standard deviation for the studied production scheduling alternatives where $\min (T O)$ dominates the others.


Figure 7.2: PSSI Pareto frontier

Process capability indices, $C_{p}$ and $C_{p k}$, are used to further compare the sustainability of different production scheduling alternatives. Process capability index is the measure of

Table 7.5: Normalized deviation statistics of trade-off functions $\left(\min \left(z^{[\Omega]}\right)\right)$

| $\Omega$ | $(1,0,0)$ |  |  | (0,1,0) |  |  | $(0,0,1)$ |  |  | (0.9,0.1,0) |  |  | (0.8,0.2,0) |  |  | (0.7,0.3,0) |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $D^{[k]}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $D^{\text {I }}$ | $D^{[2]}$ | $D^{\text {\| }}$ | $D^{[1]}$ | $D^{[2]}$ | D | $D^{\text {I }}$ | $D^{[2]}$ | D | D | $D^{[2]}$ | D |
| Ave | 0. | 0.245 | 0. | 87 | 000 | . 180 | 489 | 0.176 | 0.000 | 0.003 | 0.185 | 0.34 | 0.012 | 0.133 | 0.32 | 0.02 | 092 | 0.310 |
| St | 0.000 | 0.164 | 0.165 | 0.1 | 0.000 | 0.13 | 0.194 | 0.135 | 0.0 | 0.007 | 0.150 | 0.163 | 0.019 | 0.129 | 0.163 | 0.033 | 0.10 | 0.160 |
| Max | 0.000 | 1.000 | 0.932 | 0.870 | 0.000 | 0.751 | 0.985 | 1.000 | 0.000 | 0.080 | 0.982 | 0.932 | 0.171 | 0.982 | 0.925 | 0.345 | 0.874 | 8 |
| Min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $\Omega$ | (0.6,0.4,0) |  |  | (0.5,0.5,0) |  |  | (0.4,0.6,0)) |  |  | (0.3,0.7,0) |  |  | (0.2,0.8,0) |  |  | (0.1,0.9,0) |  |  |
| D |  | $D^{[2]}$ | $D^{[3]}$ |  | $D^{[2]}$ | $D^{[3]}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{1}$ | ${ }^{[1]}$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $D^{\text {[1] }}$ | $D^{[2]}$ | $D^{[3]}$ |
| Ave | 43 | 0.060 | 0.292 | 0.062 | 0.03 | 0.27 | 0.080 | 022 | 0.25 | 0.100 | 0.01 | 0.2 | 0.119 | 0.00 | 0.2 | 0.139 | 0.001 | 0.207 |
| Std | 0.049 | 0.074 | 0.157 | 0.063 | 0.052 | 0.156 | 0.077 | . 035 | . 15 | 0.01 | 0. 021 | 0.14 | 0.104 | 0.01 | 0.14 | 0.116 | 0.004 | 0.144 |
| Max | 0.514 | 0.628 | 0.938 | 0.580 | 0.443 | 0.938 | 0.652 | 0.416 | 0.93 | . 858 | 0.219 | 0.87 | 0.868 | 0.127 | 0.87 | 0.86 | 0.054 | 0 |
| Min | 0.000 | 0.000 | 0.0 | 00 | 0.000 | 0.000 | 00 | 0.000 | 0.00 | 000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0 |
| $\Omega$ | (0.9, $0,0.1$ ) |  |  | (0.8,0,0.2) |  |  | (0.7,0,0.3) |  |  | (0.6,0,0.4) |  |  | (0.5,0,0.5) |  |  | (0.4,0,0.6) |  |  |
| $D^{\text {k] }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{[3]}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $\mathrm{D}^{1]}$ | $D^{[2]}$ | $D^{\text {B }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{\text {3 }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{[3]}$ |
| Ave | 0.00 | 0.221 | 0.34 | 0.00 | 0.19 | 0.313 | 0.020 | . 172 | 0.2 | 0.052 | 0.14 | 0.21 | 0.106 | 0.123 | 0.1 | 0.178 | 0.107 | 0.092 |
| Std | 0.0 | 0.159 | 0.1 | 0. | 0.153 | 0.1 | 0.031 | 0.147 | 0.15 | .063 | 0.135 | 0.1 | 0.103 | 0.124 | 0.11 | 0.145 | 0.105 | 84 |
| Max | 0.052 | 0.982 | 0.930 | 0.177 | 0.920 | 0.930 | . 248 | 0.920 | 0.910 | 0.445 | 0.920 | 0.88 | 0.711 | 0.920 | 0.71 | 0.85 | 0.853 | . 529 |
| Min | 0.000 | 0.000 | 0.0 | 00 | 0.000 | 0.0 | 00 | 0.000 | 0.0 | 00 | 0.000 | 0.0 | 000 | 0.000 | 0.00 | 0 | 0.000 | 0 |
| $\Omega$ | (0.3,0,0.7) |  |  | (0.2,0,0.8) |  |  | (0.1,0,0.9) |  |  | (0,0.1,0.9) |  |  | $(0,0.2,0.8)$ |  |  | (0,0.3,0.7) |  |  |
| $D^{\text {k] }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{3}$ | $D^{[1]}$ | $D^{[2]}$ | D | $D^{[1]}$ | $D^{[2]}$ | D | ${ }^{11}$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | D | $D^{\text {[1] }}$ | $D^{12]}$ |  |
| Ave | 0.26 | 0.10 | 0.046 | 0.348 | 0.119 | 0.017 | 0.423 | 0.145 | 0.004 | 0.458 | . 1 | 0.002 | 0.424 | 0.10 | 0.0 | 0.389 | 0.080 | 0.016 |
| Std | 0.179 | 0.094 | 0.05 | 0. | 0.105 | . 027 | 0.201 | . 12 | 0.009 | 0.196 | 0.121 | 0.0 | 0.199 | 0.105 | 0.01 | 0.19 | . 08 | 28 |
| Max | 0.927 | 0.769 | 0.342 | 0.985 | 0.857 | 0.204 | 0.985 | 1.000 | 0.090 | 0.985 | 0.926 | 0.06 | 0.985 | 0.857 | 0.145 | 0.985 | 0.769 | 53 |
| Min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 00 | 0.000 | 0.00 | 00 | 0.000 | 0.0 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.000 |
| $\Omega$ | (0,0.4,0.6) |  |  | $(0,0.5,0.5)$ |  |  | (0,0.6,0.4) |  |  | (0,0.7,0.3) |  |  | (0,0.8,0.2) |  |  | (0,0.9,0.1) |  |  |
| $D^{[k]}$ | $D$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | $D$ | ${ }^{[1]}$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | $D$ | $D^{\text {II }}$ | $D^{[2]}$ | $D^{[3]}$ |
| Ave | 0.35 | 0.057 | 0.029 | 0.323 | 0.03 | 0.044 | 0.294 | 02 | 0.06 | 0.266 | . 01 | 0.082 | 0.238 | 0.00 | 0.1 | 0.216 | 0.001 | 0.131 |
| Std | 0.192 | 0.07 | 0.04 | 0. | 0.05 | 0.057 | 0.175 | . 04 | 0.070 | 16 | 0.02 | 0.08 | 0.156 | 0.01 | 0.10 | 0.1 | . 0.00 | 6 |
| Max | 0.985 | 0.548 | 0.486 | 0.951 | 0.399 | 0.51 | 0.951 | 0.399 | 0.55 | 0.951 | 0.23 | 0.75 | 0.951 | 0.132 | 0.75 | 0.95 | 0.060 | . 751 |
| Min | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.0 | . 000 | 0.000 | 0.00 | 000 | 0.000 | 0.0 | . 000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.000 |
| $\Omega$ | (0.1,0.1,0.8) |  |  | (0.1,0.2,0.7) |  |  | (0.1,0.3,0.6) |  |  | (0.1,0.4,0.5) |  |  | (0.1,0.5,0.4) |  |  | (0.1,0.6,0.3) |  |  |
| $D^{\text {k] }}$ | $D^{[1]}$ | $D^{[2]}$ | $D$ |  | $D^{[2]}$ | $D$ | [1] | $D^{[2]}$ | $D^{1}$ | ${ }^{\text {[1] }}$ | $D^{[2]}$ | D | $D^{[1]}$ | $D^{[2]}$ | D | $D^{[1]}$ | $D^{[2]}$ | $D^{[3]}$ |
| Ave | 0.389 | 0.10 | 0.01 | 0.3 | 0.07 | 0.02 | 0. | 0.052 | 0.03 | 0.28 | 0.033 | 0.0 | 0.25 | 0.020 | 0.0 | 0.229 | 0.01 | 0.09 |
| Std | 0.201 | 0.104 | 0.01 | 0.195 | 0.08 | 0.0 | 0.18 | 06 | 0.04 | 0.176 | 0.04 | 0.0 | 0.167 | 0.03 | 0.0 | 0.1 | . 02 | 4 |
| Max | 0.985 | 0.85 | 0.14 | 0.9 | 0.7 | 0.29 | 0.985 | . 548 | 0.4 | 0.935 | 0.399 | 0.5 | 0.924 | 0.36 | 0.61 | 0.878 | 0.18 | 1 |
| Min | 0.000 | 0.000 | 0.00 | 00 | 0.000 | 0.00 | 00 | 0.000 | 0.0 | 000 | 0.000 | 0.0 | 000 | 0.000 | 0.0 | . 00 | 0.000 | 0.000 |
| $\Omega$ | (0.1,0.7,0.2) |  |  | (0.1,0.8,0.1) |  |  | (0.2,0.1,0.7) |  |  | (0.2,0.2,0.6) |  |  | (0.2,0.3,0.5) |  |  | (0.2,0.4,0.4) |  |  |
| D | $D^{[1]}$ | $D^{[2]}$ | $D^{[3]}$ | $D^{[1]}$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | $D^{[3]}$ | $D^{[1]}$ | $D^{[2]}$ | D | $D^{[1]}$ | $D^{2 / 1}$ | $D^{[3]}$ |
|  | 0. | 0.003 | 0. | 0.170 | 0.001 | 0.15 |  | 0.079 | 0.02 | 0.281 | 0.051 | 0.04 | 0.249 | 0.03 | 0.066 | 0.219 | 0.018 | 㖪 |
| Std | 0.144 | 0.00 | 0.11 | 0.130 | 0.0 | 0.1 | 0.191 | 08 | 0.0 | 0.179 | 0.06 | 0.0 | 0.169 | 0.044 | 0.07 | 0.15 | 0.030 | 87 |
| Max | 0.878 | 0.085 | 0.75 | 0.868 | 0.042 | 0.751 | 0.985 | 0.769 | 0.29 | 0.935 | 0.548 | 0.4 | 0.927 | 0.399 | 0.518 | 0.875 | 0.364 | 0.642 |
| Min | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.00 | 000 | 0.000 | 0.00 | 000 | 0.000 | 0.0 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.000 |
| $\Omega$ | (0.2,0.5,0.3) |  |  | (0.2,0.6,0.2) |  |  | (0.2,0.7,0.1) |  |  | (0.3,0.1,0.6) |  |  | (0.3,0.2,0.5) |  |  | (0.3,0.3,0.4) |  |  |
| $D^{\text {k] }}$ | $D$ | $D^{[2]}$ | D | ${ }^{1 /}$ | $D^{[2]}$ | $D^{\text {l }}$ | ${ }^{\text {[1] }}$ | $D^{[2]}$ | D | $D^{[1]}$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | $D$ | $D^{[1]}$ | $D^{[2]}$ | ${ }^{[3]}$ |
| Ave | 0.189 | 0.009 | 0. | 0.162 | 0.004 | 0.1 | 0.133 | 0.003 | 0.19 | 23 | 06 | 0.06 | 0.205 | 0.03 | 0.08 | 0.177 | 0.023 | 0.111 |
| Std | 0.143 | 0.017 | 0.104 | 0.129 | 0.010 | 0.119 | 0.11 | 0.008 | 0.136 | 0.167 | 0.070 | 0.06 | 0.157 | 0.051 | 0.08 | 0.14 | 0.036 | 0.098 |
| Max | 0.875 | 0.169 | 0.751 | 0.875 | 0.091 | 0.751 | 0.868 | 0.112 | 0.789 | . 927 | 480 | 0.518 | 891 | 0.4 | 0.54 | 0.87 | 0.360 | . 642 |
| Min | 0.000 | 0.000 | 0.00 | 000 | 0.000 | 0.00 | . 000 | 0.000 | 0.00 | 000 | 0.000 | 0.00 | . 000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.000 |
| $\Omega$ | (0.3,0.4,0.3) |  |  | (0.4,0.3,0.3) |  |  | (0.3,0.5,0.2) |  |  | (0.3,0.6,0.1) |  |  | (0.4,0.1,0.5) |  |  | (0.4,0.2,0.4) |  |  |
| $D^{\text {k] }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | ${ }^{[1]}$ | $D^{[2]}$ | $D^{\mid 3}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{\text {B }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{1}$ | $D^{\text {[1] }}$ | $D^{[2]}$ | $D^{13}$ | $D^{11}$ | $D^{[2]}$ | $D^{[3]}$ |
| Ave | 0.15 | 0.014 | 0.14 | 0.112 | 0.027 | 0.1 | 0. | 010 | 0.1 | 0.107 | 0.010 | 0.21 | 0.156 | 0.0 | 0.11 | 0.1 | 0.041 | 析 |
| Std | 0.127 | 0.024 | 0.11 | 0.105 | 0.04 | 0.1 | 0.110 | 0.019 | 0.13 | . 099 | 0.018 | 0.1 | 0.135 | 0.079 | 0.095 | 0.121 | 0.058 | 0.109 |
| Max | 0.875 | 0.223 | 0.751 | 0.826 | 0.482 | 0.751 | 0.868 | 0.220 | 0.75 | 0.868 | 0.220 | 0.87 | 0.839 | 0.682 | 0.612 | 0.860 | 0.548 | 0.642 |
| Min | 0.000 | 0.000 | 0.00 | . 000 | 0.000 | 0.0 | . 000 | 0.000 | 0.00 | . 000 | 0.000 | 0.0 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.0 |
| $\Omega$ | (0.4,0.4,0.2) |  |  | (0.4,0.5,0.1) |  |  | (0.5, 0.1, 0.4$)$ |  |  | (0.5, $0.2,0.3$ ) |  |  | (0.5,0.3,0.2) |  |  | (0.5,0.4, 0.1) |  |  |
| $D$ | $D^{[1]}$ | $D^{[2]}$ | D | $D^{\text {[] }}$ | $D^{[2]}$ | $D^{[3]}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{[3]}$ |
| Ave | 0.094 | 0.022 | 0. | 0.085 | 021 | 0.23 |  | . 08 | 0.17 | . 073 | 0.057 | 0.20 | 0.065 | 0.045 | 0.23 | 0.062 | 0.03 | 0.257 |
| Std | 0.092 | 0.037 | 0.142 | 0.082 | 0.034 | 0.150 | 0.090 | 0.100 | 0.125 | 0.077 | 0.079 | 0.13 | 0.070 | 0.064 | 0.149 | 0.065 | 0.056 | 0.154 |
| Max | 0.741 | 0.482 | 0.879 | 0.654 | 0.416 | 0.938 | 0.740 | 0.874 | 0.75 | 0.654 | 0.682 | 0.879 | 0.654 | 0.653 | 0.93 | 0.58 | 0.482 | 0.938 |
| Min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.00 | 0.000 | 0.000 | 0.000 |
| $\Omega$ | (0.6,0.1,0.3) |  |  | (0.6,0.2,0.2) |  |  | (0.6,0.3,0.1) |  |  | (0.7,0.1,0.2) |  |  | (0.7,0.2,0.1) |  |  | (0.8,0.1,0.1) |  |  |
| $D^{\text {k] }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{[3]}$ | $D^{\text {[1] }}$ | $D^{[2]}$ | $D^{[3]}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{\text {[3] }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{\text {[3] }}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{13}$ | $D^{[1]}$ | $D^{[2]}$ | $D^{[3]}$ |
| Ave | 0.040 | 0.109 | 0.241 | 0.038 | 0.083 | 0.262 | 0.039 | 0.068 | 0.279 | 0.018 | 0.134 | 0.287 | 0.021 | 0.109 | 0.300 | 0.008 | 0.160 | 0.321 |
| Std | 0.051 | 0.120 | 0.149 | 0.047 | 0.100 | 0.155 | 0.047 | 0.086 | 0.157 | 0.027 | 0.132 | 0.159 | 0.029 | 0.118 | 0.160 | 0.014 | 0.143 | 0.162 |
| Max | 0.431 | 0.920 | 0.880 | 0.526 | 0.874 | 0.893 | 0.386 | 0.874 | 0.938 | 0.218 | 0.920 | 0.910 | 0.261 | 0.920 | 0.925 | 0.127 | 0.920 | 0.910 |
| Min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 7.6: PSSI for alternative production schedules

