

International Series in
Operations Research & Management Science

G.S.R. Murthy

Applications of Operations Research and Management Science

Case Studies



 Springer

International Series in Operations Research & Management Science

Volume 229

Series Editor

Camille C. Price

Stephen F. Austin State University, TX, USA

Associate Series Editor

Joe Zhu

Worcester Polytechnic Institute, MA, USA

Founding Series Editor

Frederick S. Hillier

Stanford University, CA, USA

More information about this series at <http://www.springer.com/series/6161>

G.S.R. Murthy

Applications of Operations Research and Management Science

Case Studies



Springer

G.S.R. Murthy
Indian Statistical Institute
Hyderabad, India

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-3-319-17920-9 ISBN 978-3-319-17921-6 (eBook)
DOI 10.1007/978-3-319-17921-6

Library of Congress Control Number: 2015939080

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

To my late parents

G. S. Acharyulu

and

Kanakavalli

Foreword

I, myself, am an alumnus of the Indian Statistical Institute (ISI). I studied in their M. Stat. program between 1955 and 1957 and learnt my early reasoning and mathematical skills during those years. I became acquainted with the “operations research (OR)”, a new subject being developed around that time, in lectures at ISI by a visiting American faculty under a United Nations Visiting Faculty Program. I learnt about the “travelling salesman problem (TSP)” in one of those lectures and found it so challenging that I kept thinking about ways for solving it. The director of the SQC & OR Division of ISI at that time, Mr. S. C. Sen, saw my preoccupation with the TSP and encouraged me to apply for the Fulbright Travel Grant for a year’s study at an American university, just being introduced at that time, so that I could learn OR for carrying out my research on the TSP with ISI providing funds for my living expenses if I succeeded in getting the Fulbright Travel Grant.

With Mr. S. C. Sen’s help, I received the grant and spent the 1961–1962 academic year at the Case Institute of Technology in Cleveland, Ohio, learning OR and, while there, I developed the branch and bound method for the TSP. With that excellent beginning, I was off to a research career in OR. You can understand why I consider ISI a great educational institute and why I consider myself fortunate in being educated at ISI.

Around that time, I realized that there are really two branches for research in OR. One is “theory” (including the development of methods and algorithms) and the other is “practice” dealing with applying the methods to solve decision-making problems in real-life applications. Unfortunately, in those days each of the two groups “theoreticians” and “practitioners” had very low respect for the other group. This was the reason for the story: practitioners said that “theoreticians do nothing practical,” while theoreticians said that “practitioners do practically nothing.”

Philosophically I am opposed to fences, and even though I started doing research in theory, I considered myself belonging to both the groups. However, I also had the opinion that working in practice is easier than working in theory, perhaps created by

the fact that all the textbooks in OR at that time are theory books with only minor discussion of applications.

I had to drastically revise my opinion about practice being easier than theory when I started working on the decision-making problems in the storage of arriving containers at Hong Kong International Terminals, to reduce congestion on the terminal road network, which was traced as the main reason for delaying container trucks from reaching shore cranes, thus reducing the productivity of those cranes. I found the problem of developing an appropriate mathematical model for this problem to be very hard; initially, I could not even figure out what the appropriate decision variables in the model should be.

So, I looked up journal publications on the problem. Some of them modelled the problem using binary variables representing the allocation of available storage spaces in the storage yard to each of the arriving containers, which led to a large 0–1 model for the problem. Unfortunately, I found that this model is not only impractical but also totally inappropriate, as the set of available storage spaces in the storage yard changes uncontrollably due to constant reshuffling of stored containers in daily operations.

Then, I tried to model the problem as a multi-commodity flow problem to equalize the flow of container trucks on all the arcs of the terminal road network and, over time, as much as possible to minimize congestion. Sadly, this model was also found to be inappropriate for the problem, as truck drivers resented being told what routes to take and also as the flow rates of arriving containers is quite uneven over time because of the stochastic nature of vessel arrivals.

It took us several months of getting intimate knowledge of the various daily operations inside the terminal and hard thinking to come up with a suitable mathematical model for this problem based on a substitute objective function technique.

Given this background, you can see why I commend this effort by the author, Prof. G. S. R. Murthy, in preparing this textbook to teach modelling skills to students and practitioners of OR, based on over 10 interesting and challenging applications in a wide variety of application areas. In preparing this book for his students and clients, he is carrying on the tradition of ISI as an excellent educational institution.

My best wishes for the success of his book and to his students for very successful careers.

Hyderabad, India
February 2015

Katta Gopalakrishna Murty

Preface

The case studies presented in this book have evolved from my close association with a wide variety of industries over the past three decades. As a trainer and a consultant, I had the opportunity of working with industrial personnel at all levels and could gain a good insight into the ground realities and the state of management of operations in a large number of industries. In most of the industries, there is a huge scope for promoting applications for optimal decision making. Due to routine way of functioning, the executives and managers fail to recognize and identify opportunities for improvement studies. The main problem stems from the fact that they have little or no exposure to case study models which present the practical problems from a practitioner's perspective. There has been a frequent demand from them for books with specific case studies covering modelling aspects without getting much into mathematics. This has been one of the motivating factors for writing this book.

Many academic programmes on quality management, business management, and engineering disciplines at graduate and postgraduate levels emphasize the need for practical training as part of their course curricula. The Indian Statistical Institute runs a 6-month part-time academic programme exclusively for executives from industry with the objective of promoting applications of statistics and operations research in industry. Being a teacher in academic programmes and having close association with industry, I had the opportunity of training and guiding many of these students/participants on carrying out a number of application projects. Many of these students and participants have also been asking me to write a book on realistic case study applications. This is another motivating force to write this book. Some of the case studies presented in this book are taken from the studies in which the students/participants were involved.

Another source of inspiration for writing this book is my own fascination for the subject of applying statistics and operations research to industrial problems and helping the industrial personnel in promoting the applications.

Most of the problems in industry can be modelled with fundamental techniques of statistics and operations research. Sometimes very simple tools provide very powerful solutions. The chapter on the procedure of testing vaccines (Chap. 3) is a nice

example for this. Almost all applications of operations research I have come across use linear programming or integer linear programming formulations and nothing beyond. What makes these applications difficult is recognizing the problems, identifying the decision variables and constraints, and modelling the problems. Therefore, I sincerely believe that applications can be promoted even with limited knowledge of the fundamental tools and techniques and what is required is the skill of modelling the problem. Many authors of operations research think that modelling is an art. While this is true to some extent, many of the problems are typical in their nature, and hence, knowledge of case studies should be of great help in promoting the applications.

I have written this book keeping the above thoughts in mind and using the experience I have gained over the years from my interaction with students as a teacher and with industrial personnel as a trainer and consultant. Much of the emphasis is laid on understanding the problem, identifying the decision variables, understanding and formulating the constraints, and modelling the problems. As far as the operations research applications are concerned, mathematics is kept at a minimum level. The only exception perhaps is Chap. 10 which is about constructing an efficient design for an industrial experiment that was to be performed in a foundry. For the cases concerning statistical applications, some background knowledge is expected. These cases include applications of reliability, multivariate analysis, and Markov chains. For these studies, it is recommended that the readers should focus more on understanding the problems and identifying the right tools.

Fortunately, we are in the age of powerful computing resources. Once a problem is formulated, there are a number of software packages that can be used for solving the problem. In this direction, I find that in majority of the cases, industries still use Microsoft Excel for handling data, and any solution provided using excel is the first preference as it is user-friendly and cost saving.

I have selected ten case studies for this book, and each one of them is presented as a chapter. Again Chap. 7 is an exception to this. Chapter 7 covers five cases and it is about intelligent modelling. I believe that some of the chapters will be very useful for those who aspire to become consultants or provide software solutions by developing decision support systems. For example, using the detailed presentation on how deckle optimization can be done in paper and paperboard industries (Chap. 5), one can develop a software package for the deckle optimization problem. A brief overview of the chapters and their background is presented below.

Chapter 1 deals with a material optimization problem. The problem is about minimizing sheet metal consumption of mild steel used for producing automobile parts. It uses simple integer linear programming. This application was taken up as a project by one of the participants (of the part-time programme for industrial executives) as a part of his academic curriculum.

Chapter 2 is about minimizing production time of producing printed circuit boards. The problem is a complex combinatorial optimization problem. A nice heuristic is available for this problem which uses the travelling salesman problem and the assignment problem models. A software tool was developed for the manager to make quick and near optimal decisions for this problem.

Chapter 3 deals with a test procedure for quality control of hepatitis B vaccine. This study was carried out for a biotechnology company. A part of the test procedure involved conducting in vivo test on animals which was time consuming and considered unethical. Through a systematic analysis of the past data, a statistical test procedure which uses in vitro data was proposed as an alternative to the in vivo test. Appreciating the case study, the company was provided a great relief by the concerned drug authorities.

Chapter 4 is about the development of a decision support system for a large cement manufacturing company. The problem is about macro level planning of loading the plants of the company situated in different locations with various brands of cement and planning the despatches to the company's stockists based on their monthly orders. The chapter presents only the modelling aspects. The problem is a large-scale optimization problem which is a multi-commodity network flow problem. An important contribution of this study is that it presents a method of finding optimal solution under the unknown fluctuating prices scenario.

Chapter 5 is about deckle optimization in paper and paperboard industries. Paperboard industries invest huge money in procuring software for managing deckle optimization problem. The problem falls under the one-dimensional cutting stock problem and involves solving integer linear programming problems. Intelligent modelling is essential to solve the cutting stock problems arising in the context of deckle optimization. With a brief introduction to the products and processes relevant to this case study, the chapter presents various formulations to handle the challenges. Time and material costs are two conflicting components of the deckle optimization problem. The formulations proposed aid the user to explore the consequences of weighing these two conflicting factors while choosing the strategy for optimal decision making. This is yet another study where a student of a part-time academic programme took it up for his project work.

Chapter 6 was taken up for an information technology enabling services company. It is about planning and managing the agents/associates who receive the calls and address them. It turns out that this problem is too complex to handle manually. The problem is formulated as an integer linear programming problem, and a software tool is developed to aid the management. The tool is Excel based, is simple to use, and can be effectively deployed as a decision support system. This study was initiated by an executive from the company who took it up as his project work for his part-time academic programme.

Chapter 7 is about intelligent modelling of industrial problems. Industries often approach consultants with specific questions or problems. There is a great deal of demand for consultants in this regard. Some problems, though straightforward in terms of formulation, need special techniques for solving them. This chapter presents five live cases of industrial problems, some requiring special solutions and some requiring statistical modelling.

Chapter 8 presents a model for the management of water distribution and scheduling for given availabilities and requirement of water at various crop locations in irrigation ayacut. The chapter provides a model and framework for the problem in question. The problem is formulated as a dynamic minimum cost network flow

problem and provides an approach to solve the problem using static network flow models. A need-based software is also developed to solve the network flow problems. Some issues in the programming are discussed. A postgraduate student was involved in this study.

Chapter 9 deals with the development of an alternative method to measure the land/sea ratio, an important performance measure of tyres. The company's method depends on the skill of a technical person and uses a costly measuring equipment. It takes almost 3 days to get the measurements for each day's samples. The alternative method proposed through the study uses statistical techniques and a simple computer program. The result is instantaneous, removes the subjectivity, and disposes the need of a technical expert and the need of the costly measuring equipment. This study is the project work of a master of science (statistics) student.

Chapter 10 deals with development of an efficient design for conducting a factorial experiment required at an aluminium alloy foundry. Failing to get a satisfactory design for the experiment from the literature, a design was constructed using an ad hoc method. It is transparent from the method of construction that the design provides efficient estimates for all the required main effects and interactions. The method is extended to more general situations and is translated into a systematic approach. The method consists of formulating the construction problem as certain integer linear programming problems. The main purpose of including this study as a chapter in this book is to highlight the fact that research is an important faculty of a successful consultant.

Hyderabad, India
February 2015

G.S.R. Murthy

Acknowledgements

There are a number of people who have contributed directly or indirectly in carrying out the case studies. These people include industrial personnel at various levels, my colleagues and seniors from the Indian Statistical Institute, the students of the academic programmes whom I taught and guided on their project works, a number of industrial executives who were the participants of the part-time programme run by the institute (a potential source for applications), and the organizational staff who helped in various ways. I thank each one of them for their kind cooperation.

Professor Katta Gopalakrishna Murty consolidated my passion for practical OR applications and provided constant support and encouragement. I am privileged to contribute five chapters along with other authors to his edited volume *Case Studies in Operations Research – Applications of Optimal Decision Making* recently published by Springer. Inspired by this, I expressed my desire to publish this volume with Springer. Upon his recommendation, I sent a draft proposal to Camille Price and I got a spontaneous reply from her recommending strongly to publish my case studies with Springer. I hereby record my deepest gratitude to both Professor Katta Murty and Camille Price. Further, the support I have received from Matthew Amboy and Christine Crigler is excellent. My sincere thanks to both of them. Working with the Springer team has been smooth and comfortable.

Much of the success I cherish today I owe it to two persons – my Guru Professor T. Parthasarathy and my wife Haripriya. They have been the spirit behind all my works. I record my deepest gratitude to Professor Parthasarathy. I don't need to thank Haripriya because she is my better half.