|  | $P S S I($ Ave $)$ | $P S S I(\mathrm{Std})$ | $p$-value, $t$-test <br> against $\min (T O)$ |
| :---: | :---: | :---: | :---: |
| $\min (T C T)$ | 8.52 | 0.74 | 0.000 |
| $\min (M C T)$ | 8.69 | 0.67 | 0.000 |
| $\min (C T V)$ | 7.04 | 1.27 | 0.000 |
| $\min (T O)$ | $\mathbf{9 . 0 6}$ | $\mathbf{0 . 4 8}$ | - |

the process capability to produce outputs that fall between the specification limits. Given $\mu$ and $\sigma$ as the mean and standard deviation of the process outputs, $C_{p}=\frac{(U S L-L S L)}{6 \sigma}$ is the
process capability index that measures if the process is capable of producing outputs that are centered around the center-line of the specification limits, $L S L$ and $U S L$ denote lower and upper specification limits respectively. $C_{p k}=\min \left[\frac{U S L-\mu}{3 \sigma}, \frac{\mu-L S L}{3 \sigma}\right]$ is a performance indicator that measures if the mean value of process outputs falls between the specification limits [179]. Given $L S L$ and $U S L$, greater values of $C_{p}$ and $C_{p k}$ imply that a process generate outputs which are more centered with smaller variations.

To perform process capability analyses, we first need to define the specification limits of PSSI. Generally, a sustainability score of 8-10 indicates an excellent sustainability status [150, therefore, in this case study we consider $(L S L, U S L)=(8.50,10)$. Table 7.8 shows $C_{p}, C_{p k}$, and the percentage of results being less than $L S L$ for the alternative production schedules. It is observed that $\min (T O)$ outperforms all other productions scheduling alternatives. The outputs of $\min (T O)$ not only centered around the average value but also provide greater values of $C_{p k}$ that means the process is better under control in terms of sustainability index. In order to further evaluate the capability of the production scheduling alternatives, we have provided the percentage of observations that fall bellow the lower specification limits (i.e. $\%<L S L$ ) in Table 7.8. Lower values of $\%<L S L$ demonstrate that a process has greater capability relative to the lower specification limit. We use $\%<U S L$ to evaluate the capability of solutions, because the greater PSSI the more sustainable production schedule. Therefore, those observations that fall below the $L S L$ show large deviations from the excellent sustainability status. Only $11.76 \%$ of $\min (T O)$ outputs fall below 8.5 . As it is graphically shown by Figure 7.3 , the distributions of sustainability index of $\min (T O)$ has a negative skewness which means that the mass of distribution is concentrated close to the maximum value of $10 . \min (C T V)$ has the worst sustainability status with $P S S I=7.04, C_{p}=0.20$, and $C_{p k}=-0.39$ which mean that the output of the production schedule are highly scattered away from the average value. Single objective production scheduling alternatives poorly perform with high percentage of observation below $L S L$.

Table 7.7: PSSI capabilities indices for alternative production schedules

|  | $C_{p}$ | $C_{p k}$ | $\%<L S L$ |
| :---: | :---: | :---: | :---: |
| $\min (T C T)$ | 0.35 | 0.01 | 49.00 |
| $\min (M C T)$ | 0.38 | 0.10 | 38.65 |
| $\min (C T V)$ | 0.20 | -0.39 | 87.76 |
| $\min (T O)$ | $\mathbf{0 . 5 3}$ | $\mathbf{0 . 4 0}$ | $\mathbf{1 1 . 7 6}$ |

## Sustainability at sub-cluster level

A summary of observations and a comparison of the results for the application of PSSI for the studied production scheduling alternatives is presented here. Table 7.8 shows the statistics of sustainability scores at sub-cluster level. It is observed that even though $\min (T O)$ generates the highest value of PSSI, it does not necessarily generates the highest sustainability score for each sub-cluster. That is because some of sub-clusters are driven by a single metric (e.g. waiting time is driven only by $T C T$ ) and the weighting approach plays a big role on the contribution of each metric on the final PSSI value. As it is graphically shown by Figure 7.4 $\min (T O)$ has the largest area compared to the other alternative schedules. $\min (C T V)$ has the smallest area in Figure 7.4 with the lowest score in all sub-clusters but the waiting time variance, therefore, $\min (C T V)$ can be seen as the least sustainable scheduling alternative with PSSI of 7.04.


Figure 7.3: PSSI process capability charts for alternative production schedules

Table 7.8: Statistics of sustainability scores at sub-cluster level

|  | Production Cost |  | Energy Cost |  | Labor <br> Cost |  | Inventory <br> Cost |  | Energy <br> Consumption |  | Greenhouse gas <br> Emissions |  | Waiting Time |  | Waiting Time <br> Variance |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| $\min (T C T)$ | 8.16 | 0.82 | 8.78 | 0.82 | 8.78 | 0.82 | 10.00 | 0.00 | 8.78 | 0.82 | 8.78 | 0.82 | 10.00 | 0.00 | 6.32 | 1.65 |
| $\min (M C T)$ | 8.16 | 0.84 | 9.06 | 0.67 | 9.06 | 0.67 | 8.13 | 1.33 | 9.06 | 0.67 | 9.06 | 0.67 | 8.13 | 1.33 | 8.20 | 1.34 |
| $\min (C T V)$ | 7.55 | 0.97 | 6.67 | 1.51 | 6.67 | 1.51 | 5.11 | 1.94 | 6.67 | 1.51 | 6.67 | 1.51 | 5.11 | 1.94 | 10.00 | 0.00 |
| $\min (T O)$ | 8.50 | 0.77 | 9.45 | 0.43 | 9.45 | 0.43 | 9.35 | 0.70 | 9.45 | 0.43 | 9.45 | 0.43 | 9.35 | 0.70 | 7.64 | 1.49 |

$\mu$ : Average, $\sigma$ : Standard Deviation


Figure 7.4: Sustainability scores at sub-cluster level

Figure 7.5 shows the Pareto frontiers of sustainability scores average and standard devi-
ation at the sub-cluster level. It is observed that although $\min (T O)$ dominates all the other alternatives at PSSI level, it does not necessarily needs the dominance at the sub-cluster level. In fact the PSSI model provides the best compromise among the inconsistent objectives that leads to the best outcome at the higher level, and also provides the decision maker with a desirable flexibility at the lower levels. These two together are effective tools for the decision maker to ensure the stability of the overall performance while being able to hedge against the inconsistencies at the lower levels.


Figure 7.5: Sub-cluster sustainability score Pareto frontier

## Fundamental performance metrics

In order to evaluate the performance of production schedule alternatives in terms of the fundamental performance metrics of $T C T, M C T$, and $C T V$, we perform a series of statistical analyses including statistical process control (SPC) and process capabilities analysis. Table 7.9 presents the summary of $S P C$ charts (i.e. $\bar{x}$-R charts) for the studied production schedule alternatives. For the sake of brevity the $S P C$ charts are not presented.