G.S.R. Murthy

Contents

- 1 Material Optimization in Web Manufacturing 1**
 - 1.1 Introduction 1
 - 1.2 Background 2
 - 1.3 Process 2
 - 1.4 Problem and Objectives 5
 - 1.5 Approach 6
 - 1.6 Generation of New Cutting Patterns 7
 - 1.6.1 Technical Discussions 9
 - 1.6.2 New Cutting Patterns 9
 - 1.7 Formulation 9
 - 1.8 Analysis and Results 11
 - 1.9 Software Solution 12
 - 1.10 Summary 14
 - Problems 14
 - Appendix 16
 - References 17
- 2 Optimization in PCB Manufacturing 19**
 - 2.1 Introduction 19
 - 2.2 Process 20
 - 2.3 The Decision Making Problem 24
 - 2.4 Formulation 24
 - 2.4.1 TSP Problem 26
 - 2.4.2 Assignment Problem (AP) 26
 - 2.5 The Heuristic 28
 - 2.6 Application 28
 - 2.7 Special Constraints 29
 - 2.8 Software Development 30
 - 2.8.1 Description of PCBSOft 31

2.9	Summary	33
	Problems	33
	References	35
3	A Statistical Procedure for Minimizing In Vivo Tests for Quality Control of Vaccines	37
3.1	Introduction	37
3.2	Description of the Problem	39
3.3	Statistical Analysis	40
3.3.1	Regression Approach	40
3.3.2	Stochastic Comparison Approach	42
3.3.3	Parametric Method	43
3.3.4	Nonparametric Method	45
3.4	Summary of Results and Conclusions	46
	Appendix	48
	References	49
4	Plant Loading and Dispatch Planning	51
4.1	Introduction	51
4.2	The Decision Problem	52
4.2.1	Background	52
4.2.2	Management Decisions	54
4.2.3	Objectives and Scope	55
4.3	Formulation	55
4.3.1	Initial Discussions with Management	56
4.3.2	Planning Horizon	56
4.3.3	Inputs and Notation	56
4.3.4	Decision Variables	60
4.3.5	Objective Function	66
4.3.6	Constraints	66
4.3.7	Complete Formulation	68
4.4	Complexity, Model Simplification and Size Reduction	69
4.4.1	The Case of Assumed Prices	70
4.4.2	The Case of Unknown Prices	70
4.5	Problem Size Reduction	74
4.5.1	The Reduced Problem	75
4.5.2	Effect of Reduction	77
4.6	Plant Loading	78
4.7	Past Data Analysis and Results	80
4.7.1	Basic Objects and the LINGO Model for Example 1	80
4.7.2	Solution of the Reduced Problem for Example 1	83
4.7.3	Application to the Real Problem	86
4.8	Decision Support System	86
4.9	Summary	87
	Problems	87
	References	89

5	Deckle Optimization in Paper Industry	91
5.1	Introduction	91
5.2	Products and the Process	93
5.2.1	Products	93
5.2.2	Winder	95
5.2.3	Deckle	97
5.2.4	Sheeters	97
5.3	Cutting Stock Problem	98
5.3.1	Cutting Patterns	99
5.3.2	Cutting Plans	100
5.3.3	Formulation of One-Dimensional Cutting Stock Problem	100
5.3.4	Formulations for Finding Optimal Solutions	101
5.4	Planning at Winder and Sheeters	108
5.4.1	Production Restrictions	108
5.4.2	Pooling Orders to Suit Reel Diameters	112
5.5	Formulations of the Deckle Optimization Problem	112
5.5.1	Different Options	112
5.5.2	Model for Sheet Orders with 95 cm Diameter	113
5.5.3	Model for Sheet Orders with 100 cm Diameter	117
5.5.4	Model for Sheet Orders Mixed with Both Diameters	119
5.5.5	Model for Sheet Orders Separately	121
5.5.6	Comparison of Different Alternatives	123
5.6	Sequencing Cutting Patterns in Deckle Optimization	124
5.6.1	Adjustment of Knives on the Winder	126
5.6.2	Distance Between Two Cutting Patterns	127
5.6.3	Changeover Time Between Two Cutting Patterns	127
5.6.4	Travelling Salesman Problem for Sequencing Cutting Patterns	128
5.6.5	Optimal Sequence for Deckle Problem	130
5.6.6	Complete Solution to Deckle Optimization Problem	131
5.7	Summary	132
	Problems	132
	References	134
6	A Scientific Tool for Workforce Management in BPO Companies	137
6.1	Introduction	137
6.2	Problem Background	139
6.2.1	Problem Description	139
6.3	Objectives	141
6.4	Approach	142
6.4.1	Call Volumes	143
6.4.2	The Decision Variables	143
6.4.3	Basic Assumptions	143
6.5	Formulation	144
6.5.1	Constraints of the Staffing Problem	144

6.5.2	Two-Stage Optimization and the Objective Functions	145
6.5.3	Formulation of Stage-1 Problem	145
6.5.4	Formulation of Stage-2 Problem	146
6.6	The Software Tool	146
6.6.1	Features	147
6.6.2	Main Screen	147
6.6.3	Inputs Screen	149
6.7	Application of Tool to Live Data	149
6.7.1	Plan for Sep'2010	153
6.7.2	The Company's Solution	153
6.7.3	Stage-1 Optimal Solution	154
6.7.4	Stage-2 Optimal Solution	154
6.7.5	What if the TAT Is Reduced to 8 Days?	155
6.7.6	Dynamic Review	155
6.7.7	What if Resources Are Restricted on Some Days?	155
6.7.8	Analysis of Past Data	156
6.7.9	Calls Answered on a Normal Day	156
6.8	Summary	159
	References	159
7	Intelligent Modeling and Practical Solutions	161
7.1	Introduction	161
7.2	Determination of Polypeptide Composition	162
7.2.1	Background	162
7.2.2	Statement of the Problem	163
7.2.3	Approach and Evolution of Ideas	164
7.2.4	How to Address the Question of Uniqueness?	165
7.2.5	Extension of Ideas to General Problems	166
7.2.6	Solution to the Client's Problem	167
7.2.7	Conclusions of the Peptide Study	168
7.3	Comparing Company's Product with Reference Product	168
7.3.1	Analysis of the Reference Product Data of 39 Observations	169
7.3.2	Principal Component Analysis (PCA)	172
7.3.3	Cluster Analysis (CA)	174
7.3.4	Combined Analysis of the Reference and Company's Product	177
7.3.5	Summary	178
7.4	Estimation of Reliability of CFL Bulbs	183
7.4.1	Background of CFL Problem	184
7.4.2	Life Testing	184
7.4.3	Data Collection	186
7.4.4	Objectives	186
7.4.5	Evaluation of Reliability Function	187
7.4.6	Consultant's Estimate	189

7.4.7	Simulation for Determination of Number of Test Holders . . .	191
7.4.8	Summary	192
7.5	Product Quality Rating System for Paper Boards	196
7.5.1	The Quality Rating Problem	196
7.5.2	Approach	198
7.5.3	Data Collection and Analysis	198
7.5.4	Recommendations	200
7.6	Which Pension Scheme Is Better?	204
7.6.1	The Pensioner's Problem	205
7.6.2	A Solution	205
7.6.3	Markovian Model for the Problem	205
7.6.4	Comparison of Two Schemes	208
7.6.5	Summary of Pensioner's Problem	211
7.7	Chapter Summary	211
	Problems	212
	References	212
8	Modelling Optimum Utilization of Irrigation Resources	215
8.1	Introduction	215
8.2	The Irrigation System	217
8.3	Objectives and Scope of the Project	218
8.4	Formulation and Solution to the Problem	219
8.4.1	The Minimum Flow Constraints	223
8.4.2	The Level Constraint	224
8.4.3	The Group Constraint	225
8.4.4	Handling Priorities	225
8.5	Software Development	225
8.5.1	Minimum Cost Network Flow Problem	226
8.5.2	Resolution of Cycling and Stalling	228
8.6	Summary	229
	References	229
9	Statistics for Measuring Tyre Contact Patch	231
9.1	Introduction	231
9.2	Problem Description	233
9.2.1	Contact Patch and the Land Sea Ratio	233
9.3	The First Solution Using Simulation	234
9.4	Color Spaces	236
9.4.1	RGB Color Space	237
9.4.2	HSB Color Space	237
9.4.3	YUV Color Space	238
9.4.4	XYZ Color Space	239
9.4.5	LAB Color Space	239
9.5	The Principal Component Approach	239
9.6	Approach Using Clustering Technique	242

9.7	Comparison of Results	245
9.8	Summary	245
	References	247
10	A Case Study that Lead to New Designs	249
10.1	Introduction	249
10.2	Background	250
10.2.1	Problem	250
10.2.2	Main Effects and Interactions	251
10.2.3	Model of the Factorial Experiments	252
10.3	Search for Optimal Design	253
10.3.1	Criteria for Optimal Design	254
10.4	Construction Method	255
10.4.1	Construction in 18 Runs	255
10.4.2	Construction in 12 Runs	257
10.4.3	Construction of Design for $2^3 3^3$ with Interactions in 24 Runs	259
10.5	OR Formulations	259
10.5.1	Augmenting with 2-Level Columns	259
10.5.2	Example with a $2^3 3^1$ Experiment	261
10.5.3	Augmenting with 3-Level Columns	262
10.5.4	Enlisting All Solutions	264
10.6	Summary	265
	Problems	265
	References	265
Index		267

Chapter 1

Material Optimization in Web Manufacturing

Abstract This project deals with a problem of high scrap generated in the manufacture of brake shoe webs in one of the bays of a leading automobile industry in Chennai, India. Shoe webs used in vehicle brake systems are blanked from mild steel sheets of 8 mm thickness. The company was facing a high scrap in this production and wanted to explore the possibility of reducing the sheet metal scrap in this area. The production planning is done on a monthly basis as the demand (internal) requirements are on monthly basis. The problem is formulated as that of minimizing sheet-metal consumption. It has been demonstrated, through past data, that huge savings can be made if the planning is done by deploying optimization techniques. A software has been developed for determining the monthly plans optimally. With the help of this software, planning has become extremely simple and has resulted in substantial benefits both in terms of eliminating the efforts of the manager in planning manually and reduction of material costs.

1.1 Introduction

Material optimization studies are a commonplace in applications of Operations Research (OR) methodology. Sheet-metal optimization, often classified under cutting stock problems, has numerous applications [2–4, 7, 8]. This is one such case study in this category. Though this study is a routine application of the cutting stock problem methodology, it reiterates the benefit of scientific approach in reaping rich gains in terms of material savings as well as productivity improvement. This study pertains to manufacturing of brake shoe webs in one of the bays of a leading automobile industry in Chennai, India, for their internal consumption.

1.2 Background

The company manufactures automotive brake systems for all types of vehicles and the products include a wide range of drum and disc brakes and hydraulic actuated brakes. Unit No. 27 of the company, the Press Shop Unit (PSU), is engaged in the production of a number of components of brake systems such as back plates, torque plates, dust covers, rims, servo shells, webs etc. This study pertains to production of webs which are produced from sheet-metal by blanking operation. Though a number of other components are produced from sheet-metal, there was a specific request from the company to look into the problem of high scrap arising out of web production (Fig. 1.1).



Fig. 1.1 Shoe webs in parts of brake subsystems. The *left hand side* picture shows the webs (webs are at the bottom of these components) and the *right hand side* shows after they are assembled into brake systems

1.3 Process

It is important to understand the web production process to see how the huge material savings may be achieved through this study. Shoe webs are semicircular pieces cut from sheet metal (see Figs. 1.2 and 1.3). Process of making shoe webs is briefly described below. This project is concerned with three types of webs. Basically, all the three types have similar shape but their dimensions are different. The three sizes are 420, 433 and 450 mm. These are referred to as 420-webs, 433-webs and 450-webs in the company's terminology (see Fig. 1.3).

Huge sheets of mild steel of 8 mm thickness are bought from companies like SAIL and JSPL. These sheets are usually available in certain standard sizes, viz., $1250\text{ mm} \times L\text{ mm}$, $1400\text{ mm} \times L$ and $1500\text{ mm} \times L$, where the length L (in mm) is either 4800 or 5000 or 6300.

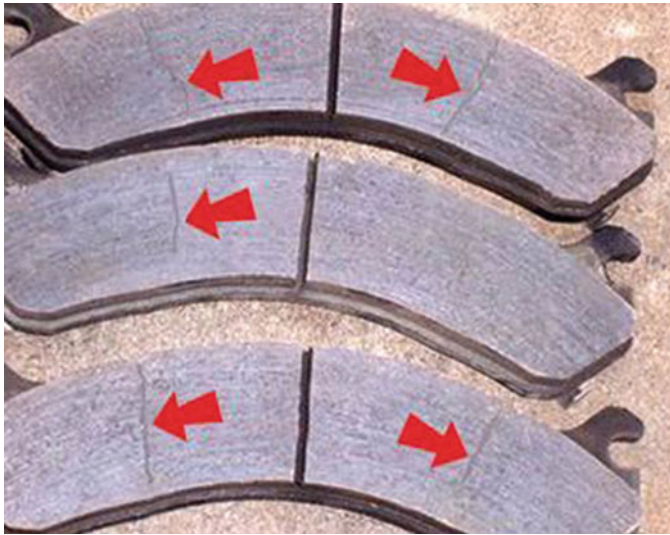


Fig. 1.2 Web when blanked from a sheet metal is in the form of an arc. Picture shows three webs, one below the other

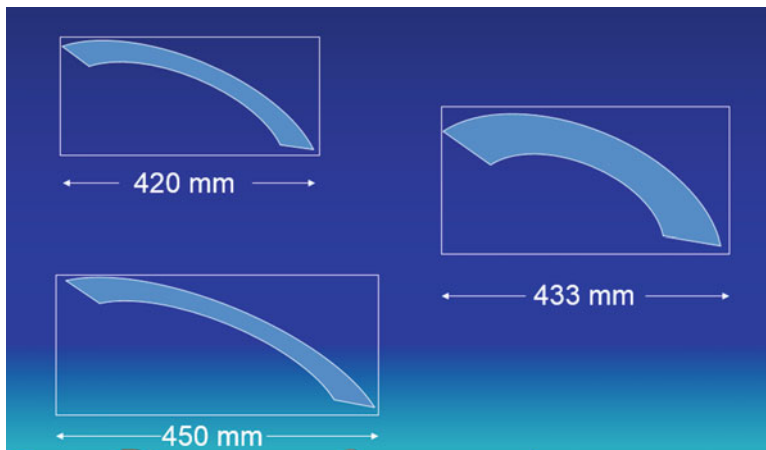
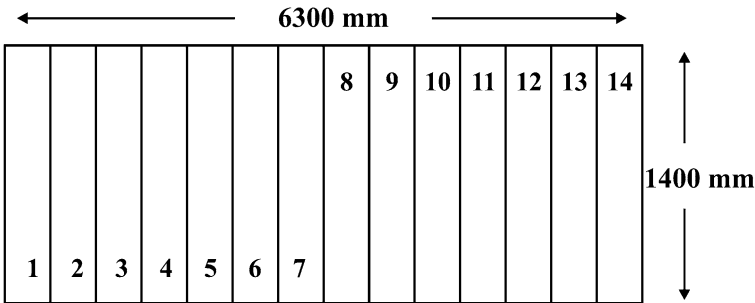


Fig. 1.3 Three different web sizes produced in press shop unit

A cutting pattern is the way in which a sheet-metal is cut to produce webs. One of the cutting patterns used by PSU to produce 450-webs from 1400×6300 mm sheet is P_3 (see Fig. 1.4). This is produced as follows. Using gas-cutting operation, the sheet is first cut into two smaller size-sheets of dimensions 1400×3150 mm and 1400×3150 mm. The length of the sheet cannot exceed 4000 mm as it would lead

to handling problems in the subsequent operation otherwise. Each of the sheets obtained from gas-cutting is then slit into smaller sheets of dimension 1400×450 mm using a shearing machine (this operation is called shearing operation). The width of the *shearing blade* is 2500 mm and it is not possible to shear sheets if their width exceeds 2250 mm (the balance 250 mm goes for certain allowances). We shall refer to the sheets produced from the shearing operation as **blanking sheets**. According to cutting pattern P_3 , a sheet of size 6300×1400 mm is converted into 14 blanking sheets, all of which are used for producing 450-webs, using gas cutting and shearing operations (see Fig. 1.4).



Pieces numbered 1 to 14 are blanking sheets of size 450 mm x 1400 mm each.

Fig. 1.4 Cutting pattern P_3 converts a sheet of size 6300×1400 mm into 14 blanking sheets all of which are used for producing 450-webs. Each of these blanking sheets yields 13 450-webs after blanking

Each blanking sheet is fed into the *blanking machine* by the operator and the die of the machine *blanks* the webs (see Fig. 1.5). In this operation, the number of webs produced from a blanking sheet depends on the length of the blanking sheet. Consider cutting pattern P_3 . According to this, a sheet of size 6300×1400 mm is slit into 14 blanking sheets and each of these blanking sheets yields 13 450-webs. Thus, P_3 yields 182 webs of size 450 mm. The same sheet of size 6300×1400 mm can be cut in a different cutting pattern. According to another cutting pattern, P_4 (see Fig. 1.7), this sheet can be slit into 14 blanking sheets with 7 of them meant for 450-webs and the remaining 7 for 433-webs. This leaves an end piece of size 1400×119 mm which will be a part of the scrap. Cutting pattern P_4 yields 98 433-webs and 91 450-webs. The flow chart of operations of making webs is shown pictorially in Fig. 1.6.

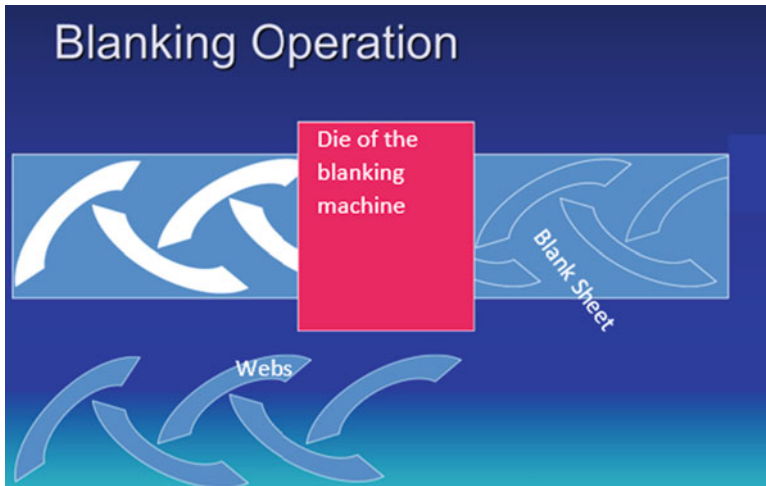


Fig. 1.5 Blanking operation. At the bottom of the figure are the webs after blanking. The right most web on the blanking sheet is also part of the scrap as it does not have the full shape of a web. If only the blanking sheet is little longer, this scrap could have been avoided

1.4 Problem and Objectives

The production planning of webs is done on a monthly basis. The production requirements for every month, referred to as monthly **production schedule**, are supplied to PSU by Brake Assembly Unit well in advance. On the basis of this information and the inventory of sheet-metal, PSU determines the cutting plan for the month. A **cutting plan** specifies what are the cutting patterns to be used and how many of each type. The scrap, the left-over portion of sheets after blanking, is sold at a low price (by weight). Production records revealed that the scrap generated in web production was very high. It was also noticed that cutting plans were being prepared manually based on intuitive calculations. So the need arose to develop a computer software that can determine the cutting plans to be used as well as the sheet-metal requirements. The project was taken up, at the request of the management, with the following objectives:

- (i) To explore the possibility of reducing the material consumption by generating new cutting patterns,
- (ii) To develop a computer software that can be used for (a) planning procurement of sheet-metal based on production requirements so as to minimize the material consumption, and (b) generating optimal cutting plans (for any given production requirements) with available sheet metal.

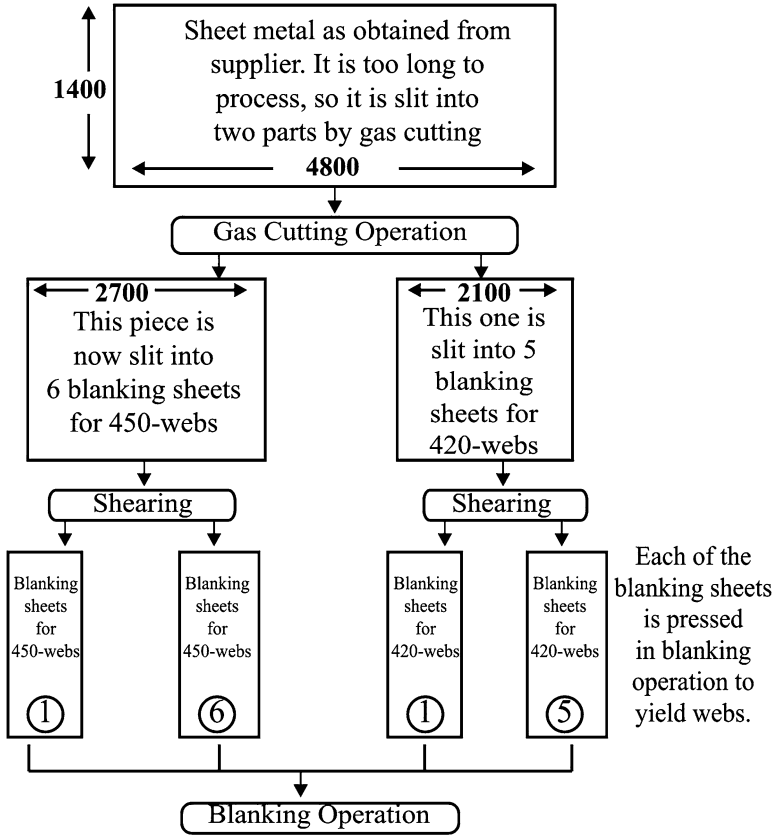


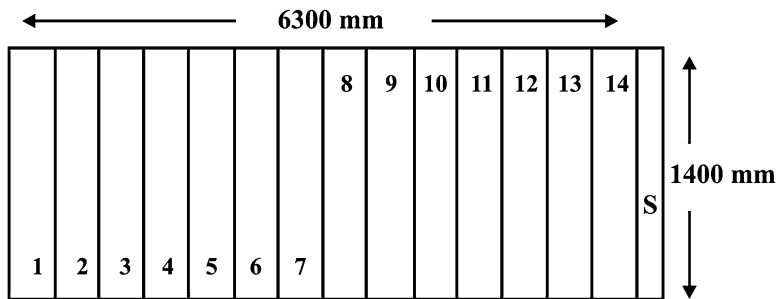
Fig. 1.6 Schematic description of the process

1.5 Approach

The approach here is a typical one. We first examine the existing cutting patterns and augment them with some new patterns by looking at the features specific to this problem and the observation made earlier on the significance of the length of blanking sheet.

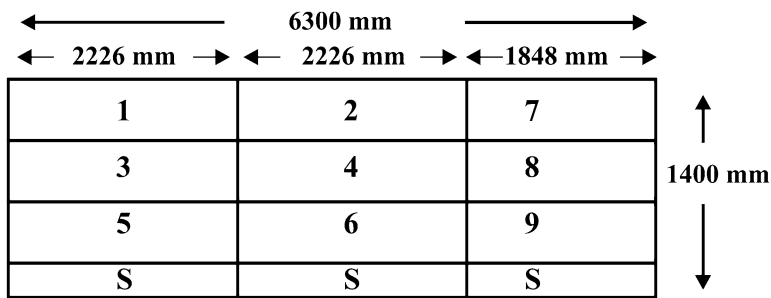
Then, the problem is formulated as a routine integer linear programming problem [1, 5, 6] by considering the material constraints as well as operational constraints. *Obviously the solutions obtained here are not the global optimal solutions because the cutting patterns used here are not exhaustive. Throughout this report, the optimality is used in this restricted sense.*

The effectiveness of the new patterns generated is tested by trying them on the past data. That is, take the production requirements for the past months and work out the optimal cutting plans with and without new cutting patterns and then compare them.



Pieces numbered 1 to 7 are blanking sheets of width 433.
Pieces numbered 8 to 14 are blanking sheets of width 450.
Piece labelled ‘S’ is of width 119 is a part of scrap.

Fig. 1.7 Cutting pattern P_4 . This cutting pattern uses sheet of size 6300×1400 mm. This sheet is slit into 14 blanking sheets, 7 of them meant for 433-webs and the remaining for 450-webs



Pieces numbered 1 to 6 are of size 450mm x 2226mm.
Pieces numbered 7 to 9 are of size 450mm x 1848mm.
The pieces labelled ‘S’ have width 50mm.

Fig. 1.8 Cutting pattern P_{21}

1.6 Generation of New Cutting Patterns

In all the existing cutting patterns (P_1 to P_{20}), the shearing operation is done along the width of the original sheets. Take for example cutting pattern P_3 (see Fig. 1.4). The original sheet of size 1400×6300 mm is slit into 14 blanking sheets each of size 1400×450 mm. Similarly, in cutting pattern P_4 , the same 1400×6300 mm is slit into seven blanking sheets of size 1400×450 mm and seven blanking sheets of size 1400×433 mm. All the existing cutting patterns are such that the width of the original sheet-metal (1400mm in the case of P_3 and P_4) becomes the length of the blanking sheet.

In P_4 , a piece of size 1400×119 mm becomes a part of scrap (see Fig. 1.7). Because of this piece, the planning personnel were under the impression that P_3 was efficient pattern but P_4 was not (because there is no piece which is scrapped). However, this is a misleading notion that had set in the minds of the planning personnel. To see this, notice that the blanking sheets of P_3 having a length of 1400 mm yield 13 450-webs. But even a blanking sheet of length 1346 mm will also yield 13 450-webs. So, in P_3 , effectively used portion of the sheet-metal is of size 1346×6300 mm. The remaining piece of size 54×6300 mm essentially goes as a part of the scrap.