As it is shown by table 7.9 single objective production scheduling alternatives of $\min (T C T)$, $\min (M C T)$, and $\min (C T V)$ generate no deviations on their intended objective but very large deviations on the others. Furthermore, single objective production scheduling alternatives generate lower PSSI compared to $\min (T O)$. On the other hand $\min (T O)$ not only provide the two best PSSI, respectively, but also small and uniform deviations on the fundamental performance metrics.

As it was discussed earlier, the first step in trade-off balancing is to establish an "ideal point", the coordinates of the ideal point are given by the optimum values of all objectives. It is obvious with inconsistent objectives the ideal point is not feasible to achieve, therefore, the ideal point is only a point of reference. The second step is to establish an "anti-ideal" point. The coordinates of the anti-ideal point are given by the worst values of all objectives. The objective of trade-off balancing is to find the closest efficient solution to the ideal-point. To

Table 7.9: Average performance and control limits of fundamental performance metrics

|  |  | $\min (T C T)$ | $\min (M C T)$ | $\min (C T V)$ | $\min (T O)$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $x-L C L$ | TCT | 0.000 | 0.151 | 0.442 | 0.045 |
|  | MCT | 0.203 | 0.000 | 0.142 | 0.027 |
|  | CTV | 0.327 | 0.147 | 0.000 | 0.199 |
|  | PSSI | 8.32 | 8.50 | 6.72 | 8.92 |
| $\bar{x}$ | $T C T$ | 0.000 | 0.187 | 0.490 | 0.065 |
|  | MCT | 0.245 | 0.000 | 0.176 | 0.045 |
|  | CTV | 0.368 | 0.180 | 0.000 | 0.236 |
|  | PSSI | 8.52 | 8.69 | 7.04 | 9.06 |
| $x-U C L$ | $T C T$ | 0.000 | 0.223 | 0.537 | 0.084 |
|  | MCT | 0.287 | 0.000 | 0.211 | 0.062 |
|  | CTV | 0.410 | 0.213 | 0.000 | 0.272 |
|  | PSSI | 8.71 | 8.87 | 7.36 | 9.19 |
| $R-L C L$ | TCT | 0.000 | 0.383 | 0.504 | 0.209 |
|  | MCT | 0.446 | 0.000 | 0.366 | 0.190 |
|  | CTV | 0.441 | 0.348 | 0.000 | 0.385 |
|  | PSSI | 2.07 | 1.95 | 3.43 | 1.41 |
| $R$ | $T C T$ | 0.000 | 0.603 | 0.796 | 0.329 |
|  | MCT | 0.702 | 0.000 | 0.576 | 0.299 |
|  | CTV | 0.695 | 0.548 | 0.000 | 0.606 |
|  | PSSI | 3.27 | 3.08 | 5.40 | 2.22 |
| $R-U C L$ | $T C T$ | 0.000 | 0.823 | 1.085 | 0.449 |
|  | MCT | 0.958 | 0.000 | 0.786 | 0.408 |
|  | CTV | 0.949 | 0.748 | 0.000 | 0.828 |
|  | PSSI | 4.46 | 4.20 | 7.36 | 3.03 |
| ${ }^{*} L C L$ : Lower control limit, $U C L$ : Upper control lim |  |  |  |  |  |

measure the distances between a solution and the ideal point, a family of distance functions are introduced as follows: $L_{p}(\omega, k)=\left[\sum_{k=1}^{3}\left(w_{k} D^{[k]}\right)^{p}\right]^{1 / p}$, where $w_{k}$ is the preference of the decision maker for metric $k$ [40]. When $p=1, L_{1}$ measures the longest distance (geometrically speaking) between solution and the ideal point. Let assume that all three fundamental metrics have the same importance for the decision maker, that is, $L_{1}=\frac{\sum_{k=1}^{3} D^{[k]}}{3}$. Table 7.10 shows the SPC statistics of $L_{1}$ for the studied alternatives.

It is observed that $\min (T O)$ generates the minimum average $L_{1}$ of 0.115 with the tightest bounds of [0.099, 0.131]. Therefore, it can be concluded that minimizing the trade-offs among the fundamental production scheduling performance indicators of $T C T, M C T$, and $C T V$ not only results in a sustainable production schedule but also provides a better control over the system performance at the metric level.

In addition to $S P C$ analyses, we perform process capabilities analyses for the fundamental performance metrics. In order to perform capability analyses, we first need to define the specification limits of each performance indicator. Equations 7.17 and 7.18 represent the $L S L$ and $U S L$ of performance indictor $k$ using the performance of $\min (T o)$ (because it has $\min \left(L_{1, h}\right)$. This definition not only provides a tight specification limits with only one standard

Table 7.10: $L_{1}$ average performance and control limits

|  | $\min (T C T)$ | $\min (M C T)$ | $\min (C T V)$ | $\min (T O)$ |
| :---: | :---: | :---: | :---: | :---: |
| $x$-LCL | 0.178 | 0.107 | 0.196 | $\mathbf{0 . 0 9 9}$ |
| $\bar{x}$ | 0.204 | 0.122 | 0.222 | $\mathbf{0 . 1 1 5}$ |
| $x$-UCL | 0.231 | 0.138 | 0.248 | $\mathbf{0 . 1 3 1}$ |
| $R$-LCL | 0.276 | 0.162 | 0.276 | 0.171 |
| $R$ | 0.435 | 0.256 | 0.435 | 0.270 |
| $R$-USL | 0.593 | 0.349 | 0.594 | 0.368 |
| std | 0.098 | $\mathbf{0 . 0 5 6}$ | 0.100 | 0.060 |
| *std: standard deviation |  |  |  |  |

deviation but also drives the specification limits towards 0 which is desirable for minimizing the deviation from the best value for each performance indicator. Table 7.11 shows the specification limits used in this study.

$$
\begin{array}{r}
L S L^{[k]}=\max \left[0, \mu_{\min (T O)}^{[k]}-\sigma_{\min (T O)}^{[k]}\right] \\
U S L^{[k]}=\mu_{\min (T O)}^{[k]}+\sigma_{\min (T O)}^{[k]} \tag{7.18}
\end{array}
$$

Table 7.11: Specification limits of performance indicators

|  | $\min (T O)$ |  | Specification limits |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $\mu$ | $\sigma$ | $L S L$ | $U S L$ |
| $T C T$ | 0.065 | 0.070 | 0.000 | 0.135 |
| $M C T$ | 0.045 | 0.064 | 0.000 | 0.109 |
| $C T V$ | 0.236 | 0.149 | 0.087 | 0.385 |
| $\min (T O)=\min \left(T O^{[0.5,0.3,0.2]}\right)$ |  |  |  |  |

Given the specification limits shown in table 7.11, we calculate $C_{p}$ and $C_{p k}$ of the fundamental performance metrics for each scheduling alternative. The results of the capability analyses are shown by table 7.12 .

Table 7.12: Capability results for the scheduling alternatives

|  |  | $\min (T C T)$ | $\min (M C T)$ | $\min (C T V)$ | $\min (T O)$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $C_{p}$ | $T C T$ | $\inf$ | 0.17 | 0.12 | 0.32 |
|  | $M C T$ | 0.11 | $\inf$ | 0.14 | 0.29 |
|  | $C T V$ | 0.32 | 0.37 | $-\inf$ | 0.34 |
| $C_{p k}$ | $T C T$ | $\inf$ | -0.13 | -0.62 | 0.31 |
|  | $M C T$ | -0.28 | $\inf$ | -0.17 | 0.23 |
|  | $C T V$ | 0.04 | 0.23 | $-\inf$ | 0.34 |
|  | $T C T$ | 0.00 | 65.33 | 96.83 | 15.49 |
|  | $M C T$ | 79.72 | 0.00 | 69.13 | 15.56 |
|  | $C T V$ | 45.81 | 6.09 | 100.00 | 15.24 |

It is observed that the single objective optimizations poorly perform in terms of $C_{p}$ and $C_{p k}$ throughout all three fundamental metrics. $\min (T O)$ exhibits a uniform and acceptable
performance. The outputs of $\min (T O)$ not only centered around the average value but also provide greater values of $C_{p k}$ that means the process is better under control. In order to further evaluate the capability of different solutions, we have provided the percentage of observations that fall above the upper specification limits (i.e. $\%>U S L$ ) in table 7.12 . Lower values of $\%>U S L$ demonstrate that a process has greater capability relative to the upper specification limit. We use $\%>U S L$ to evaluate the capability of solutions, because in a minimization problem the objective is to minimize the deviations from the minimum possible value. Therefore, those observations that fall above the $U S L$ show large deviations from the best solution and are of extreme importance in decision making. Although single objective function show a very heterogeneous performance on different metrics i.e. no deviation on their intended objectives but very large percentage of outputs above $U S L$ for the others. On the other hand $\min (T O)$ evenly perform with roughly $\% 15.50$ above $U S L$ for all three metrics.

### 7.5 Conclusion

Production scheduling is an operational level of decision making that covers not only the manufacturing stage of the product life-cycle but also the use stage of the processes lifecycle. Our comprehensive literature review showed that a systematic model for evaluating the sustainability of a production schedule with respect to all aspects of triple bottom line $(T B L)$ is virtually absent. To fill this gap, in this chapter, we proposed a quantitative yet generic sustainability evaluation scheme for production scheduling. Our proposed Production Scheduling Sustainability Index ( $P S S I$ ) is a hierarchical procedure that decomposes the objectives defined in $T B L$ into three clusters of (i) Economic-oriented scheduling, (ii) Environment-oriented scheduling, and (iii) social-oriented scheduling. Each cluster is then decomposed into multiple sub-clusters each covering a specific area of impact in the production. Finally, each sub-cluster is divided into multiple measurable fundamental metrics. Once the fundamental metrics are measured an aggregation procedure including, normalization and weighting is adopted to calculate PSSI. The weighting scheme can reflect the decision maker's preferences according to the policies, regulations, and/or market necessities.

Through an extensive literature review, maximum completion time ( $M C T$ ), total completion time ( $T C T$ ), and completion time variance ( $C T V$ ) were identified as the most fundamental performance metrics that drive many other aspects of the production. We statistically showed that the objectives of $\min (T C T), \min (M C T)$, and $\min (C T V)$ are inconsistent, that is, a schedule optimizing all metrics is infeasible. We proposed a trade-off balancing scheme in order to minimize the distance between the production scheduling performance and an ideal yet infeasible point at which all the fundamental performance metrics are at their optimum values. We then utilized the normalized deviation of each metric from its best possible value to construct the sustainability scores at metric-level of PSSI.

In order to evaluate the efficiency and effectiveness of the proposed models, a comprehensive case study on a permutation flow shop scheduling problem was performed. Multiple production scheduling alternative were compared in terms of sustainability index and the fundamental performance metrics as well. Our results showed that scheduling for sustainability (i.e. $\max (P S S I)$ ) outperforms all the studied production scheduling alternatives with maximum average and minimum standard deviation of PSSI. It was observed that although $\max (P S S I)$ results in the best and tightest bounds for PSSI, it allows larger variation ranges for the fundamental metrics which provide a less sensitive schedule.

Our future studies will be in two directions including: (i) Since the weighting scheme can dramatically affect the skewness of PSSI distribution, further studies are required to find the
most suited weighting scheme for production scheduling sustainability. (ii) Considering the $N P$-completeness of most production scheduling problems, it is extremely time consuming to determine optimum schedules for a large number of jobs. Therefore, developing heuristics is necessary for the large scale problems. Our next step will be on developing a heuristic for balancing trade-offs among $\min (T C T), \min (M C T)$, and $\min (C T V)$.

## Chapter 8

## Conclusions and Future Work

### 8.1 Conclusions

This dissertation addresses an open challenge in the field of operating room planning and scheduling: inconsistencies among objectives at different levels of decision making process including strategic in long-term, tactical in medium-term, and operational in short-term. We proposed balancing trade-offs as an alternative objective function to address this gap. Table 8.1 presents the summary of this dissertation conclusions.

In Chapter 2, at the strategic level, we identified throughput and total cost as the fundamental performance measures driving many other areas of the system. We demonstrated the inconsistency between the objectives of maximizing throughput and minimizing total cost. A bi-level chance-constrained objective function was proposed to simultaneously address the inconsistency between objectives and variations in patient arrivals (demand) and surgical case times as disturbances blurring the decision space. We introduced two thresholds $T L$ and $\gamma$ as the tolerable deviations from the optima of overtime and throughput. Using $T L$ and $\gamma$, we provide OR managers a flexible tools to manage the number of patients in the waiting list within the financial means of the hospital. We compared the results of the proposed models with those of single-attribute models. The bi-level model outperformed the single attribute models by increasing the throughput by $50.84 \%$ with only $3.18 \%$ increase in the total cost. SPC results showed that the results of the bi-level objective function is better under control compared to those of the single-attribute objective functions.

In Chapter 3, at the tactical level, we identified surgery properties (i.e., priority and type), OR utilization, load distribution, and total cost as managerial concerns. Traditional bin-packing problem with $L P T$ as the initial sequence is commonly used to assign patients to ORs with the objective of maximizing utilization. However, the traditional bin-packing is only capable of addressing OR utilization among the aforementioned managerial concerns. To fill this gap, we proposed a novel algorithm called $P T D$ to form the initial sequence for the bin-packing model. The results showed that $P T D$ successfully addressed the surgery priorities by scheduling surgeries with high priorities in the head of the sequence. PTD also improved utilization by $10.4 \%$ compared to $L P T$. A metric called smoothness index $(S I)$ was proposed to evaluate the evenness of load distribution among ORs. PTD outperformed $L P T$ by $23 \%$ improvement in load distribution. The $P T D$ algorithm provides a powerful tool to manage the priority of surgeries, reducing the number of setups and turnovers, while increasing the utilization of the system.

In Chapter 4, at the operational level, patient flow mean $(P t F)$ and patient flow variance (PtFV) were identified as two key performance indicator affecting the outcomes of the OR
peri-operative process. $\min (P t F)$ and $\min (P t F V)$ are of importance for different stakeholders in the OR peri-operative process. By modeling OR as a single-machine production system, we analytically proved the inconsistencies of $\min (P t F)$ and $\min (P t F V)$, thus, any scheduling method generates some levels of trade-offs in the systems. We proposed a minimizing tradeoff function as the alternative objective function. By proving the conditions of optimality, a novel heuristic was proposed to find the optimal schedule in $O(n \log n)$. The performance of the proposed model was compared with those of single-attribute and existing bi-criteria models. The proposed models strictly dominated other studied models. The proposed model on average generated the minimum loss among the studied models with loss average and standard deviation of $(0.1767,0.0232)$ compared to $(0.2537,0.1480)$ and $(0.2600,0.1521)$ those of $\min (P t F)$ and $\min (P t F V)$. Minimizing trade-offs between $\min (P t F)$ and $\min (P t F V)$ shifted the outcome of the OR peri-operative process towards a more predictable state at which the outcomes are also at the minimum distance from their optima.

In Chapter 5, a mixed integer linear programming called $B M$ was proposed in order to minimize blockings between ORs and the downstream resources of ICUs and PACUs. The proposed $B M$ model considers the availability of the dowunstream resources before assigning patients to ORs. This function synchronize the ORs outflow with that of the downstream resources such that after performing a surgery in an OR, there is always a downstream unit available to host the patient. Therefore, the expensive OR times are not used for recovery purposes. The $B M$ model not only synchronizes ORs with the downstream resources, but also balances the daily occupancy of downstream units throughout the planning horizon. We compared the performance of the $B M$ model with an existing model that we called $P B$, and also with the studied hospital base model. The results showed that the $B M$ model outperformed the $P B$ and the base model by $85 \%$ and $94 \%$, respectively, in terms of the number of blockings. By evenly distributing the workload among ORs and among the weekdays, the $B M$ model provided on average a capacity buffer of at least 5.7 units in the downstream stages. This capacity buffer mitigates the negative impacts of variations in demand/case times, or it can be used to treat more patients. The results demonstrated the necessity of considering the availability of the downstream resources while assigning patients to ORs as opposed to the traditional approach of maximizing the ORs utilization and assuming that the downstream resources abound.

In Chapter 6, the operational risk management in the OR peri-operative context was studied. Risk definition, risk sources, and risk consequences were meticulously elaborated. Risk was defined as "the variations on the distributions of $O R$ peri-operative process outcomes, their likelihood, and their subjective values". Variations in demand and case times, and the inconsistencies among objectives were identified as the risk sources. Unpredictability of outcomes and their deviations from the optima were identified as risk consequences. Finally, balancing trade-offs was proposed as a risk mitigating strategy. The impacts of this risk mitigating strategy was studied in a surgical case sequencing problem as the test bed in which total completion time ( $T C T$ ) and maximum completion time $(M C T)$ were identified as the key performance indicators. It is worth mentioning that $\min (M C T)$ and $\min (T C T)$ are inconsistent objectives. $N E H$ and $L R$ are the best known heuristics to $\min (M C T)$ and $\min (T C T)$, respectively. The performance of the $N E H$ and $L R$ were compared with those of $C F D(\alpha)$ the best known heuristic to minimize the trade-offs between $\min (M C T)$ and $\min (T C T)$. Conditional value at risk, i.e., $\alpha$-CVaR was used to quantify the risk. The results showed that $C F D(0.2)$ outperformed the $N E H$ and $L R$ in terms of loss average with average loss of $5.12 \%$ compared to $50 \%$ and $49.8 \%$ those of the $N E H$ and $L R$, respectively. $C F D(0.3)$ outperformed the $N E H$ and $L R$ in terms of loss standard deviation with value of $3.5 \%$ com-
pared to $31.6 \%$ and $31.3 \%$ those of the $N E H$ and $L R$, respectively. $C F D(0.3)$ outperformed the $N E H$ and $L R$ in terms of $\alpha$-CVaR with average value of $6.03 \%$ compared to $27.41 \%$ and $37.43 \%$ those of the NEH and $L R$, respectively. Trade-off balancing is capable of reflecting the decision maker's preferences into the objective function by which the decision maker has a powerful yet flexible tool to shift the outcomes of the OR peri-op process towards a more predictable state with average ( $0.95-C V a R$ ) of $6.8 \%$, and acceptable subjective values for key performance indicators with only $3.27 \%$ and $1.94 \%$ deviations from the optimal solutions to $\min (M C T)$ and $\min (T C T)$, respectively.

In Chapter 7, we addressed the managerial concerns defined in the triple bottom line (TBL) including economic, environmental, and social aspect of the production. Through extensive literature review, $T C T, M C T$, and completion time variance ( $C T V$ ) were identified as the fundamental performance measures driving different aspects of $T B L$ during implementation of the production schedules. We showed the inconsistencies among the objectives of $\min (T C T), \min (M C T)$, and $\min (C T V)$. Therefore, any schedule generates some levels of trade-offs. We proposed a decomposition structure based on the normalized deviations to link the fundamental performance measures with the objectives defined in the $T B L$. The decomposition structure was followed by am aggregation procedure involving weighting to convert the normalized deviation into a sustainability score that holistically evaluate the sustainability performance of the production schedules. The proposed production scheduling sustainability index (PSSI) outperformed the single-attribute models with PSSI $=9.06$ compared to compared to $8.52,8.69$, and 7.04 , those of $\min (T C T), \min (M C T)$ and $\min (C T V)$, respectively. The PSSI model not only generates schedules with the highest sustainability scores, but also shift the performance of the system in terms of the fundamental performance measures towards a more controllable state. We deliberately kept the PSSI model as generic as possible such that any practitioner would be able to modify the model with regard to the constraints of their production systems, i.e., job shop, flow shop, and OR peri-op process as well.

Balancing trade-offs is a flexible yet powerful tool for operating room managers to address different stakeholders' preferences, multiple levels of decision making process, inconsistent objectives, and mitigating risks of non-realization.

Table 8.1: Summary of Conclusions

| Q O O O 0 | Problem | Existing Models | Proposed Models | Difference | Influence |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 2 | Variations in demand and case times | Newsvendor Model, Chance constrained programming | Bi-level Chanceconstrained models | Outperforming the Newsvendor model by $50.84 \%$ in terms of throughput with only $3.18 \%$ increase in the total cost. | - Elaborating the relationship between Total Cost and Throughput <br> - Introducing thresholds as a managerial tool to manage inconsistent objectives. |
| 3 | Surgery priorities, surgery types, and total cost | Bin-Packing with $\quad L P T$ as the initial sequence | Bin-Packing with $\quad P T D$ as the initial sequence | Increasing OR utilization by $10.4 \%$, -Providing a schedule which is insensitive to setup cost <br> - Evenly distributing the workloads among ORs | The PTD algorithm provides a powerful tool to manage the priority of surgeries, reducing the number of setups and turnovers, while increasing the utilization of the system. |
| 4 | Inconsistency between patient flow mean and variance | $\begin{aligned} & \min (P t F), \\ & \min (P t F V), \\ & \min (\text { bi-criteria }) \end{aligned}$ | Trade-off balancing model | Dominating the single-attribute and bicriteria models <br> - Resulting in minimum loss of (0.1767, 0.0232 ) compared to ( $0.2537,0.1480$ ) and (0.2600,0.1521) those of $\min (\mathrm{PtFV})$ and $\min (\mathrm{PtF})$ | - Elaborating the relationship between patient flow mean and variance. <br> - Providing a tool to simultaneously address conflicting interests. |
| 5 | Blockings between stages | $P B$ model, Hospital base model | $B M$ model | - Outperforming PB and the base models in terms of number of blockings by $85 \%$ and $94 \%$ respectively. <br> - Evenly distributing the workload throughout the week. Less weekend shift. Providing capacity for roughly 250 more patients. | Demonstrating the relationship between the OR utilization and downstream resources availability. |
| 6 | Variations in demand and case times, and inconsistencies among objectives | $L R, N E H$ | Trade-off balancing models $C F D(\alpha)$ with $\alpha$-CVaR as the risk measure | - Outperforming $N E H$ and $L R$ in terms of Loss average with average loss of $5.12 \%$ compared to $50 \%$ and $49.8 \%$ those of NEH and $L R$. <br> - Outperforming $N E H$ and $L R$ in terms of Loss standard deviation with value of $3.5 \%$ compared to $31.6 \%$ and $31.3 \%$ those of $N E H$ and $L R$. <br> - Outperforming $N E H$ and $L R$ in terms of $\alpha$-CVaR with average value of $6.03 \%$ compared to $27.41 \%$ and $37.43 \%$ those of $N E H$ and $L R$. | - Providing a quantitative risk measure in OR peri-op process. <br> - Marking the boundaries of Risk in OR peri-op process. <br> - Shifting the outcomes of the peri-op process toward a state with CVaR of $6.8 \%$ and only $3.27 \%$ and $1.94 \%$ deviations from the optima of Util and LoS. |
| 7 | Inconsistencies among objectives defined in the triple bottom lines | $\begin{aligned} & \min (T C T) \\ & \min (M C T) \\ & \min (C T V) \end{aligned}$ | $\begin{aligned} & \min (T O) \\ & P S S I \text { model } \end{aligned}$ | - Dominating single-attribute models in terms of fundamental performance metrics with average value of L1 of 0.115 compared to $0.204,0.122$, and 0.222 those of $\min (T C T), \min (M C T)$ and $\min (C T V)$, respectively. <br> - Outperforming single-atribute models in terms of sustainability index with average value of 9.06 compared to $8.52,8.69$, and 7.04 those of $\min (T C T), \min (M C T)$ and $\min (C T V)$, respectively. | - Defining production scheduling sustainability index (PSSI) for the first time. <br> - Bridging the fundamental performance indicators with the objectives defined in the $T B L$. <br> - Providing managerial guidelines to simultaneously address sustainability and process control concerns. |

### 8.2 Direction for future works

OR schedules drive the peri-operative process performance, and disturbances (e.g., emergencies, overruns, cancelations) negatively affect the performance of the peri-operative process by decreasing the predictability of the outcomes. We propose mitigating the uncertainty through the application of scheduling and adaptive control mechanisms grounded control theory as shown by Figure 8.1, and illustrated by Equation (8.1) and Equation (8.2) as a linear time variant system. Where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{D}_{\mathbf{1}}$, and $\mathbf{D}_{\mathbf{2}}$ are the state-space matrices. $t \geq 0, \mathbf{x}(t) \in \mathbb{R}^{n}$ is the state, $\mathbf{x}(0) \in \mathbb{R}^{n}$ is the initial conditions, $\mathbf{u}(t) \in \mathbb{R}^{m}$ is the control, $\mathbf{y}(t) \in \mathbb{R}^{k}$ is the measured performance, and $\mathbf{d}(t) \in \mathbb{R}^{q}$ is the unmeasured disturbance.

$$
\begin{align*}
\dot{\mathbf{x}}(t) & =\mathbf{A x}(t)+\mathbf{B u}(t)+\mathbf{D}_{\mathbf{1}} \mathbf{d}(t)  \tag{8.1}\\
\mathbf{y}(t) & =\mathbf{C x}(t)+\mathbf{D u}(t)+\mathbf{D}_{\mathbf{2}} \mathbf{d}(t) \tag{8.2}
\end{align*}
$$



Figure 8.1: Adaptive control structure for OR planning and scheduling

The objective of this proposed approach is to design the control $\mathbf{u}$ such that the negative impacts of the disturbance $\mathbf{d}$ is minimized or eliminated.