An Important Observation for Yield

The length of a blanking sheet has a significant influence on the yield of a cutting pattern. For example, to produce one 450-web, the length of the blanking sheet should be a minimum of 290 mm and for every additional web, an additional length of 86 or 90 mm in the blanking sheet is required. Thus, for two webs the length required is 376 mm, for three webs the length required is 466 mm and so on. At the time of initiating this project, the company was using 20 cutting patterns. These plans are labeled as P_1 to P_{20} and their yields are shown in Table 1.2. The length of blanking sheet of any of these cutting patterns is either 1250 or 1400 or 1500 mm. Consider any blanking sheet of length of length 1400 mm. This yields 13 450-webs. For 13 450-webs, the length required is 1346 mm and for 14 450-webs the length required is 1432 mm. Thus, the length of a blanking sheet determines the number of webs per blanking sheet. Table 1.1 summarizes the yields of blanking sheets of different lengths for the three types of webs. *Therefore, the length of blanking sheet plays a crucial role in the generation of effective cutting patterns. It is this observation that has led to potential savings in this project.*

Now consider a new cutting pattern P_{21} shown in Fig. 1.8. This is obtained by slitting the 1400×6300 mm size sheet into three pieces of sizes 1400×2226 mm, 1400×2226 mm and 1400×1848 mm. Then, using shearing operation, convert them into six blanking sheets of size 450×2226 mm and three blanking sheets of size 450×1848 mm. This is feasible as the shearing operation is possible up to a maximum length of 2250 mm (see Sect. 1.3).

Using the blanking operation, we can extract 23 450-webs from a blanking sheet of size 450×2226 mm (in fact 2226 mm is the minimum length to produce 23 450-webs, see Table 1.1) and 18 450-webs from the blanking sheet of size 450×1848 mm (the minimum length required to extract 18 450-webs is, however, 1786 mm only). Thus, we get a total of 192 webs of size 450 mm from P_{21} . In contrast, the number of 450-webs that can be extracted from P_3 is only 182. This is because from a blanking sheet of size 450×1400 mm we can extract only 13 450-webs (the minimum length required to extract 13 450-webs is 1346 mm). Also, the productivity is more with P_{21} as the number of blanking sheets to be used is only 9 as opposed to 14 for the P_3 cutting pattern.

1.6.1 Technical Discussions

It has been noted that P_{21} is a better cutting pattern compared to P_3 . However, one should explore the feasibility of implementing P_{21} from production view point. Management discussed this point in a meeting involving a group of technical personnel. One of them expressed the doubt that the formation of granule structure in the sheet-metal might prohibit cutting the blanking sheets across the length as proposed in P_{21} . But after the deliberations, it was concurred that this would not pose any problem. One more issue raised in this connection was that even though the shearing machine is designed to shear sheets of lengths up to a 2.5 m, trials indicated that shearing sheets of length beyond 1.8 m was causing a bending in the blanking sheets which is undesirable. It was agreed that this problem could be sorted out without much difficulty and that implementing P_{21} could be made feasible.

1.6.2 New Cutting Patterns

Encouraged by the management's decision to make the new type of cutting patterns feasible, nine new cutting patterns - numbered from P_{21} to P_{29} - were evolved. The ready reckoner Table 1.1 was prepared to aid the development of new cutting patterns. This table gives the minimum length of a blanking sheet required to produce a given number of webs. The yields of the new cutting patterns as well as those of the existing ones are summarized in Table 1.2.

1.7 Formulation

To formulate the problem we shall introduce relevant notation. Let S_1, S_2, \dots, S_m be the distinct sheet sizes (by dimensions) used by the cutting patterns P_1 to P_{29} . Let A_i be the area of sheet of size S_i , $i = 1, 2, \dots, m$. Define the sets $\alpha_i = \{j : P_j \text{ uses sheet of size } S_i\}$, $i = 1, 2, \dots, m$. Let a_{j1} be the number of 420-webs that are extracted from cutting pattern P_j . Similarly, define a_{j2} and a_{j3} for the 433-webs and 450-webs respectively. Then the matrix $Y = ((a_{jk}))$ is the yield matrix where the j th row of Y is the yield of the cutting pattern P_j , $j = 1, 2, \dots, 29$. Last three columns of Table 1.2 present the yield matrix Y .

Consider the production requirements of an arbitrary month. Let w_1 be the number of 420-webs required for the month, and let w_2 and w_3 be the corresponding quantities for 433- and 450-webs respectively. Let q_i be the number of sheets of size S_i available for the month, $i = 1, 2, \dots, m$. Let x_j be the number of sheets to be cut according to cutting pattern P_j , $j = 1, 2, \dots, n$, where n is the number of all cutting patterns under consideration.

Table 1.1 Length table for yield of webs

No. of webs	Min. length (mm)			No. of webs	Min. length (mm)		
	420	433	450		420	433	450
1	260	259	290	13	1310	1267	1346
2	350	338	376	14	1400	1346	1432
3	435	427	466	15	1485	1435	1522
4	525	506	552	16	1575	1514	1608
5	610	595	642	17	1660	1603	1698
6	700	674	728	18	1750	1682	1784
7	785	763	818	19	1835	1771	1874
8	875	842	904	20	1925	1850	1960
9	960	931	994	21	2010	1939	2050
10	1050	1010	1080	22	2100	2018	2136
11	1135	1099	1170	23	2185	2107	2226
12	1225	1178	1256	24	2275	2186	2312

Given w_1 , w_2 and w_3 , the total area of these webs alone is a constant, say X . Let M be the total area of the sheet metal used. As sheet thickness is a constant, scrap is directly proportional to $(M - X)$. Therefore, for given w_1 , w_2 and w_3 , minimizing the scrap is equivalent to minimizing the total area of the sheet-metal, M . Therefore, the decision problem is to determine the vector $x = (x_1, x_2, \dots, x_n)$ so as to

$$\begin{aligned} & \text{minimize } \sum_{i=1}^m A_i \sum_{j \in \alpha_i} x_j \\ & \text{subject to} \\ & \sum_{j=1}^{29} a_{jk} x_j \geq w_k, \quad k = 1, 2, 3, \end{aligned} \tag{1.1}$$

$$\sum_{j \in \alpha_i} x_j \leq q_i, \quad i = 1, 2, \dots, m, \tag{1.2}$$

x_j 's are nonnegative integers.

We shall refer to the vector x as monthly cutting plan (MCP). In the above formulation, the objective function is the total area of the sheets consumed by the MCP x . The first set of constraints (1.1) ensures that x meets the schedule requirements while the second set of constraints (1.2) ensures that x uses only the available sheets for the month.

Management asserts that the availability of standard size sheets is usually not a problem. *Therefore, the above formulation can be used to plan the procurement of standard size sheets by solving the problems without the second set of constraints.*

Table 1.2 Details of cutting patterns

Cutting pattern	Sheet size (m × m)	Sheet code	Sheet area (m ²)	Number of webs		
				420	433	450
P_1	4.8 × 1.4	S_1	6.720	70	0	78
P_2	4.8 × 1.4	S_1	6.720	70	84	0
P_3	6.3 × 1.4	S_2	8.820	0	0	182
P_4	6.3 × 1.4	S_2	8.820	0	98	91
P_5	6.3 × 1.4	S_2	8.820	210	0	0
P_6	5.0 × 1.4	S_3	7.000	0	0	143
P_7	5.0 × 1.4	S_3	7.000	0	70	78
P_8	5.0 × 1.4	S_3	7.000	0	154	0
P_9	6.3 × 1.25	S_4	7.785	0	0	154
P_{10}	6.3 × 1.25	S_4	7.785	0	84	77
P_{11}	6.3 × 1.25	S_4	7.785	180	0	0
P_{12}	5.0 × 1.25	S_5	6.250	0	0	111
P_{13}	5.0 × 1.25	S_5	6.250	0	60	66
P_{14}	5.0 × 1.25	S_5	6.250	0	132	0
P_{15}	4.8 × 1.25	S_6	6.000	60	72	0
P_{16}	4.8 × 1.25	S_6	6.000	60	0	66
P_{17}	6.3 × 1.5	S_7	9.450	0	0	196
P_{18}	6.3 × 1.5	S_7	9.450	0	105	98
P_{19}	6.3 × 1.5	S_7	9.450	225	0	0
P_{20}	4.8 × 1.4	S_1	6.720	0	154	0
P_{21}	6.3 × 1.4	S_2	8.820	0	0	192
P_{22}	5.0 × 1.4	S_3	7.000	0	0	151
P_{23}	6.3 × 1.4	S_2	8.820	135	66	0
P_{24}	4.8 × 1.4	S_1	6.720	0	150	0
P_{25}	6.3 × 1.4	S_2	8.820	0	204	0
P_{26}	5.0 × 1.4	S_3	7.000	0	159	0
P_{27}	5.0 × 1.4	S_3	7.000	0	51	102
P_{28}	5.0 × 1.4	S_3	7.000	0	0	147
P_{29}	6.3 × 1.4	S_2	8.820	0	96	96

1.8 Analysis and Results

In Sect. 1.6 it was pointed out that there is an impression that some of the existing cutting patterns are efficient (P_3 for example) and some are not. Actually, the efficiency depends on the requirements. It might happen that for a given requirement, P_3 may not be in the selection of optimum cutting plan. This aspect will be observed in this section. Further, the scope for minimizing the sheet metal with and without the inclusion of new cutting patterns is explored. The savings with the inclusion of new cutting patterns is quite substantial.

Consider the monthly schedules for one full year from April to March presented in Table 1.3. We shall work out the MCPs for each of these months in two ways: (i) by using only the existing cutting patterns (i.e., P_1 to P_{20}) and (ii) by using all the cutting patterns (i.e., P_1 to P_{29}). Since it was difficult to get records on sheets' availability (i.e., q_i 's), the problems are solved by dropping the sheets' availability constraints (1.2).

The problems have been solved by using the formulation given in the earlier section (with appropriate modifications) and the professional OR package LINGO (www.lindo.com/products/lingo/). The results are summarized in Table 1.4.

We see from Table 1.4 approximately 100,000 rupees can be saved every month by considering all the patterns.

It must be noted that in the above comparison, we have compared the best solutions using P_1 to P_{20} with the best solutions using P_1 to P_{29} . However, at the time of initiating this project, the MCPs actually used by the company are not the best solutions using P_1 to P_{20} . This is because the MCPs were worked out manually based on the intuitive calculations and not by using the OR formulation. We shall now analyze the actual MCPs used. Consider the data for the month of August. While the

Table 1.3 Monthly schedules during Apr to Mar'97

Month	Requirements			Month	Requirements		
	420	433	450		420	433	450
Apr	45800	32000	33800	Oct	43000	40000	42000
May	50000	36000	21000	Nov	43000	21000	46000
Jun	36000	36800	45000	Dec	38000	21000	41000
Jul	28000	43000	50000	Jan'97	39000	15200	40000
Aug	34000	42000	44000	Feb'97	50000	0	56000
Sep	40000	40000	40000	Mar'97	40000	0	40000

schedule for this month specified was $w_1 = 39000$, $w_2 = 42000$ and $w_3 = 44000$, the actual production figures were 38475, 42602 and 40309 respectively. The actual MCP, sheets used and total sheet sizes used are given Table 1.5.

Taking the actual production figures as requirements, optimal plans were worked out under four situations: (i) using only P_1 to P_{20} with sheet availability restricted to sheets actually used, (ii) using only P_1 to P_{20} with no restrictions on sheets' availability, (iii) using P_1 to P_{29} with sheets' availability restricted to actual sheets used and (iv) using P_1 to P_{29} with no restrictions on sheets' availability.

Table 1.6 presents a comparative picture. A notable observation is that if we allow all P_1 to P_{29} patterns, there is a huge reduction in the total sheet area even when the sheets' availability is restricted to sheets actually used (sheet area reduces to 5392.10 from 5500.6 m²).