## Appendix A

## Appendices

## A. 1 Proof of Lemma 3

$\min (T C T)$ and $\min (T A D C)$ are inconsistent objectives

Proof by contradiction: Assume that there exists a sequence $\boldsymbol{\sigma}$ with assignment matrix $\mathbf{X}$ that simultaneously minimizes $T C T$ and $T A D C$. Therefore, Equations (A.1) and (A.2) hold.

$$
\begin{align*}
& L B_{1}(\mathbf{P})=\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})\right)=\mathbf{W}_{1} \mathbf{X P}  \tag{A.1}\\
& L B_{2}(\mathbf{P})=\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)=\mathbf{W}_{2} \mathbf{X P} \tag{A.2}
\end{align*}
$$

We can rewrite equations A.1 and A.2 into equation A.3).

$$
\begin{equation*}
\left(\mathbf{W}_{1}+\mathbf{W}_{2}\right) \mathbf{X P}=L B_{1}(\mathbf{P})+L B_{2}(\mathbf{P}) \tag{A.3}
\end{equation*}
$$

Let $\mathbf{P}=\left[p_{1}, p_{2}, p_{3}\right]^{T}$ be the processing times of three jobs, since $\mathbf{W}_{k}$ is sequence-independent, Equations A.1 and A.2 are rewritten into matrix form as follows:

$$
\begin{align*}
& {\left[\begin{array}{l}
3 \\
2 \\
1
\end{array}\right]^{T}\left[\begin{array}{lll}
x_{11} & x_{12} & x_{13} \\
x_{21} & x_{22} & x_{23} \\
x_{31} & x_{32} & x_{33}
\end{array}\right]\left[\begin{array}{l}
p_{1} \\
p_{2} \\
p_{3}
\end{array}\right]=\left[L B_{1}(\mathbf{P})\right]}  \tag{A.4}\\
& {\left[\begin{array}{l}
0 \\
2 \\
2
\end{array}\right]^{T}\left[\begin{array}{lll}
x_{11} & x_{12} & x_{13} \\
x_{21} & x_{22} & x_{23} \\
x_{31} & x_{32} & x_{33}
\end{array}\right]\left[\begin{array}{l}
p_{1} \\
p_{2} \\
p_{3}
\end{array}\right]=\left[L B_{2}(\mathbf{P})\right]} \tag{A.5}
\end{align*}
$$

Summation of Equation A.4 and Equation A.5 yields Equation A.6):

$$
\left[\begin{array}{l}
3  \tag{A.6}\\
4 \\
3
\end{array}\right]^{T}\left[\begin{array}{lll}
x_{11} & x_{12} & x_{13} \\
x_{21} & x_{22} & x_{23} \\
x_{31} & x_{32} & x_{33}
\end{array}\right]\left[\begin{array}{l}
p_{1} \\
p_{2} \\
p_{3}
\end{array}\right]=\left[L B_{1}(\mathbf{P})+L B_{2}(\mathbf{P})\right]
$$

Without loss of generality assume $p_{1}=p_{2}<p_{3}$;
By $S P T$ rule; $L B_{1}(\mathbf{P})=3 p_{1}+2 p 2+p 3$, and by Algorithm $2 ; L B_{2}(\mathbf{P})=2 p_{1}+2 p_{2}$, therefore, the right hand side of Equation A.6) is $\left[L B_{1}(\mathbf{P})+L B_{2}(\mathbf{P})\right]=\left[5 p_{1}+4 p_{2}+p_{3}\right]$. We can rewrite Equation A.6 in the equation form as:

$$
\begin{gather*}
3 p_{1}\left(x_{11}+x_{12}+x_{13}\right)+4 p_{2}\left(x_{21}+x_{22}+x_{23}\right)+3 p_{3}\left(x_{31}+x_{32}+x_{33}\right)=5 p_{1}+4 p_{2}+p_{3} \\
\sum_{i=1}^{n} x_{j, i}=1 \Rightarrow 3 p_{1}+4 p_{2}+3 p_{3}=5 p_{1}+4 p_{2}+p_{3} \Rightarrow p_{1}=p_{3} \\
p_{1}=p_{3} \Rightarrow \Leftarrow p_{1}<p_{3} \tag{A.7}
\end{gather*}
$$

By Equation A.7) $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})\right)$ and $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ are consistent if $p_{1}=p_{3}$ which contradicts with the assumption $p_{1}<p_{3}$, thus Lemma 3 is proved. $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{1}(\boldsymbol{\sigma}, \mathbf{P})\right)$ and $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}\left(\gamma_{2}(\boldsymbol{\sigma}, \mathbf{P})\right)$ are consistent if and only if $p_{1}=p_{2}=\ldots, p_{n}$ which is trivial.

## A. 2 Proof of Lemma 4

$\stackrel{+}{\boldsymbol{\sigma}}_{Z}(\mathbf{P})$; the optimal solution to $\min (Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$ is $V$-shaped.
If $\boldsymbol{\sigma}$ is the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$ and is NOT $V$-shaped, then there are three consecutive jobs $(i, j, k)$ such that $p_{j}>p_{i}$ and $p_{j}>p_{k}$, as depicted in Figure 4.1. Let $\boldsymbol{\sigma}=\{\sigma(1), \sigma(2), \ldots, \sigma(n)\}$ be a sequence and $\boldsymbol{\sigma}^{\prime}$ the sequence obtained from $\boldsymbol{\sigma}$ by exchanging $\sigma(i)$ and $\sigma(i+1)$, then

$$
\begin{equation*}
\delta=Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})-Z\left(\alpha, \boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)=\left(p_{\sigma(i)}-p_{\sigma(i+1)}\right)\left(\frac{\alpha}{R_{1}}+\frac{(1-\alpha)(2 i-n-1)}{R 2}\right) \tag{A.8}
\end{equation*}
$$

We prove that $\boldsymbol{\sigma}$ cannot be the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$ because exchanging either $i$ and $j$ or $j$ and $k$ will result in a smaller $Z(\alpha, \mathbf{P})$. Let $\boldsymbol{\sigma}^{\prime}$ be the sequence generated by exchanging jobs $i$ and $j$. Similarly, let $\boldsymbol{\sigma}^{\prime \prime}$ be the sequence generated by exchanging jobs $j$ and $k$. We need to prove the followings:
I. $Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})-Z\left(\alpha, \boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)<0 \Rightarrow Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})-Z\left(\alpha, \boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)>0$
II. $Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})-Z\left(\alpha, \boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)<0 \Rightarrow Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})-Z\left(\alpha, \boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)>0$

Proof for II. is similar to the proof for I.) therefore, we only prove I.. Assume that job $i$ is in position $e$ in $\boldsymbol{\sigma}$;

$$
\begin{array}{r}
\text { By assumption } \quad Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})-Z\left(\alpha, \boldsymbol{\sigma}^{\prime}, \mathbf{P}\right)<0 \Rightarrow \\
\underbrace{\left(p_{i}-p_{j}\right)}_{<0}\left(\frac{\alpha}{R_{1}}+\frac{(1-\alpha)(2 e-n-1)}{R 2}\right)<0 \\
\Longrightarrow \frac{\alpha}{R_{1}}+\frac{(1-\alpha)(2 e-n-1)}{R 2}>0 \\
\begin{aligned}
\frac{(1-\alpha)}{}(\alpha, \boldsymbol{\sigma}, \mathbf{P})-Z\left(\alpha, \boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)=\underbrace{\left(p_{j}-p_{k}\right)}_{>0}(\underbrace{\frac{\alpha}{R_{1}}}_{>0}+\frac{(1-\alpha)(2 e-n-1)}{R 2} & \underbrace{\left.\frac{2(1-\alpha)}{R_{2}}\right)}_{>0} \\
& \Longrightarrow Z(\alpha, \boldsymbol{\sigma}, \mathbf{P})-Z\left(\alpha, \boldsymbol{\sigma}^{\prime \prime}, \mathbf{P}\right)>0
\end{aligned} \tag{A.10}
\end{array}
$$

Equation A. 10 clearly states that $\boldsymbol{\sigma}^{\prime \prime}$ is strictly better than $\boldsymbol{\sigma}$, therefore, the optimal sequence to $\min _{\boldsymbol{\sigma} \in \boldsymbol{\Omega}}(Z(\alpha, \boldsymbol{\sigma}, \mathbf{P}))$ must be $V$-shaped.

Copyright ${ }^{\circledR}$ Amin Abedini, 2019.

## Bibliography

[1] M. Hartman, A. B. Martin, J. Benson, A. Catlin, and N. H. E. A. Team, "National health spending in 2011: overall growth remains low, but some payers and services show signs of acceleration," Health Affairs, vol. 32, no. 1, pp. 87-99, 2013.
[2] R. Kogan, "The myth of the out-of-control federal government," Off the Charts Blog, Center on Budget and Policy Priorities, 2012.
[3] A. J. Weiss and A. Elixhauser, "Overview of hospital stays in the united states, 2012: Statistical brief\# 180," 2014.
[4] B. Denton, J. Viapiano, and A. Vogl, "Optimization of surgery sequencing and scheduling decisions under uncertainty," Health Care Management Science, vol. 10, no. 1, pp. 13-24, 2007.
[5] B. Moore, K. Levit, and A. Elixhauser, "Costs for hospital stays in the united states, 2012: Statistical brief\# 181," 2014.
[6] B. Cardoen, E. Demeulemeester, and J. Beliën, "Operating room planning and scheduling: A literature review," European Journal of Operational Research, vol. 201, no. 3, pp. 921-932, 2010.
[7] A. Pfuntner, L. Wier, and C. Steiner, "Costs for hospital stays in the united states, 2011: Statistical brief\# 168," 2013.
[8] D. A. Etzioni, J. H. Liu, M. A. Maggard, and C. Y. Ko, "The aging population and its impact on the surgery workforce," Annals of Surgery, vol. 238, no. 2, p. 170, 2003.
[9] M. Zeleny and J. L. Cochrane, Multiple criteria decision making. University of South Carolina Press, 1973.
[10] P. Santibáñez, M. Begen, and D. Atkins, "Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a british columbia health authority," Health Care Management Science, vol. 10, no. 3, pp. 269-282, 2007.
[11] A. Testi, E. Tanfani, and G. Torre, "A three-phase approach for operating theatre schedules," Health Care Management Science, vol. 10, no. 2, pp. 163-172, 2007.
[12] C. S. Hultman, "Procedural portfolio planning in plastic surgery, part 1: strategic changes in clinical practice to increase physician revenue, improve operative throughput, and maintain patient satisfaction," Annals of Plastic Surgery, vol. 76, pp. S344-S346, 2016.
[13] A. J. Fong, M. Smith, and A. Langerman, "Efficiency improvement in the operating room," Journal of Surgical Research, vol. 204, no. 2, pp. 371-383, 2016.
[14] V. Roshanaei, C. Luong, D. M. Aleman, and D. R. Urbach, "Collaborative operating room planning and scheduling," INFORMS Journal on Computing, vol. 29, no. 3, pp. 558-580, 2017.
[15] S. M. Ballard and M. E. Kuhl, "The use of simulation to determine maximum capacity in the surgical suite operating room," in Proceedings of the 38th Conference on Winter simulation, pp. 433-438, Winter Simulation Conference, 2006.
[16] B. Cardoen and E. Demeulemeester, "Capacity of clinical pathways: a strategic multilevel evaluation tool," Journal of Medical Systems, vol. 32, no. 6, pp. 443-452, 2008.
[17] M. Arenas, A. Bilbao, R. Caballero, T. Gomez, M. Rodríguez, and F. Ruiz, "Analysis via goal programming of the minimum achievable stay in surgical waiting lists," Journal of the Operational Research society, vol. 53, no. 4, pp. 387-396, 2002.
[18] S. Chaabane, N. Meskens, A. Guinet, and M. Laurent, "Comparison of two methods of operating theatre planning: application in belgian hospital," Journal of Systems Science and Systems Engineering, vol. 17, no. 2, pp. 171-186, 2008.
[19] I. Ozkarahan, "Allocation of surgeries to operating rooms by goal programing," Journal of Medical Systems, vol. 24, no. 6, pp. 339-378, 2000.
[20] E. van Veen-Berkx, S. G. Elkhuizen, B. Kuijper, G. Kazemier, and D. O. R. B. Collaborative, "Dedicated operating room for emergency surgery generates more utilization, less overtime, and less cancellations," The American Journal of Surgery, vol. 211, no. 1, pp. 122-128, 2016.
[21] X. Li, N. Rafaliya, M. F. Baki, and B. A. Chaouch, "Scheduling elective surgeries: the tradeoff among bed capacity, waiting patients and operating room utilization using goal programming," Health Care Management Science, vol. 20, no. 1, pp. 33-54, 2017.
[22] E. van Veen-Berkx, S. G. Elkhuizen, S. van Logten, W. F. Buhre, C. J. Kalkman, H. G. Gooszen, G. Kazemier, and D. O. R. B. Collaborative, "Enhancement opportunities in operating room utilization; with a statistical appendix," Journal of Surgical Research, vol. 194, no. 1, pp. 43-51, 2015.
[23] A. Jebali and A. Diabat, "A stochastic model for operating room planning under capacity constraints," International Journal of Production Research, vol. 53, no. 24, pp. 7252-7270, 2015.
[24] J. T. Blake and M. W. Carter, "A goal programming approach to strategic resource allocation in acute care hospitals," European Journal of Operational Research, vol. 140, no. 3, pp. 541-561, 2002.
[25] S. Ceschia and A. Schaerf, "Dynamic patient admission scheduling with operating room constraints, flexible horizons, and patient delays," Journal of Scheduling, vol. 19, no. 4, pp. 377-389, 2016.
[26] Y. Wang, J. Tang, Z. Pan, and C. Yan, "Particle swarm optimization-based planning and scheduling for a laminar-flow operating room with downstream resources," Soft Computing, vol. 19, no. 10, pp. 2913-2926, 2015.
[27] Q.-L. Lin, H.-C. Liu, D.-J. Wang, and L. Liu, "Integrating systematic layout planning with fuzzy constraint theory to design and optimize the facility layout for operating theatre in hospitals," Journal of Intelligent Manufacturing, vol. 26, no. 1, pp. 87-95, 2015.
[28] F. Dexter, J. T. Blake, D. H. Penning, and D. A. Lubarsky, "Calculating a potential increase in hospital margin for elective surgery by changing operating room time allocations or increasing nursing staffing to permit completion of more cases: a case study," Anesthesia \& Analgesia, vol. 94, no. 1, pp. 138-142, 2002.
[29] R. Guido and D. Conforti, "A hybrid genetic approach for solving an integrated multiobjective operating room planning and scheduling problem," Computers \& Operations Research, vol. 87, pp. 270-282, 2017.
[30] B. Addis, G. Carello, A. Grosso, and E. Tànfani, "Operating room scheduling and rescheduling: a rolling horizon approach," Flexible Services and Manufacturing Journal, vol. 28, no. 1-2, pp. 206-232, 2016.
[31] L. W. Robinson and R. R. Chen, "Scheduling doctors' appointments: optimal and empirically-based heuristic policies," IIE Transactions, vol. 35, no. 3, pp. 295-307, 2003.
[32] Z. Zhang and X. Xie, "Simulation-based optimization for surgery appointment scheduling of multiple operating rooms," IIE Transactions, vol. 47, no. 9, pp. 998-1012, 2015.
[33] A. Duenas, C. Di Martinelly, G. Y. Tütüncü, and J. Aguado, "A multi-objective hospital operating room planning and scheduling problem using compromise programming," in Mexican International Conference on Artificial Intelligence, pp. 379-390, Springer, 2016.
[34] R. Aringhieri, P. Landa, P. Soriano, E. Tànfani, and A. Testi, "A two level metaheuristic for the operating room scheduling and assignment problem," Computers \& Operations Research, vol. 54, pp. 21-34, 2015.
[35] E. Marcon and F. Dexter, "Impact of surgical sequencing on post anesthesia care unit staffing," Health Care Management Science, vol. 9, no. 1, pp. 87-98, 2006.
[36] E. Marcon and F. Dexter, "An observational study of surgeons' sequencing of cases and its impact on postanesthesia care unit and holding area staffing requirements at hospitals," Anesthesia \& Analgesia, vol. 105, no. 1, pp. 119-126, 2007.
[37] J. Vissers, I. J. Adan, and J. A. Bekkers, "Patient mix optimization in tactical cardiothoracic surgery planning: a case study," IMA journal of Management Mathematics, vol. 16, no. 3, pp. 281-304, 2005.
[38] D.-N. Pham and A. Klinkert, "Surgical case scheduling as a generalized job shop scheduling problem," European Journal of Operational Research, vol. 185, no. 3, pp. 1011-1025, 2008.
[39] I. R. Shafarevich and A. O. Remizov, Linear algebra and geometry. Springer Science \& Business Media, 2012.
[40] C. Romero, F. Amador, and A. Barco, "Multiple objectives in agricultural planning: a compromise programming application," American Journal of Agricultural Economics, vol. 69, no. 1, pp. 78-86, 1987.
[41] S. Nigel, The Blackwell encyclopedic dictionary of operations management. Wiley Blackwall, 1996.
[42] P. J. Montana and B. H. Charnov, Barron's management book, Fourth Edition. Barron's Educational Series, 2011.
[43] M. Pinedo, Scheduling. Springer, 2012.
[44] F. Dexter, A. Willemsen-Dunlap, and J. D. Lee, "Operating room managerial decisionmaking on the day of surgery with and without computer recommendations and status displays," Anesthesia \& Analgesia, vol. 105, no. 2, pp. 419-429, 2007.
[45] F. Dexter and R. D. Traub, "How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time," Anesthesia $\mathcal{E}$ Analgesia, vol. 94, no. 4, pp. 933-942, 2002.
[46] A. Abedini, W. Li, and H. Ye, "Stochastic bi-level optimization models for efficient operating room planning," Procedia Manufacturing, vol. 26, pp. 58-69, 2018.
[47] V. R. Fuchs, "Major trends in the us health economy since 1950," New England Journal of Medicine, vol. 366, no. 11, pp. 973-977, 2012.
[48] C. McIntosh, F. Dexter, and R. H. Epstein, "The impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: a tutorial using data from an australian hospital," Anesthesia 83 Analgesia, vol. 103, no. 6, pp. 1499-1516, 2006.
[49] F. Guerriero and R. Guido, "Operational research in the management of the operating theatre: a survey," Health Care Management Science, vol. 14, no. 1, pp. 89-114, 2011.
[50] M. Lamiri, X. Xie, A. Dolgui, and F. Grimaud, "A stochastic model for operating room planning with elective and emergency demand for surgery," European Journal of Operational Research, vol. 185, no. 3, pp. 1026-1037, 2008.
[51] B. T. Denton, A. J. Miller, H. J. Balasubramanian, and T. R. Huschka, "Optimal allocation of surgery blocks to operating rooms under uncertainty," Operations Research, vol. 58, no. 4-part-1, pp. 802-816, 2010.
[52] S. Chattopadhyay, D. R. Edwards, Y. Yu, and A. Hamidisepehr, "An assessment of climate change impacts on future water availability and droughts in the kentucky river basin," Environmental Processes, vol. 4, no. 3, pp. 477-507, 2017.
[53] P. T. VanBerkel and J. T. Blake, "A comprehensive simulation for wait time reduction and capacity planning applied in general surgery," Health Care Management Science, vol. 10, no. 4, pp. 373-385, 2007.
[54] D. Masursky, F. Dexter, S. A. Isaacson, and N. A. Nussmeier, "Surgeons' and anesthesiologists' perceptions of turnover times," Anesthesia \& Analgesia, vol. 112, no. 2, pp. 440-444, 2011.
[55] J. D. Little, "A proof for the queuing formula: $\mathrm{L}=\lambda \mathrm{w}$," Operations Research, vol. 9, no. 3, pp. 383-387, 1961.
[56] R. E. Wachtel and F. Dexter, "Tactical increases in operating room block time for capacity planning should not be based on utilization," Anesthesia \&3 Analgesia, vol. 106, no. 1, pp. 215-226, 2008.
[57] R. H. Epstein and F. Dexter, "Rescheduling of previously cancelled surgical cases does not increase variability in operating room workload when cases are scheduled based on maximizing efficiency of use of operating room time," Anesthesia $\mathcal{E}^{3}$ Analgesia, vol. 117, no. 4, pp. 995-1002, 2013.
[58] F. Dexter, A. Macario, and R. D. Traub, "Enterprise-wide patient scheduling information systems to coordinate surgical clinic and operating room scheduling can impair operating room efficiency," Anesthesia $\mathcal{G}$ Analgesia, vol. 91, no. 3, pp. 617-626, 2000.
[59] W. S. Lovejoy and Y. Li, "Hospital operating room capacity expansion," Management Science, vol. 48, no. 11, pp. 1369-1387, 2002.