1.9 Software Solution

Besides optimizing on the material consumption, the major benefit of this project is that the process of planning is simplified. Originally, a software was provided to determine the MCPs. This software used a professional OR package. Recently, a solution to the problem is developed in Microsoft Excel which makes things much

Table 1.4 Savings with the use of new cutting patterns

Month	Total sheet area (m ²)		Savings	
	Using only	Using all	Area	Material cost
	P_1 to P_{20}	P_1 to P_{29}	(m ²)	(Rs.)
Apr	4945.29	4862.87	82.42	83108
May	4680.16	4626.42	53.74	54189
Jun	5283.37	5174.63	108.74	109648
Jul	5456.54	5336.33	120.21	121214
Aug	5375.90	5271.58	104.32	105191
Sep	5350.51	5254.50	96.01	96812
Oct	5571.22	5471.01	100.21	101047
Nov	4938.17	4831.64	106.53	107420
Dec	4489.97	4391.12	98.85	99675
Jan'97	4230.24	4135.69	94.55	95339
Feb'97	4806.28	4674.04	132.24	133344
Mar'97	3610.75	3519.87	90.88	91639

Table 1.5 Actual MCP used in Aug

Cutting pattern	Sheet		No. of sheets used	No. of webs produced		
	Size	Code		420	433	450
P_3	6.3×1.4	S_2	21	0	0	3822
P_4	6.3×1.4	S_2	34	0	3332	3094
P_6	5.0×1.4	S_3	43	0	0	6149
P_8	5.0×1.4	S_3	255	0	39270	0
P_{17}	6.3×1.5	S_7	139	0	0	27244
P_{19}	6.3×1.5	S_7	171	38475	0	0
Total				38475	42602	40,309

Total sheet area consumed = 5500.6 m²

Table 1.6 Comparison of optimal MCPs with the actual MCP

Item	Actual	Best of P_1 to P_{20} constraints		Best of P_1 to P_{29} constraints	
		With	Without	With	Without
No. of webs					
w_1	38475	38475	38480	38475	38655
w_2	42602	42630	42616	42678	42606
w_3	40309	40325	40324	40314	40320
Total sheet area	5500.6	5500.6	5422.0	5392.1	5319.4
Cost (Rs)	5546530	5546530	54567299	5437124	5363822
Savings ⁺ (Rs)		0	79231	109406	182708

Data pertains to August

‘With constraints’ means availability of sheet-metal is restricted as given in Table-5; and

‘Without constraints means no restriction on the availability of sheet-metal

+ comparison against the actual cost (= cost – 5546530)

easier and simpler for the users so that it is accessible to wide range of users. The Excel solution uses the excel solver and a user-friendly template is designed for the problem. The template is shown in Fig. 1.10. The template is self explanatory. The problem can be programmed using excel solver and the program can be

linked to the push button ‘Solve.’ Up on giving all the required inputs, the solution will be flashed in the template sheet within seconds. With the help of this excel solver, what used to take a day or two for the manager to plan the MCP takes less than 5 min now.

1.10 Summary

In this case study, we have taken up the problem of reducing scrap of sheet-metal in the manufacture of brake shoe webs. The problem is studied as the routine cutting stock problem. Designing nine new cutting patterns, it is shown, through past data, that the material savings could be substantial if the new cutting patterns are adopted and the OR methodology is followed.

The present method of preparing the material requirements and the corresponding cutting plans manually is laborious and far from optimality. Even with the existing cutting patterns (P_1 to P_{20}), substantial savings (Rs.80000/- for August) could have been made had the material planning been done using OR methodology.

Further, we see that huge savings in material costs (Rs.182000/- for August) can be achieved by supplementing the existing cutting patterns with the new ones. A distinct feature of the new cutting patterns is that the number of shearing operations as well as the number of blanking sheets to be processed is less in the case of new cutting patterns compared to the existing ones. Hence, using new cutting patterns will be more productive.

The biggest benefit that the company derived through this project is the excel solver for making the monthly production plans for the shoe manufacturing.

Problems

1.1. A metal fabric industry gets orders for copper disks of three different diameters, namely, 11, 17 and 23 cm. These are cut from rectangular busbars of standard size 100×60 cm. The disks are obtained by a blanking operation. Busbar costs Rs.10000/- per square meter (we take the cost per square meter as the thickness of the busbar remain same which is 8 mm) and the scrap (the leftover material after blanking) is sold at Rs.4000/- per square meter. One order received is 150 pieces of 23 cm disks, 225 pieces of 17 cm disks and 90 pieces of 11 cm disks. Figure 1.9 shows a cutting pattern for this problem.

- (a) Generate 10 cutting patterns for this problem.
- (b) Using your 10 cutting patterns, formulate the problem and solve it.
- (c) Develop a nice template in excel to solve the above problem with your cutting patterns so that it can be used for any order quantities.

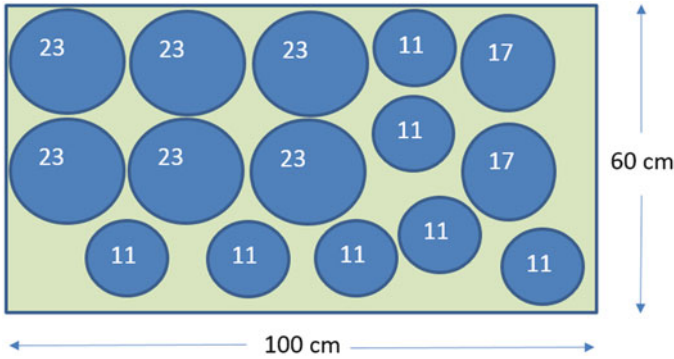


Fig. 1.9 A cutting pattern for copper disks

- (d) The actual problem has the following restrictions. Disks can be cut only from square shapes. That is, to cut a disk of diameter d cm, it is necessary to first cut a square of side $(d + 0.3)$ cm, and then blank the disk from the square. The additional 0.3 cm here is required for cutting allowance. Generate the cutting patterns for this actual problem. Formulate the problem and solve it.
- (e) The process to produce these disks in the company is to use the standard bars separately for each disk size. In other words, use k_d sheets for producing disks of diameter d , $d = 11, 17, 23$. Determine the smallest values of k_d s. Compare your solutions with this solution of the company.
- (f) Consider the problem in the following situation. Like in the shoe web problem, suppose the process of producing disks is to cut the master sheet into long bars of size $100 \times (d + 1)$ cm, where d is either 11, 17 or 23, and then blank disks from each of these bars so that from bar of size $100 \times (d + 1)$ cm disks of diameter d are blanked. Develop a method in which this problem can be converted into a one-dimensional cutting stock problem. This requires a good research into the problem. Try.

1.2. There was another sheet metal optimization study that was carried out at Chennai, India. This was for the Integral Coach Factory which produces railway coaches. There were 14 sheet metal parts that go into the windows of the coaches. Try and see if you can get shapes and sizes of such parts from the net and carry out a project.

References

1. Chen, D.-S., Batson, R.G., Dang, Y.: *Applied Integer Programming: Modeling and Solution*. Wiley, Hoboken (2010)
2. Farley, A.: Mathematical programming models for cutting stock problems in the clothing industry. *J. Oper. Res. Soc.* **39**, 41–53 (1988)
3. Farley, A.: The cutting stock problem in the canvas industry. *Eur. J. Oper. Res.* **44**, 247–255 (1990)
4. Haessler, R.W.: A new generation of paper machine trim programs. *TAPPI J.* **71**, 127–130 (1988)
5. Nemhauser, G.L., Wolsey, L.A.: *Integer and Combinatorial Optimization*. Wiley, New York (1998)
6. Papadimitriou, C.H., Steiglitz, K.: *Combinatorial Optimization: Algorithms and Complexity*. Dover, Mineola (1998)
7. Sweeney, P.E., Haessler, R.W.: One-dimensional cutting stock decisions for rolls with multiple quality grades. *Eur. J. Oper. Res.* **44**, 224–231 (1990)
8. Sweeney, P.E., Paternoster, E.R.: Cutting and packing problems: a categorized, application-oriented research bibliography. *J. Oper. Res. Soc.* **43**, 691–706 (1992)

Chapter 2

Optimization in PCB Manufacturing

Abstract This project was taken up for M/s Sree International Limited, Chennai, and it deals with a case study in which we try to optimize the manufacturing process of mounting components on bare printed circuit boards. Following the ideas and heuristic suggested in Chang et al. (Optimizing the radial component insertion on PCB's. In: Yu G (ed) Industrial applications of combinatorial optimization. Kluwer, Boston, 1997) the problem of minimizing production time is formulated as a combination of traveling salesman and assignment problems. Some special difficulties encountered in the current problem are tackled by formulating them using linear constraints. The most important contribution of this project is the development of a software package, a decision support tool, for making decisions directly related to productivity.

2.1 Introduction

It is hard think of any electronic gadget without a printed circuit board (PCB). Cell phones, calculators, computers, television sets, aircraft engines, what not; every one of these uses PCBs. A PCB is a flat plate or base of insulating material containing a pattern of conducting material and components such as resistors, capacitors, connectors, etc. placed and soldered to it. It is an electrical circuit. The conducting material is commonly copper which is coated with solder or plated with tin or tin-lead alloy. The usual insulating material is epoxy laminate. A PCB without components is called a printed wiring board and this shall be referred to as **bare PCB** in this chapter (Fig. 2.1). PCBs can be either single-sided or double-sided. Single-sided boards, the most common style in mass-produced consumer electronic products, have all components on one side of the board. With two-sided boards, the conductors, or copper traces, can travel from one side of the board to the other. This project deals with single-sided boards. Components are mounted or inserted using automatic machines.

Combinatorial optimization [4, 6, 7] plays an important role in manufacturing of PCBs. Chang, Hwang and Murty [3] proposed a heuristic to optimize the radial component insertion operations in the manufacturing process of printed circuit boards. The process in their study involved mounting radial components into holes in PCB through a computer controlled machine. The objective being minimization of production time, the decision problem was: (i) sequencing the holes in the PCB for insertion, and (ii) the assignment of component types to stations that hold them. Giving a general formulation, the authors of [3] proposed a heuristic to solve problems with similar structure. The method involves solving an alternating series of traveling salesman and station assignment problems [5, 8].

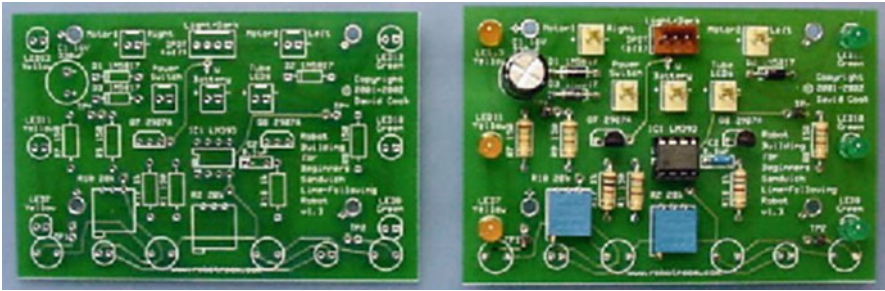


Fig. 2.1 Printed circuit boards: The picture on the *left* is bare PCB and the one on the *right hand side* is a PCB

In this project, we apply this heuristic to a similar problem that we have encountered at M/s Sree International Limited, Chennai, an ancillary to printed circuit board manufacturing companies. The company is engaged in converting bare PCBs received from their clients into PCBs by *populating* them with components through surface mounting operations. The process under consideration yields some simplification in the methodology. The heuristic in this context reduces to solving travelling salesman and assignment problems. Further restrictions of the process in certain special cases are also tackled by a special formulation (see Sect. 2.7).

2.2 Process

One of the operations in the manufacture of PCBs is the *stuffing* operation. In this operation, various types of components such as resistors, diodes, ICs etc. are mounted on the surface of bare PCBs. The PCBs thus produced are known as stuffed PCBs or simply PCBs. Mounting is done by a computer controlled automatic machine. The machine comprises feeders, a head and a flat board, and is connected to a micro processor. The machine has a flat table and a movable head. One end of the table consists of several feeders. Bare PCB is placed on the flat table at specified location on the table. A picture of one such machine is shown in Fig. 2.2. Feeders are loaded with strips of components so that each feeder holds only one type of compo-

nents. When the operation is started, the machine head moves from its *home position* to a feeder, picks up a component and places it on the PCB at the specified location on the bare PCB. The placement is done one component at a time. That is, the head picks up a component, places it at its specified location and then goes to the feeder where it has to pick the next component, picks up the next component, mounts it at the specified location, then goes for the next component and so on. After placing all components, the head returns to its home position from where it started its operations. One such mounting operation can be seen in action in the video clipping at <https://www.youtube.com/watch?v=RjoxCprf3Kk> (cited on 31 Oct 2014).

Definition 2.1. Feeder Location. Each feeder has a *front position*, called the feeder location, from where the head picks up a component. The moment a component is picked up from a feeder, the next component on the strip of that feeder moves forward to the front position of the feeder. It is important to understand this operation from the view point of modeling this problem. In Fig. 2.2, the feeder locations are at the heads of the arrows showing feeders.

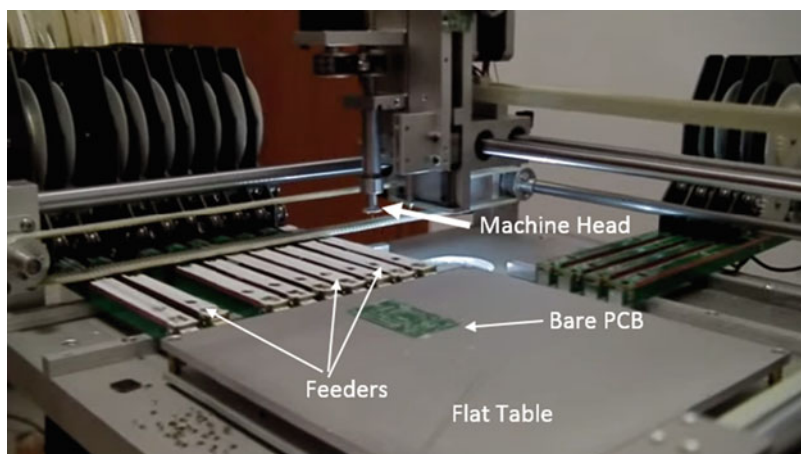


Fig. 2.2 A PCB component mounting machine with feeders and a head

The number of components to be mounted on PCB typically varies from 50 to 200 for the PCBs processed at the company. To understand the decision making problem, a miniature example is chosen for discussion. Imagine a PCB has to be mounted with seven components with three identical components of one type (call these as type A components), three identical components of another type, say type B, and one component of a third type, say type C. The components are embedded in strips and for each type there is a strip of components of that type. Imagine that the machine has eight feeders, labeled as F_1, F_2, \dots, F_8 (in the actual machine at the company the number of feeders is 24). The set up of operations is shown pictorially in Fig. 2.3. In this set up, a strip of type A components is loaded on feeder F_3 , strip of type B components on feeder F_5 and strip of type C components on feeder F_7 .

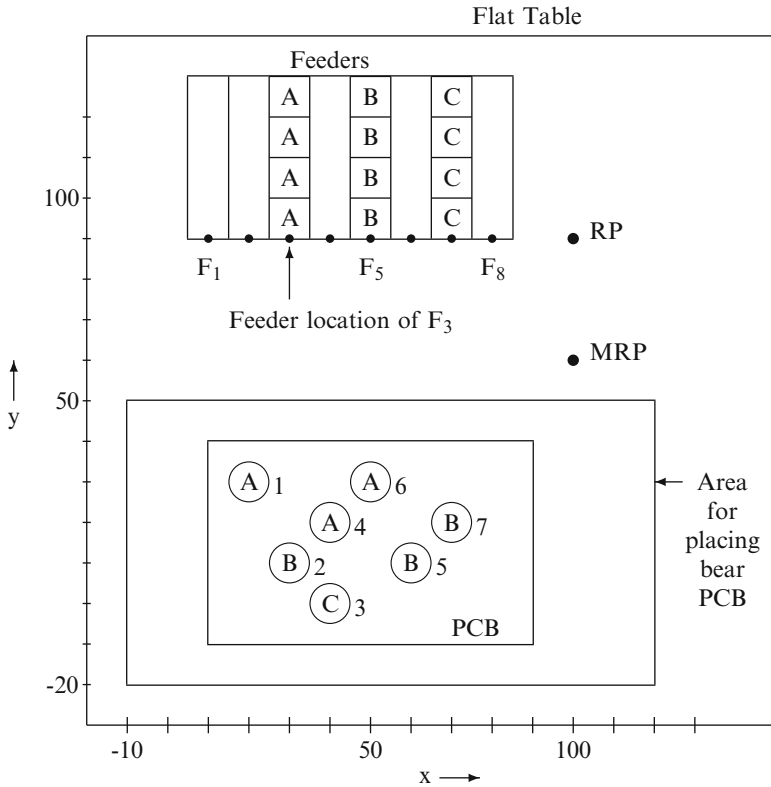


Fig. 2.3 Schematic picture of PCB mounting machine

Definition 2.2. Feeder Allocation. Assigning types (of component strips) to feeders is called *feeder allocation*.

In the example under discussion, type A is assigned to feeder F_3 , type B to feeder F_5 and type C to feeder F_7 . It is required that the seven components are to be placed at specified locations on the bare PCB. These locations are identified with two-dimensional coordinates with reference to a location point on the flat table treated as origin $O(0,0)$. This is the reference point to all coordinates in the two-dimensional plane. The bare PCB on which these components are to be mounted is placed at a particular location with reference to the origin $O(0,0)$. The locations where the components are to be placed are specified by the two-dimensional coordinates. All these locations are identified uniquely with reference to $O(0,0)$. Also, the coordinates of feeder locations and the home position of the machine head are with reference to $O(0,0)$. In Fig. 2.3, MRP (machine reference point) stands for the home position of the head. Feeder locations are shown with '•'. Take the component locations as the centers of the respective circles in the figure.

In the actual operation, the machine head moves from MRP to RP and then moves to a feeder location. This happens only for the first component to be mounted.

For the second component onwards, the machine head moves directly from component location to feeder location to pick up a component and from feeder location to a component location to place a component.

Production of PCBs

The company receives orders in batches. That is, each client will send a lot of bare PCBs and the company's job is to mount the components on each of the bare PCBs in the lot. PCBs in each lot are identical in nature. The operations for converting one bare PCB into a PCB, is same for every bare PCB in the lot.

To convert a bare PCB into a PCB the following inputs are necessary and these are given as inputs to the micro processor to which the machine is connected.

1. Coordinates of home position of the head.
2. Coordinates of the feeder locations.
3. Coordinates of each component location along with the type of the component.

In addition to the above, the manager has to specify the feeder allocation and a mounting sequence (that is, the order in which the components are to be mounted) to the micro processor. Once this is done, a bare PCB is placed and the operation (production) is started. We shall illustrate the operation with our example. Refer to mounting sequence and feeder allocation shown in Fig. 2.3. The mounting sequence is specified by the numbers 1–7 against the component locations. To start with, the machine head is in its home position **MRP** and when the operation starts, the head moves to **RP** (known as reference point). Since the first component to be mounted is of type A, the machine head, then, moves from **RP** to F_3 , picks up one component, moves to location 1 (that is where the first component is to be mounted) and mounts the component there (when a component is mounted it gets fixed there automatically). Recall that when a component is picked up from a feeder, the next component in that feeder moves forward automatically and places itself at the feeder location.

Since the next component to be mounted is a type B component, the machine head moves (from location 1) to F_5 , picks up a component, moves to location 2 and mounts it there. Next, the head moves to F_7 , picks up a component and mounts it in location 3. This way all the components are mounted. After mounting the last component, the head goes back to its home position. The PCB is then replaced with another bare PCB and the operation is repeated. *The time taken for mounting all the components is directly proportional to the distance traversed by the machine head.*

The number of components to be mounted on a PCB depends on the type of PCB. For most PCB types, it varies between 75 and 100 but this number can be as small as 50 and as large as 200.

2.3 The Decision Making Problem

The company receives orders on a variety of PCBs. Depending on the client and the order, the lot size varies from 1000 to 5000. Consider an order for 3000 PCBs, that is, the lot size is 3000. If 1 min can be saved for mounting components on one bare PCB, it amounts to saving 3000 min of production time. Thus, the problem of reducing production time on one PCB results in potential savings. Since the time for mounting the components is directly proportional to the distance traversed by the head, reduction in time can be achieved by minimizing the distance covered by the head. Thus, the problem is to minimize the distance covered by the head in mounting all components on one bare PCB for any given lot. Further if the problem is solved for a given PCB, then the same solution can be applied to all repeated orders on that PCB in future.

Prior to this study, the feeder allocation and the mounting sequence were being determined by the manager using his intuition after a physical examination of the locations. For the purpose of demonstration, we applied the methodology to a PCB with 88 components comprising of 39 types to be mounted. The company carried out this mounting in three separate stages. The 88 components were partitioned into three separate batches, and each batch was processed separately. The first batch (Batch 1) consisted 52 components comprising 14 types, the second batch (Batch 2) consisted 19 components of 13 types and the third batch (Batch 3) consisted 17 components of 12 types. Part of the input data for Batch 1 components is shown in Fig. 2.4.

2.4 Formulation

Conceptually this problem is similar to the one considered in [3]. In this section we first formulate the problem in a general set-up and then formulate two subproblems in Sects. 2.4.1 and 2.4.2. These formulations are used to find local optimal solutions to the problem.

Suppose we have m feeders and n components comprising p types. Let the feeders be denoted by F_1, F_2, \dots, F_m and let the types of components be denoted by T_1, T_2, \dots, T_p . Then, any feeder allocation a is an $m \times p$ matrix whose (i, j) th element is given by

$$a_{ij} = \begin{cases} 1 & \text{if } T_j \text{ is assigned to } F_i, \\ 0 & \text{otherwise.} \end{cases}$$

Any mounting sequence $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ is a permutation of $1, 2, \dots, n$.

Clearly, the distance covered by machine head in mounting the components is a function of mounting sequence π and feeder allocation a only. Let this function be denoted by $f(\pi, a)$. Then, $f(\pi, a)$ is the total distance *travelled* by the head to mount all the components. The optimization problem is to find a π^* and an a^* such that

Basic Data for Scheduling PCB Surface Mounting Operations

No of components	: 52
No of feeders	: 24
No of types	: 14
MRP coordinates (x,y)	: (-112.55, -40.82)
RP coordinates (x,y)	: (-169.13, -100.65)

Component Types and Coordinates				Feeder Coordinates		
Component		Coordinates		Feeder	Coordinates	
No.	Type	x	y	Number	x	y
1	1	40.25	0.30	1	-166.60	-97.92
2	1	14.30	5.30	2	-151.60	-97.92
3	2	74.26	7.50	3	-136.60	-97.92
4	2	67.55	7.40	4	-121.60	-97.92
5	2	85.32	-2.63	5	-106.60	-97.92
⋮	⋮	⋮	⋮	⋮	⋮	⋮
52	14	3.90	1.29	24	178.40	-97.92

Fig. 2.4 Basic data for scheduling PCB surface mounting operations

$$f(\pi^*, a^*) = \min_{\pi, a} f(\pi, a). \quad (2.1)$$

A (π^*, a^*) satisfying the above property will be called a global optimal solution to the problem in question. Finding a global optimal solution to the problem is an NP-hard problem [1] and it is practically impossible to find one. However, *local optimal* solutions can be defined and computed using existing algorithms and professional OR solvers. But this needs an approach. As mentioned earlier, in this project we used the approach suggested in [3] to find local optimal solutions. The approach uses two standard OR models – the traveling salesman problem and the assignment problem [8] – and depends on the following crucial observation.

Crucial Observation

The problem of minimizing $f(\pi, a)$ does not fall into any of the standard models if we try to find π and a simultaneously. But if the mounting sequence π is fixed, then $f(\pi, a)$ is only a function of a and minimizing $f(\pi, a)$ for fixed π is an assignment problem. This problem is formulated in Sect. 2.4.2. On the other hand, if a feeder

allocation a is fixed, then $f(\pi, a)$ is only a function of the mounting sequence π and minimizing $f(\pi, a)$ for fixed a is a travelling salesman problem. This problem is formulated in Sect. 2.4.1. Furthermore, for any π_0 and a_0

$$f(\pi_0, a_0) \geq \min_a f(\pi_0, a) \quad \text{and} \quad f(\pi_0, a_0) \geq \min_{\pi} f(\pi, a_0). \quad (2.2)$$

2.4.1 TSP Problem

Let C_1, C_2, \dots, C_n denote the component locations. When there is no ambiguity, the same notation is used to refer the components as well. Fix a feeder allocation a . Then, from the observation made above, minimizing $f(\pi, a)$ is a travelling salesman problem. If we imagine the locations MRP, RP, C_1, C_2, \dots, C_n as the cities, then the machine head is our salesman. The costs of travelling between any two cities are given by

$$c(\text{MRP}, B) = \begin{cases} -M + d(\text{MRP}, \text{RP}), & \text{if } B = \text{RP} \\ M, & \text{otherwise} \end{cases}$$

where M is a large positive number and $d(P_1, P_2)$ is the Euclidian distance between the two points in the two-dimensional plane representing P_1 and P_2 ,

$$c(\text{RP}, C_k) = d(\text{RP}, F_i) + d(F_i, C_k) \text{ if type assigned to } F_i \text{ is that of } C_k,$$

$$c(C_k, C_l) = d(C_k, F_i) + d(F_i, C_l) \text{ if type assigned to } F_i \text{ is that of } C_l,$$

for $k \neq l$, $c(C_k, C_k) = M$ and $c(C_k, \text{MRP}) = d(C_k, \text{MRP})$, $k = 1, 2, \dots, n$.

Problem **TSP** is to find a π^* such that

$$f(\pi^*, a) = \min_{\pi} f(\pi, a).$$

At the start of the operations the head always moves from MRP to RP and after mounting the last component it goes back to MRP. This explains the use of M and $-M$ in the definitions of costs above.