[60] M. Olivares, C. Terwiesch, and L. Cassorla, "Structural estimation of the newsvendor model: an application to reserving operating room time," Management Science, vol. 54, no. 1, pp. 41-55, 2008.
[61] R. E. Wachtel and F. Dexter, "Review of behavioral operations experimental studies of newsvendor problems for operating room management," Anesthesia $\mathcal{E}^{\circ}$ Analgesia, vol. 110, no. 6, pp. 1698-1710, 2010.
[62] A. Charnes and W. W. Cooper, "Chance-constrained programming," Management Science, vol. 6, no. 1, pp. 73-79, 1959.
[63] O. V. Shylo, O. A. Prokopyev, and A. J. Schaefer, "Stochastic operating room scheduling for high-volume specialties under block booking," INFORMS Journal on Computing, vol. 25, no. 4, pp. 682-692, 2012.
[64] A. Seijas-Macías and A. Oliveira, "An approach to distribution of the product of two normal variables," Discussiones Mathematicae Probability and Statistics, vol. 32, no. 12, pp. 87-99, 2012.
[65] K. Oldham and J. Spanier, "The fractional calculus theory and applications of differentiation and integration to arbitrary order," in Mathematics in Science and Engineering, vol. 111, Elsevier, 1974.
[66] D. C. Montgomery, Introduction to statistical quality control. John Wiley \& Sons, 2007.
[67] A. Abedini, H. Ye, and W. Li, "Operating room planning under surgery type and priority constraints," Procedia Manufacturing, vol. 5, pp. 15-25, 2016.
[68] I. Marques, M. E. Captivo, and M. V. Pato, "An integer programming approach to elective surgery scheduling," OR Spectrum, vol. 34, no. 2, pp. 407-427, 2012.
[69] V. N. Hsu, R. De Matta, and C.-Y. Lee, "Scheduling patients in an ambulatory surgical center," Naval Research Logistics (NRL), vol. 50, no. 3, pp. 218-238, 2003.
[70] H. Fei, N. Meskens, and C. Chu, "A planning and scheduling problem for an operating theatre using an open scheduling strategy," Computers \&3 Industrial Engineering, vol. 58, no. 2, pp. 221-230, 2010.
[71] R. Valente, A. Testi, E. Tanfani, M. Fato, I. Porro, M. Santo, G. Santori, G. Torre, and G. Ansaldo, "A model to prioritize access to elective surgery on the basis of clinical urgency and waiting time," BMC Health Services Research, vol. 9, no. 1, p. 1, 2009.
[72] A. Macario, "What does one minute of operating room time cost?," Journal of Clinical Anesthesia, vol. 22, no. 4, pp. 233-236, 2010.
[73] National Center for Health Statistics. http://www.cdc.gov/nchs/fastats/inpatientsurgery.htm, Accessed: 2018-09-30.
[74] V. V. Vazirani, Approximation algorithms. Springer Science \& Business Media, 2013.
[75] G. Dósa, "The tight bound of first fit decreasing bin-packing algorithm is ffd (i) $\leq$ 11/9opt (i)+6/9," in International Symposium on Combinatorics, Algorithms, Probabilistic and Experimental Methodologies, pp. 1-11, Springer, 2007.
[76] J. Zhou and F. Dexter, "Method to assist in the scheduling of add-on surgical casesupper prediction bounds for surgical case durations based on the log-normal distribution," Anesthesiology: The Journal of the American Society of Anesthesiologists, vol. 89, no. 5, pp. 1228-1232, 1998.
[77] I. for Healthcare Improvement, "Optimizing patient flow: moving patients smoothly through acute care settings," IHI Innovation Series white paper, pp. 1-11, 2003.
[78] A. Jebali, A. B. H. Alouane, and P. Ladet, "Operating rooms scheduling," International Journal of Production Economics, vol. 99, no. 1-2, pp. 52-62, 2006.
[79] F. Dexter and J. Ledolter, "Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historic data," Anesthesiology: The Journal of the American Society of Anesthesiologists, vol. 103, no. 6, pp. 1259-1167, 2005.
[80] C. Mancilla and R. H. Storer, "Stochastic sequencing and scheduling of an operating room," Theses and Dissertations, Lehigh University, Department of Industrial and Systems Engineering (November 14, 2009), 2009.
[81] C. P. Childers and M. Maggard-Gibbons, "Understanding costs of care in the operating room," JAMA Surgery, vol. 153, no. 4, pp. e176233-e176233, 2018.
[82] F. Dexter, R. H. Epstein, R. D. Traub, and Y. Xiao, "Making management decisions on the day of surgery based on operating room efficiency and patient waiting times," Anesthesiology: The Journal of the American Society of Anesthesiologists, vol. 101, no. 6, pp. 1444-1453, 2004.
[83] A. T. Pedroja, "The tipping point: the relationship between volume and patient harm," American Journal of Medical Quality, vol. 23, no. 5, pp. 336-341, 2008.
[84] M. D. McHugh and C. Ma, "Hospital nursing and 30-day readmissions among medicare patients with heart failure, acute myocardial infarction, and pneumonia," Medical Care, vol. 51, no. 1, p. 52, 2013.
[85] D. R. Baker, P. J. Pronovost, L. L. Morlock, R. G. Geocadin, and C. G. Holzmueller, "Patient flow variability and unplanned readmissions to an intensive care unit," Critical Care Medicine, vol. 37, no. 11, pp. 2882-2887, 2009.
[86] P. P. Wang, "Optimally scheduling n customer arrival times for a single-server system," Computers \& Operations Research, vol. 24, no. 8, pp. 703-716, 1997.
[87] E. N. Weiss, "Models for determining estimated start times and case orderings in hospital operating rooms," IIE Transactions, vol. 22, no. 2, pp. 143-150, 1990.
[88] N. G. Hall and W. Kubiak, "Proof of a conjecture of schrage about the completion time variance problem," Operations Research Letters, vol. 10, no. 8, pp. 467-472, 1991.
[89] W. Kubiak, "Completion time variance minimization on a single machine is difficult," Operations Research Letters, vol. 14, no. 1, pp. 49-59, 1993.
[90] J. J. Kanet, "Minimizing variation of flow time in single machine systems," Management Science, vol. 27, no. 12, pp. 1453-1459, 1981.
[91] S. Eilon and I. Chowdhury, "Minimising waiting time variance in the single machine problem," Management Science, vol. 23, no. 6, pp. 567-575, 1977.
[92] U. Bagchi, "Simultaneous minimization of mean and variation of flow time and waiting time in single machine systems," Operations Research, vol. 37, no. 1, pp. 118-125, 1989.
[93] G. Hardy, J. Littlewood, and G. Polya, "Inequalities," Cambridge University Press, 1934.
[94] P. De, J. B. Ghosh, and C. E. Wells, "On the minimization of completion time variance with a bicriteria extension," Operations Research, vol. 40, no. 6, pp. 1148-1155, 1992.
[95] H. Markowitz, "Portfolio selection," The Journal of Finance, vol. 7, no. 1, pp. 77-91, 1952.
[96] A. Abedini, W. Li, and H. Ye, "An optimization model for operating room scheduling to reduce blocking across the perioperative process," Procedia Manufacturing, vol. 10, pp. 60-70, 2017.
[97] D. Gupta, "Surgical suites' operations management," Production and Operations Management, vol. 16, no. 6, pp. 689-700, 2007.
[98] J. T. Blake and M. W. Carter, "Physician and hospital funding options in a public system with decreasing resources," Socio-Economic Planning Sciences, vol. 37, no. 1, pp. 45-68, 2003.
[99] A. Baumgart, G. Schuepfer, A. Welker, H.-J. Bender, and A. Schleppers, "Status quo and current trends of operating room management in germany," Current Opinion in Anesthesiology, vol. 23, no. 2, pp. 193-200, 2010.
[100] F. Dexter, "Impact on operating room efficiency of reducing turnover times and anesthesia-controlled times," Annals of Surgery, vol. 245, no. 2, pp. 336-337, 2007.
[101] D. C. Krupka and W. S. Sandberg, "Operating room design and its impact on operating room economics," Current Opinion in Anesthesiology, vol. 19, no. 2, pp. 185-191, 2006.
[102] F. Dexter, "Operating room utilization: information management systems," Current Opinion in Anesthesiology, vol. 16, no. 6, pp. 619-622, 2003.
[103] M. Van Houdenhoven, E. W. Hans, J. Klein, G. Wullink, and G. Kazemier, "A norm utilisation for scarce hospital resources: evidence from operating rooms in a dutch university hospital," Journal of Medical Systems, vol. 31, no. 4, pp. 231-236, 2007.
[104] C. Banditori, P. Cappanera, and F. Visintin, "A combined optimization-simulation approach to the master surgical scheduling problem," IMA Journal of Management Mathematics, vol. 24, no. 2, pp. 155-187, 2013.
[105] B. Zhang, P. Murali, M. Dessouky, and D. Belson, "A mixed integer programming approach for allocating operating room capacity," Journal of the Operational Research Society, vol. 60, no. 5, pp. 663-673, 2009.
[106] M. J. Persson and J. A. Persson, "Analysing management policies for operating room planning using simulation," Health Care Management Science, vol. 13, no. 2, pp. 182191, 2010.
[107] C. Price, B. Golden, M. Harrington, R. Konewko, E. Wasil, and W. Herring, "Reducing boarding in a post-anesthesia care unit," Production and Operations Management, vol. 20, no. 3, pp. 431-441, 2011.
[108] L. D. Procter, D. L. Davenport, A. C. Bernard, and J. B. Zwischenberger, "General surgical operative duration is associated with increased risk-adjusted infectious complication rates and length of hospital stay," Journal of the American College of Surgeons, vol. 210, no. 1, pp. 60-65, 2010.
[109] H. Nathan and J. B. Dimick, "Medicare's shift to mandatory alternative payment models: why surgeons should care," JAMA surgery, vol. 152, no. 2, pp. 125-126, 2017.
[110] A. Macario, "Are your operating rooms 'efficient'?," OR Manager, vol. 23, no. 12, pp. 16-18, 2007.
[111] P. A. Najjar and S. W. Ashley, "How should surgeons interpret operating room costs?: Valuing our time," JAMA Surgery, vol. 153, no. 4, pp. e176234-e176234, 2018.
[112] A. M. Stey, R. H. Brook, J. Needleman, B. L. Hall, D. S. Zingmond, E. H. Lawson, and C. Y. Ko, "Hospital costs by cost center of inpatient hospitalization for medicare patients undergoing major abdominal surgery," Journal of the American College of Surgeons, vol. 220, no. 2, pp. 207-217, 2015.
[113] A. Moosavi and S. Ebrahimnejad, "Scheduling of elective patients considering upstream and downstream units and emergency demand using robust optimization," Computers E Industrial Engineering, vol. 120, pp. 216-233, 2018.
[114] W. Li, B. R. Nault, and H. Ye, "Trade-off balancing in scheduling for flow shop production and perioperative processes," European Journal of Operational Research, vol. 273, no. 3, pp. 817-830, 2019.
[115] J. Kingman, "The single server queue in heavy traffic," in Mathematical Proceedings of the Cambridge Philosophical Society, vol. 57, pp. 902-904, Cambridge University Press, 1961.
[116] P. Hopkin, Fundamentals of risk management: understanding, evaluating and implementing effective risk management. Kogan Page Publishers, 2018.
[117] U. Jüttner, H. Peck, and M. Christopher, "Supply chain risk management: outlining an agenda for future research," International Journal of Logistics: Research and Applications, vol. 6, no. 4, pp. 197-210, 2003.
[118] J. G. March and Z. Shapira, "Managerial perspectives on risk and risk taking," Management Science, vol. 33, no. 11, pp. 1404-1418, 1987.
[119] J. Castaing, A. Cohn, B. T. Denton, and A. Weizer, "A stochastic programming approach to reduce patient wait times and overtime in an outpatient infusion center," IIE Transactions on Healthcare Systems Engineering, vol. 6, no. 3, pp. 111-125, 2016.
[120] J. W. Pratt, "Risk aversion in the small and in the large," in Uncertainty in Economics, pp. 59-79, Elsevier, 1978.
[121] R. Litterman, "Hot spots and hedges," The Journal of Portfolio Management, vol. 23, no. 5, pp. 52-75, 1996.
[122] R. T. Rockafellar, S. Uryasev, et al., "Optimization of conditional value-at-risk," Journal of Risk, vol. 2, pp. 21-42, 2000.