It is easy to observe that the cost matrix given above is not symmetric. *Thus, the TSPs encountered in the current problem are asymmetrical where as the corresponding ones in [3] are symmetric.*

2.4.2 Assignment Problem (AP)

In the problem considered in [3] it was possible to assign the same type of components to more than one feeder. However, with the mounting machine at M/s Sree International Limited, it is not possible to pick the same component from more than one feeder *dynamically*. If strips of components of same type are loaded in more

than one feeder, then the machine will continue to draw the components from one of these feeders until all the components in that feeder are exhausted. Only then the head will approach the next feeder. Since this complicates the problem, the practice has been to use only one feeder for each type of components. In view of the complexities involved, it has been decided not to relax the constraint. In other words, it will be assumed that strips of components of same type will be loaded only in one of the feeders. Note that $\sum_{i=1}^m a_{ij}$ is the number of feeders in which strips of type T_j are loaded. In view of the points just mentioned, we have the following constraint.

$$\sum_{i=1}^m a_{ij} = 1, \quad j = 1, 2, \dots, p. \quad (2.3)$$

In this subsection we shall assume that the mounting sequence π is fixed. Let $t(k)$ denote the type of component C_k (we may assume without loss of generality that types are numbered as $1, 2, \dots, p$). To mount the component C_{π_1} the head moves from MRP to RP, RP to the feeder containing type $t(\pi_1)$ and from that feeder to location of C_{π_1} . Let a denote the assignment. The distance travelled by the head to mount C_{π_1} is given by

$$d(\text{MRP}, \text{RP}) + \sum_{i=1}^m a_{it(\pi_1)} d(\text{RP}, F_i) + \sum_{i=1}^m a_{it(\pi_1)} d(F_i, C_{\pi_1}). \quad (2.4)$$

Next, to mount C_{π_2} , the distance travelled by head from C_{π_1} to C_{π_2} is given by

$$\sum_{i=1}^m a_{it(\pi_2)} d(C_{\pi_1}, F_i) + \sum_{i=1}^m a_{it(\pi_2)} d(F_i, C_{\pi_2}). \quad (2.5)$$

Hence we can write

$$\begin{aligned} f(\pi, a) &= d(\text{MRP}, \text{RP}) + \sum_{i=1}^m a_{it(\pi_1)} [d(\text{RP}, F_i) + d(F_i, C_{\pi_1})] \\ &\quad + \sum_{i=1}^m a_{it(\pi_2)} [d(C_{\pi_1}, F_i) + d(F_i, C_{\pi_2})] \\ &\quad + \dots + \sum_{i=1}^m a_{it(\pi_n)} [d(C_{\pi_{n-1}}, F_i) + d(F_i, C_{\pi_n})] \\ &\quad + d(C_{\pi_n}, \text{MRP}) \\ &= d(\text{MRP}, \text{RP}) + d(C_{\pi_n}, \text{MRP}) + \sum_i \sum_j a_{ij} \bar{c}_{ij}, \text{ say.} \end{aligned} \quad (2.6)$$

Clearly, for fixed π , $f(\pi, a)$ is a linear function of the entries in the matrix a specifying the assignment of component types to feeders. Hence, for fixed π , Problem **AP** is to find $a = (a_{ij})$ to:

$$\begin{aligned} &\min_a f(\pi, a) \\ &\text{subject to} \\ &\sum_{i=1}^m a_{ij} = 1, \quad j = 1, 2, \dots, p, \\ &\sum_{j=1}^k a_{ij} \leq 1, \quad i = 1, 2, \dots, m, \\ &a_{ij} \text{'s are 0 or 1.} \end{aligned}$$

This problem is an assignment problem.

2.5 The Heuristic

The heuristic suggested in [3] is as follows. Start with a mounting sequence π^0 and determine feeder allocation a^1 by solving **AP** with $\pi = \pi^0$. Keeping a^1 fixed, solve **TSP** to get an optimal sequence π^1 . Then, with π^1 fixed, solve **AP** and obtain optimal feeder allocation a^2 . Next, with a^2 fixed, once again solve **TSP**. From (2.2), we have

$$f(\pi^0, a^1) \geq f(\pi^1, a^1) \geq f(\pi^1, a^2) \geq \dots$$

This process of solving **TSP** and **AP** alternatively is continued until there is no further reduction in the distance $f(\pi, a)$. Obviously, this process ends in a finite number of steps leading to a *local optimal solution* to the problem.

2.6 Application

The heuristic method was applied to the specific problem mentioned in Sect. 2.3. This problem involved 88 components comprising of 39 types. Since the mounting was divided into three batches (see Sect. 2.3), problem of each batch has been treated as separate instance. In order to make comparisons, the mounting sequence and feeder allocation actually implemented have been noted. These will be denoted by π^0 and a^0 respectively. Feeder allocation of Batch 2 components has posed some special difficulties. This problem is discussed in the next section.

The effectiveness of the solution obtained by the heuristic depends on the choice of initial mounting sequence. For the purpose of solving the above problems, the sequences π^0 s actually implemented have been chosen as the initial sequences. It should be mentioned here that the mounting sequences used by the company are such that components of same type are mounted successively. For example, one of the company's choices for the seven component problem described in Sect. 2.2 is $\pi^0 = (1, 4, 6, 2, 5, 7, 3)$. Note that 1, 4 and 6 are type A components, and these are mounted successively; 2, 5 and 7 are type B components and these are also mounted successively.

The traveling salesman and assignment problems in the sequel are all solved using Quant Systems (QS) package [2]. To work out the cost matrices of these problems and to prepare the input files to QS, special computer programs were written in PASCAL.

Consider Batch 1 problem. The distance actually covered by head in this case is equal to $f(\pi^0, a^0) = 12912$ mm (π^0 and a^0 were used by the company for producing the PCB with 88 components). Starting with π^0 and solving **AP** we get a^1 . Fixing a^1 and solving **TSP** for π results in π^1 which is different from π^0 . Solving **AP** keeping π^1 fixed, we get a^2 . In this case it turned out that $a^2 = a^1$. Hence the algorithm terminated here with $f(\pi^1, a^1) = 11058$ mm. Therefore, the company wasted at least 16.76 % ($= \frac{12912-11058}{11058} \times 100$) of the time in the case of Batch 1.

To get a rough idea about the impact of changing the initial sequence on the distance covered by head, one more sequence σ^0 was tried. This sequence σ^0 was obtained by solving another travelling salesman problem with component locations C_1, C_2, \dots, C_{52} as cities and Euclidian distances between them as costs. Since this TSP is symmetric (see Problem 2.5), it was solved using Vogel's package [9, 10]. Here too the algorithm terminated with a local optimal solution in the third iteration. The solution sequence was $\sigma^0, b^1, \sigma^1, b^2 = b^1$. In this case $f(\sigma^1, b^1) = 11055$ mm. Results are summarized in the exhibit on page 30.

2.7 Special Constraints

In Batch 2, there are 19 components of 13 types. Of these 13 types, 10 can be assigned to any of the feeders. The remaining three types need bigger size feeders. Each of these bigger components can be loaded by combining two adjacent feeders into one. Clearly, this imposes a restriction on the assignment problem **AP**. We will show that this problem can be tackled by imposing additional linear constraints on the assignment problem. Assume that types 1, 2, and 3 require bigger feeders. Consider type 1. This type will be treated as two new types, say, $1'$ and $1''$. Types 2 and 3 are treated in a similar fashion. Thus, we now have 16 types.

These 16 types are to be assigned to 24 feeders in such a way that i' and i'' are assigned to adjacent feeders, $i = 1, 2, 3$. This special constraint poses nonlinear constraints on the assignment variables. However, these constraints are translated into linear constraints as follows. Note that $\sum_i ia_{ij}$ is the feeder number assigned to type j . Hence the constraint

$$\sum_i ia_{i1''} = 1 + \sum_i ia_{i1'} \quad (2.7)$$

ensures that $1'$ and $1''$ are assigned to adjacent feeders. Thus the assignment problem **AP** in this case is

$$\min_a f(\pi, a) \quad (2.8)$$

subject to

$$\sum_{i=1}^m a_{ij} = 1, \quad j = 1', 1'', 2', 2'', 3', 3'', 4, \dots, 13$$

$$\sum_{j=1}^k a_{ij} \leq 1, \quad i = 1, 2, \dots, 24$$

$$\sum_i ia_{ij''} = 1 + \sum_i ia_{ij'}, \quad j = 1, 2, 3,$$

a_{ij} 's are 0 or 1.

Exhibit**Results of Combinatorial Optimization****Batch 1 Using π^0 as Initial Sequence**

π^0 and a^0 are the present sequence and allocation in use at the company.

The heuristic yields a^1, π^1, a^1 . $f(\pi^0, a^0) = 12912\text{mm}$, $f(\pi^1, a^1) = 11058\text{mm}$.

Possible savings $\leq \frac{12912-11058}{11058} \times 100 = 16.76\%$.

Heuristic Solutions Using σ^0 as Initial Sequence

The heuristic yields b^1, σ^1, a^1 . $f(\sigma^0, b^1) = 11152\text{mm}$, $f(\sigma^1, b^1) = 11055\text{mm}$.

Possible savings $\leq \frac{12912-11055}{11055} \times 100 = 16.79\%$.

Batch 2 Using π^0 as Initial Sequence

π^0 and a^0 are the present sequence and allocation in use at the company.

The heuristic yields a^1, π^0 . $f(\pi^0, a^0) = 5228\text{mm}$, $f(\pi^0, a^1) = 4485\text{mm}$.

Possible savings $\leq \frac{5228-4485}{4485} \times 100 = 13\%$.

Batch 3 Using π^0 as Initial Sequence

π^0 and a^0 are the present sequence and allocation in use at the company.

The heuristic yields $a^1, \pi^1, a^2, \pi^2, a^3, \pi^3, a^4, \pi^4, a^5, \pi^5, a^5$.

$f(\pi^0, a^0) = 4408\text{mm}$, $f(\pi^0, a^1) = 3926\text{mm}$, $f(\pi^1, a^1) = 3737\text{mm}$,

$f(\pi^1, a^2) = 3698\text{mm}$, $f(\pi^2, a^2) = 3692\text{mm}$, $f(\pi^2, a^3) = 3690\text{mm}$,

$f(\pi^3, a^3) = 3689\text{mm}$, $f(\pi^3, a^4) = 3688\text{mm}$, $f(\pi^4, a^4) = 3673\text{mm}$,

$f(\pi^4, a^5) = 3670\text{mm}$, $f(\pi^5, a^5) = 3666\text{mm}$.

Possible savings $\leq \frac{4408-3666}{3666} \times 100 = 20.24\%$.

By introducing dummy types, the above problem can be transformed into an assignment problem with additional linear constraints (as in (2.8)). It becomes a 0 – 1 integer programming problem. However, since the number of 0 – 1 variables is small, we were able to solve the models encountered in our application, to optimality, within very reasonable times, using the integer programming solvers in [2]. For larger size models, special algorithms based on Lagrangian Relaxation could be used to solve these constrained assignment problems efficiently.

2.8 Software Development

This project is incomplete or has no meaning if we do not provide a software tool that will solve the problem for each type of PCB and provide an optimal or a local optimal solution. In fact, the most important contribution of this project is the development of the software tool. If modeling the problem and providing the solution method is one part of this project, developing the software tool is equally (if not more) significant and important part. In the rest of this section we shall briefly outline the details of this software tool which we named PCBSOft.

As described in the methodology, we have a series of APs and TSPs to be solved. Though it is possible to develop our own programs to solve APs and TSPs, it would be better, and in fact essential, to use professional OR packages to solve these problems as this would save enormous amount of time and efforts. Solving APs and TSPs is an important step in solving the problem in question. But, what is more important is designing and development of a total package that will take minimum inputs from the user and produce final solution to the user without any further interaction. PCBSOFT was developed in such a way that all that the user has to do is to store the information shown in Fig. 2.4 in an ascii file, and execute the package; no further interaction is required. Once the solution is obtained, PCBSOFT calls for users attention with a beep and prompts him to pick up the solution. It must be mentioned here that at the time of this software development, most of the software was available on DOS operating system and the programming flexibility was very much restricted.

Before looking at various components of PCBSOFT, a brief understanding of the steps involved in meeting package requirements, listed below, will be useful.

- Reading the input data
- Construction (formulation) of APs and TSPs arising in the course of determining optimal solution. This involves computing of cost/distance matrices from the basic data and for specific mounting sequence/feeder allocation. This step is somewhat complex and needs thorough understanding and meticulous programming.
- Developing interface program with professional OR package used to solve the APs and TSPs. This program should (i) prepare the input data files for the OR package, (ii) prepare instructions to the OR package, (iii) execute the OR package and extract solutions.
- Evaluating the solutions in each iteration, keeping track of the objective values, and checking for termination criterion.
- Providing final solution to the user.

2.8.1 Description of PCBSOFT

PCBSOFT comprises a group of programs some of which are command files (these are executable files of type *.com) and the others are batch files. These programs and their functions are described below. The flow and control of these programs in PCBSOFT is shown in Fig. 2.5. PCBSOFT uses Quant Systems (QS), a professional OR package, to solve APs and TSPs. All the programs of PCBSOFT are originally written in PASCAL and later converted into executable files (*.com files).

Stp.com. PCBSOFT uses the files done.fil, base.pcb, route.fil, nroute.fil, feed.fil, nfeed.fil as dummy files during its execution. These files are created each time the package is used to solve a new instance of the problem. The job of clearing the old files is done by this stp.com program. Each time we find a new (better) feeder allocation, it is stored in nfeed.fil. Similarly, each time a new (better) mounting sequence is found, it is stored in nroute.fil.

As.com. This program prepares the input file for solving the assignment problem on QS software for a given mounting sequence. Reading the data from base.pcb, and the mounting sequence from nroute.fil, as.com computes the cost matrix of AP and prepares the input file (work.in) for the QS package. It also creates the instruction file inp.pcb for QS.

Pcb.bat. This program runs the QS package and prompts QS to pick up the necessary inputs from the inp.pcb file. The file inp.pcb contains the information on (i) whether the problem to be solved is AP or TSP, (ii) the input data (cost/distance matrix, size of the problem, etc.) are in the file work.in and (iii) the output (of QS) should be stored in work.out file. The output of QS is diverted to the file work.out.

Asnsolu.com. This program reads the work.out file, picks up the optimal solution to AP (feeder allocation) and stores it in the file nfeed.fil.