[123] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," Mathematical Finance, vol. 9, no. 3, pp. 203-228, 1999.
[124] K. Dowd, Measuring market risk. John Wiley \& Sons, 2007.
[125] J. Philippe, "Value at risk: the new benchmark for managing financial risk," $N Y$ : McGraw-Hill Professional, 2001.
[126] C. Acerbi and D. Tasche, "Expected shortfall: a natural coherent alternative to value at risk," Economic Notes, vol. 31, no. 2, pp. 379-388, 2002.
[127] B. W. Nocco and R. M. Stulz, "Enterprise risk management: Theory and practice," Journal of Applied Corporate Finance, vol. 18, no. 4, pp. 8-20, 2006.
[128] M. Choudhry and M. Wong, An introduction to value-at-risk. John Wiley \& Sons, 2013.
[129] G. C. Pflug, "Some remarks on the value-at-risk and the conditional value-at-risk," in Probabilistic constrained optimization, pp. 272-281, Springer, 2000.
[130] M. Nawaz, E. Enscore, and I. Ham, "A heuristic algorithm forthe m-machine/n-job ow-shop sequencing problem. omega," The International Journal of the Managment Science, vol. 11, no. 1, 1983.
[131] J. Liu and C. R. Reeves, "Constructive and composite heuristic solutions to the $p / / \sum c i$ scheduling problem," European Journal of Operational Research, vol. 132, no. 2, pp. 439452, 2001.
[132] W. Li, V. L. Mitchell, and B. R. Nault, "Inconsistent objectives in operating room scheduling," in IIE Annual Conference. Proceedings, p. 727, Institute of Industrial and Systems Engineers (IISE), 2014.
[133] M. R. Garey, D. S. Johnson, and R. Sethi, "The complexity of flowshop and jobshop scheduling," Mathematics of Operations Research, vol. 1, no. 2, pp. 117-129, 1976.
[134] J. Hoogeveen and T. Kawaguchi, "Minimizing total completion time in a two-machine flowshop: analysis of special cases," Mathematics of Operations Research, vol. 24, no. 4, pp. 887-910, 1999.
[135] J. Elkington, "Partnerships from cannibals with forks: The triple bottom line of 21stcentury business," Environmental Quality Management, vol. 8, no. 1, pp. 37-51, 1998.
[136] EPA, United State Environmental Agency, Sutainable Manufacturing, 2017. https:// www.epa.gov/sustainability/sustainable-manufacturing,(accessed February 20, 2019).
[137] I. Jawahir, F. Badurdeen, and K. Rouch, Innovation in sustainable manufacturing education. Universitätsverlag der TU Berlin, 2015.
[138] A. Huang and F. Badurdeen, "Metrics-based approach to evaluate sustainable manufacturing performance at the production line and plant levels," Journal of Cleaner Production, vol. 192, pp. 462-476, 2018.
[139] I. Jawahir and R. Bradley, "Technological elements of circular economy and the principles of 6 r -based closed-loop material flow in sustainable manufacturing," Procedia CIRP, vol. 40, pp. 103-108, 2016.
[140] Y. He, Y. Li, T. Wu, and J. W. Sutherland, "An energy-responsive optimization method for machine tool selection and operation sequence in flexible machining job shops," Journal of Cleaner Production, vol. 87, pp. 245-254, 2015.
[141] X. Gong, T. De Pessemier, W. Joseph, and L. Martens, "An energy-cost-aware scheduling methodology for sustainable manufacturing," Procedia CIRP, vol. 29, pp. 185-190, 2015.
[142] Y. Chu, F. You, J. M. Wassick, and A. Agarwal, "Integrated planning and scheduling under production uncertainties: Bi-level model formulation and hybrid solution method," Computers $\mathcal{B}^{3}$ Chemical Engineering, vol. 72, pp. 255-272, 2015.
[143] M. L. Pinedo, Scheduling: theory, algorithms, and systems. Springer, 2016.
[144] J.-Y. Ding, S. Song, and C. Wu, "Carbon-efficient scheduling of flow shops by multiobjective optimization," European Journal of Operational Research, vol. 248, no. 3, pp. 758-771, 2016.
[145] S. A. Mansouri, E. Aktas, and U. Besikci, "Green scheduling of a two-machine flowshop: Trade-off between makespan and energy consumption," European Journal of Operational Research, vol. 248, no. 3, pp. 772-788, 2016.
[146] C. Gahm, F. Denz, M. Dirr, and A. Tuma, "Energy-efficient scheduling in manufacturing companies: a review and research framework," European Journal of Operational Research, vol. 248, no. 3, pp. 744-757, 2016.
[147] A. Giret, D. Trentesaux, and V. Prabhu, "Sustainability in manufacturing operations scheduling: A state of the art review," Journal of Manufacturing Systems, vol. 37, pp. 126-140, 2015.
[148] X. Gong, M. Van der Wee, T. De Pessemier, S. Verbrugge, D. Colle, L. Martens, and W. Joseph, "Energy-and labor-aware production scheduling for sustainable manufacturing: A case study on plastic bottle manufacturing," Procedia CIRP, vol. 61, pp. 387-392, 2017.
[149] M. Akbar and T. Irohara, "Scheduling for sustainable manufacturing: A review," Journal of Cleaner Production, vol. 205, pp. 866-883, 2018.
[150] F. Badurdeen, M. A. Shuaib, T. Lu, and I. S. Jawahir, "Sustainable value creation in manufacturing at product and process levels: Metrics-based evaluation," in Handbook of Manufacturing Engineering and Technology, pp. 3343-3375, Springer, 2015.
[151] W. Li, H. Dai, and D. Zhang, "The relationship between maximum completion time and total completion time in flowshop production," Procedia Manufacturing, vol. 1, pp. 146-156, 2015.
[152] A. Jayal, F. Badurdeen, O. Dillon Jr, and I. Jawahir, "Sustainable manufacturing: Modeling and optimization challenges at the product, process and system levels," CIRP Journal of Manufacturing Science and Technology, vol. 2, no. 3, pp. 144-152, 2010.
[153] J. M. Framinan, J. N. Gupta, and R. Leisten, "A review and classification of heuristics for permutation flow-shop scheduling with makespan objective," Journal of the Operational Research Society, vol. 55, no. 12, pp. 1243-1255, 2004.
[154] S. Reza Hejazi and S. Saghafian, "Flowshop-scheduling problems with makespan criterion: a review," International Journal of Production Research, vol. 43, no. 14, pp. 28952929, 2005.
[155] I. A. Chaudhry and A. A. Khan, "A research survey: review of flexible job shop scheduling techniques," International Transactions in Operational Research, vol. 23, no. 3, pp. 551-591, 2016.
[156] S. J. Pulluru and R. Akkerman, "Water-integrated scheduling of batch process plants: Modelling approach and application in technology selection," European Journal of Operational Research, vol. 269, no. 1, pp. 227-243, 2018.
[157] N. G. Hall, W. Kubiak, and S. P. Sethi, "Earliness-tardiness scheduling problems, ii: Deviation of completion times about a restrictive common due date," Operations Research, vol. 39, no. 5, pp. 847-856, 1991.
[158] N. G. Hall and M. E. Posner, "Earliness-tardiness scheduling problems, i: weighted deviation of completion times about a common due date," Operations Research, vol. 39, no. 5, pp. 836-846, 1991.
[159] G.-S. Liu, Y. Zhou, and H.-D. Yang, "Minimizing energy consumption and tardiness penalty for fuzzy flow shop scheduling with state-dependent setup time," Journal of Cleaner Production, vol. 147, pp. 470-484, 2017.
[160] K. Gowrishankar, C. Rajendran, and G. Srinivasan, "Flow shop scheduling algorithms for minimizing the completion time variance and the sum of squares of completion time deviations from a common due date," European Journal of Operational Research, vol. 132, no. 3, pp. 643-665, 2001.
[161] F. Gao, M. Liu, J.-J. Wang, and Y.-Y. Lu, "No-wait two-machine permutation flow shop scheduling problem with learning effect, common due date and controllable job processing times," International Journal of Production Research, vol. 56, no. 6, pp. 2361-2369, 2018.
[162] G. Li, M.-L. Luo, W.-J. Zhang, and X.-Y. Wang, "Single-machine due-window assignment scheduling based on common flow allowance, learning effect and resource allocation," International Journal of Production Research, vol. 53, no. 4, pp. 1228-1241, 2015.
[163] K. R. Baker and G. D. Scudder, "Sequencing with earliness and tardiness penalties: a review," Operations Research, vol. 38, no. 1, pp. 22-36, 1990.
[164] A. Bruzzone, D. Anghinolfi, M. Paolucci, and F. Tonelli, "Energy-aware scheduling for improving manufacturing process sustainability: A mathematical model for flexible flow shops," CIRP annals, vol. 61, no. 1, pp. 459-462, 2012.
[165] R. Neugebauer, M. Wabner, H. Rentzsch, and S. Ihlenfeldt, "Structure principles of energy efficient machine tools," CIRP Journal of Manufacturing Science and Technology, vol. 4, no. 2, pp. 136-147, 2011.
[166] F. Shrouf, J. Ordieres-Meré, A. García-Sánchez, and M. Ortega-Mier, "Optimizing the production scheduling of a single machine to minimize total energy consumption costs," Journal of Cleaner Production, vol. 67, pp. 197-207, 2014.
[167] G. Mouzon, M. B. Yildirim, and J. Twomey, "Operational methods for minimization of energy consumption of manufacturing equipment," International Journal of Production Research, vol. 45, no. 18-19, pp. 4247-4271, 2007.
[168] M. Dai, D. Tang, A. Giret, M. A. Salido, and W. D. Li, "Energy-efficient scheduling for a flexible flow shop using an improved genetic-simulated annealing algorithm," Robotics and Computer-Integrated Manufacturing, vol. 29, no. 5, pp. 418-429, 2013.
[169] H. Luo, B. Du, G. Q. Huang, H. Chen, and X. Li, "Hybrid flow shop scheduling considering machine electricity consumption cost," International Journal of Production Economics, vol. 146, no. 2, pp. 423-439, 2013.
[170] K. Fang, N. Uhan, F. Zhao, and J. W. Sutherland, "A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction," Journal of Manufacturing Systems, vol. 30, no. 4, pp. 234-240, 2011.
[171] W. Li, A. Zein, S. Kara, and C. Herrmann, "An investigation into fixed energy consumption of machine tools," in Glocalized solutions for sustainability in manufacturing, pp. 268-273, Springer, 2011.
[172] M. Mori, M. Fujishima, Y. Inamasu, and Y. Oda, "A study on energy efficiency improvement for machine tools," CIRP Annals, vol. 60, no. 1, pp. 145-148, 2011.
[173] S. Tiwari, Y. Daryanto, and H. M. Wee, "Sustainable inventory management with deteriorating and imperfect quality items considering carbon emission," Journal of Cleaner Production, vol. 192, pp. 281-292, 2018.
[174] A. Che, S. Zhang, and X. Wu, "Energy-conscious unrelated parallel machine scheduling under time-of-use electricity tariffs," Journal of Cleaner Production, vol. 156, pp. 688697, 2017.
[175] Q. Liu, M. Zhan, F. O. Chekem, X. Shao, B. Ying, and J. W. Sutherland, "A hybrid fruit fly algorithm for solving flexible job-shop scheduling to reduce manufacturing carbon footprint," Journal of Cleaner Production, vol. 168, pp. 668-678, 2017.
[176] Y. Zhang, J. Wang, and Y. Liu, "Game theory based real-time multi-objective flexible job shop scheduling considering environmental impact," Journal of Cleaner Production, vol. 167, pp. 665-679, 2017.
[177] R. Leisten and C. Rajendran, "Variability of completion time differences in permutation flow shop scheduling," Computers \& Operations Research, vol. 54, pp. 155-167, 2015.
[178] M. Voorneveld, "Characterization of pareto dominance," Operations Research Letters, vol. 31, no. 1, pp. 7-11, 2003.
[179] D. C. Montgomery, Statistical quality control, vol. 7. Wiley New York, 2009.

## Vita

Amin Abedini was born in Aligodarz, Iran, in 1984. Amin finished his undergraduate studies in Mechanical Engineering at Iran University of Science and Technology in 2011. Later in 2013, he received his M.Sc. degree in Mechanical Engineering from Urmia University, Iran. In 2015, he joined the Department of Mechanical Engineering at the University of Kentucky pursuing his Ph.D., his research interests include resource allocation, planning, and scheduling.