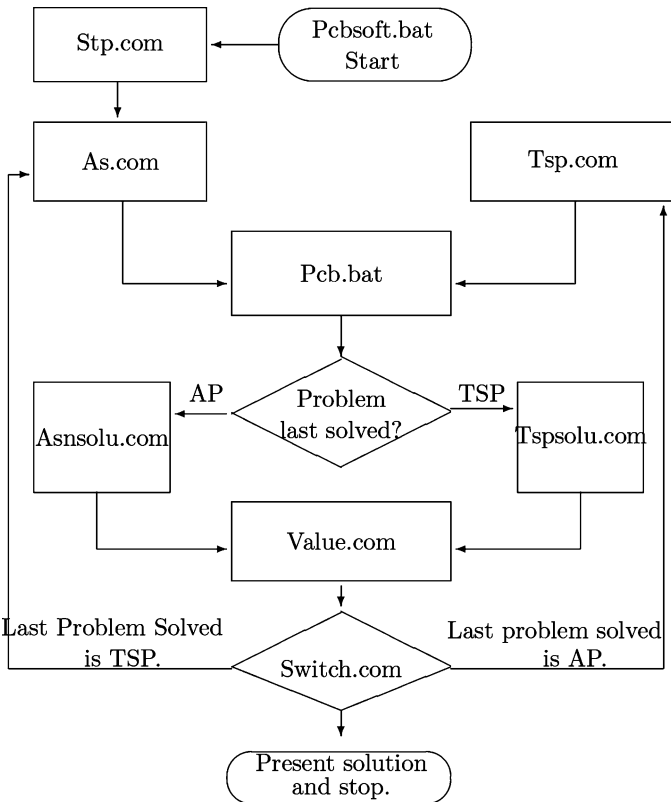


Fig. 2.5 Flow and control of various programs in PCBSOFT

- Value.com.** This program evaluates the total distance covered by the head in mounting components on the PCB for any given feeder allocation and component mounting sequence. The basic data are read from `Base.pcb` file.
- Tsp.com.** This program prepares the input file for solving the travelling salesman problem on QS software for a given feeder allocation. Reading the data from `base.pcb`, and the feeder allocation from `nfeed.fil`, `tsp.com` computes the distance matrix of TSP and prepares the input file (`work.in`) for the QS package. It also creates the instruction file `inp.pcb` for QS.
- Tspsolu.com.** This program reads the `work.out` file, picks up the optimal solution to TSP (mounting sequence) and stores it in the file `nfeed.fil`.
- Switch.com.** This program checks the termination criterion for optimality. If optimal solution is reached, then it terminates the execution of `pcbsoft.bat` by producing the solution to the user. Otherwise, it updates the files for next iteration.

2.9 Summary

In this paper we have presented a case study of combinatorial optimization in PCB manufacturing from Indian Industry. The methodology is based on a heuristic proposed by Chang, Hwang and Murty [3]. The heuristic approach involves solving series of travelling salesman and assignment problems alternatively.

We have also presented a formulation in which the process calls for usage of different sizes of feeders. The usage of different types of feeders poses problems in formulating the assignment problems. Our formulation takes care of these problems by converting them into constrained assignment problems, for which practically efficient algorithms exist.

Under present practice the feeder allocation and sequencing of components is done manually. This takes about a day for the problem that is considered in this project. It took about 1 h to obtain the optimal (local) sequence and feeder allocation on a 486 PC. The losses in productivity in Batch 1, Batch 2 and Batch 3 are at least 16.76 %, 13 % and 20.24 % respectively. Since the company gets repeated orders of the same PCB design, about 2000 boards per month, the savings would be quite substantial by making use of this study and the programs developed here.

The manufacture of PCBs typically requires hundreds of operations. The productivity at most of these operations can be improved substantially by optimizing them using appropriate combinatorial optimization models. We report this study only as an illustration of what can be achieved in PCB manufacturing.

Problems

2.1. Consider the problem introduced in Sect. 2.2 with eight feeders and seven components in three types. A layout for this baby problem is given in Fig. 2.6. Designate the component locations by C_1 (2,7), C_2 (5,7), C_3 (3,5), C_4 (7,5), C_5 (2,4), C_6 (4,4), C_7 (6,3). Do the following exercises.

- (a) Find the coordinates of all points relevant for the problem.
- (b) Fix the sequence $\pi_0 = (3, 1, 5, 4, 6, 2, 7)$ and consider the corresponding assignment problem. Determine the cost matrix for this problem.

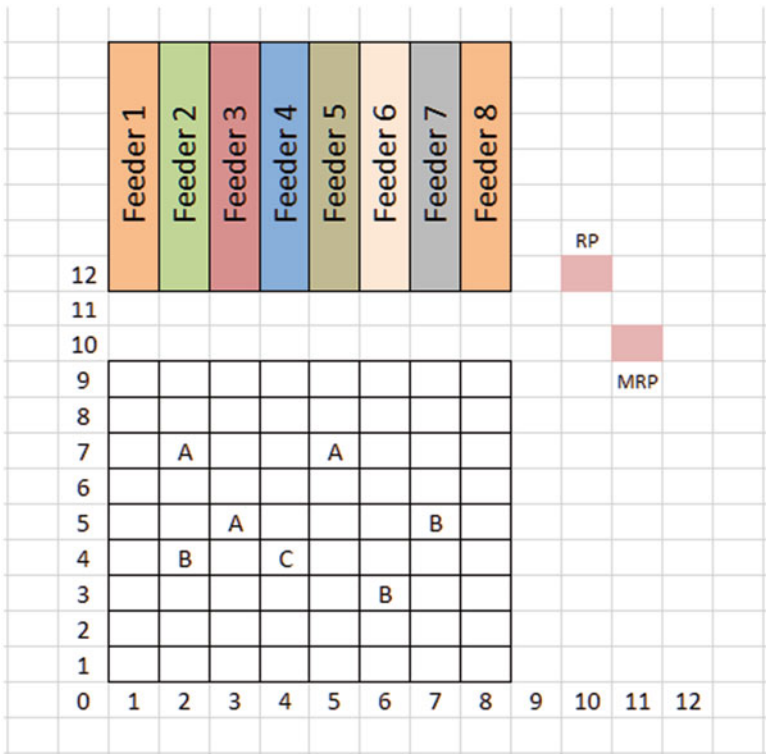


Fig. 2.6 Baby example for practice

- 2.2. This will actually be a project. Note that 39 types cannot be accommodated in 24 feeders, hence the PCB has to be split into batches. How many batches and how to group the components into batches?
- 2.3. Consider the feeder allocation problem with 5 types of components and the number of components to be mounted is 9. Also assume that the number of feeders is 5. Fix the mounting sequence as 834756219. Let the assignment be $a = (a_{ij})$.
- (a) Find the expression for the feeder number to which type 5 components are assigned.
- (b) Write down the constraints which ensure that first and second types are assigned to the adjacent feeders.
- (c) Write down the constraint which ensures that types 1 and 2 are not adjacent.
- (d) Compute the distance traversed by the head for the assignment a .

2.4. Derive the expressions (2.4)–(2.6).

2.5. The local optimal solution to the PCB problem depends on the initial solution. To find an initial solution, one approach is to ignore the feeder locations and treat the problem as an ordinary travelling salesman problem with distance defined as the Euclidean distance between them and solve for the initial sequence. Once this sequence is obtained, then follow the algorithm described in this chapter. Create a sample problem of your choice and study the impact of this approach.

References

1. Bovet, D., Pierluigi, C.: Introduction to the Theory of Complexity. Prentice-Hall, New York (1994)
2. Chang, Y.-L., Sullivan, R.S.: Quant Systems. Prentice-Hall, Englewood Cliffs (1988–1989)
3. Chang, S.Y., Hwang, H.-C., Murty, K.G.: Optimizing the radial component insertion on PCB's. In: Yu, G. (ed.) Industrial Applications of Combinatorial Optimization. Kluwer, Boston (1997)
4. Foulds, L.R.: Graph Theory Applications. Springer, New York Inc. (1992)
5. Gutin, G., Punnen, A.P.: The Traveling Salesman Problem and Its Variations. Kluwer Academic, Dordrecht/Boston (2002)
6. Murty, K.G.: Operations Research: Deterministic Optimization Models. Prentice Hall, Englewood Cliffs (1995)
7. Papadimitriou, C.H., Steiglitz, K.: Combinatorial Optimization: Algorithms and Complexity. Dover, Mineola (1998)
8. Taha, H.A.: An Introduction to Operations Research. Prentice Hall, Upper Saddle River (2007)
9. Volgenant, A., Jonker, R.: A branch and bound algorithm for the symmetric traveling salesman problem based on the 1-tree relaxation. *Eur. J. Oper. Res.* **9**, 83–89 (1982)
10. Volgenant, A., Jonker, R.: The symmetric traveling salesman problem and edge exchanges in minimal 1-trees. *Eur. J. Oper. Res.* **12** 394–403 (1983)

Chapter 3

A Statistical Procedure for Minimizing In Vivo Tests for Quality Control of Vaccines

Abstract Biotechnology and pharmaceutical industries are required to use in vivo tests, in addition to in vitro tests, for drug evaluation. These in vivo tests are carried out on nonhuman animals, and drug regulating authorities insist on rigorous animal testing before issuing license for human use. For various reasons (Wales, Animal testing. http://en.wikipedia.org/wiki/Animal_testing), there is a global move towards minimizing the in vivo tests. This article presents a case study carried out at M/s Bharat Biotech International Limited, Hyderabad, on minimizing in vivo tests for evaluating efficacy of recombinant Hepatitis B vaccine produced by the company. Exploring the past data using different statistical methods, the study recommends a statistical procedure to assess the efficacy of the vaccine produced in each batch based on in vitro test result and provides a decision rule on when to conduct in vivo test. The benefits of the study are discussed. The approach suggested in this work may be extended to similar cases where in vivo tests are involved. It is hoped that this study will be contributing to the cause of minimizing in vivo tests.

3.1 Introduction

The significance of statistical applications in pharmaceutical industry has been ever growing. Statistical analyses have helped the industry in solving many of their critical problems such as establishing drug efficacy through clinical trials, process control and process optimization problems while manufacturing drugs, developing new formulations and testing them, etc. [2, 4, 5, 10, 11]. Health governing bodies such as US FDA, World Health Organization and Drug Authority of India insist on the use of proper statistical methods and data analyses in drug related studies [14]. This chapter presents a case study carried out at M/s Bharath Biotech International Limited, Hyderabad, India, on one of their products, REVAC B+[®], a recombinant Hepatitis B vaccine. Globally, Hepatitis B virus infections are a major cause of cirrhosis

and liver cancer and result in an estimated 780,000 deaths annually (World Health Organization Fact Sheet No 204, Updated July 2014, <http://www.who.int/mediacentre/factsheets/fs204/en/>); Hepatitis B vaccine is an effective way of minimizing such deaths. In fact, hepatitis vaccine is considered by many as the first vaccine against cancer. Production of each vaccine batch takes not less than 2–3 months and the cost of vaccine produced is significant. After the vaccine is produced, it is mandatory for the company to test the vaccine and get approval from the National Regulatory Authority's Testing labs, Central Drugs Laboratory (CDL). Only after formal approval from CDL, the vaccine can be released into the market for sale. The test procedure involves subjecting a sample of the vaccine to two tests: (i) in vivo and (ii) in vitro. The in vivo tests are conducted on laboratory animals. In this case of testing Hepatitis B vaccine, the in vivo test is carried out on mice (Fig. 3.1). The in vivo test takes about a month. It is laborious, sensitive and expensive.



Fig. 3.1 Animal testing: Around 50–100 million vertebrate animals are used in experiments annually. Supporters of the use of animals in experiments argue that virtually every medical achievement in the twentieth century relied on the use of animals in some way and that even sophisticated computers are unable to model interactions between molecules, cells, tissues, organs, organisms, and the environment, making animal research necessary in many areas. On the other hand, animal rights and animal welfare organizations question the legitimacy of animal testing, and argue that it is cruel, poor scientific practice, poorly regulated, that medical progress is being held back by misleading animal models, that some of the tests are outdated, that it cannot reliably predict effects in humans, that the costs outweigh the benefits, or that animals have the intrinsic right not to be used or harmed

The in vitro test is conducted in the laboratory. It is performed with cells or biological molecules studied outside their normal biological context; for example proteins are examined in solution, or cells in artificial culture medium. As opposed to in vivo test, the in vitro test takes only one day. There has been

a continual effort by the developing countries to minimize testing in animals ([1, 6], <http://www.pcrm.org/research/animaltestalt/animaltesting/strategies-to-reduce-animal-testing-in-us-epas-hpv>, http://en.wikipedia.org/wiki/Animal_testing).

Once the tests are completed, the company has to forward samples from every batch of vaccine along with the test results to CDL. Using this sample, the two tests are repeated at CDL and based on the outcome of the results and their comparison with the results of the company, CDL communicates its decision on whether the batch of vaccine is fit to be released into the market. The whole process of testing and getting the approval from CDL takes almost 2–3 months.

The shelf life of the vaccine is for 3 years, and the company loses about 3 months of shelf life in getting the approval. Most of the time is lost in carrying out the in vivo test. In view of this, the ethical and economic considerations and the global move towards minimizing animal testing, it was felt appropriate to look for correlations between in vitro and in vivo tests. Examining the historical data, the company felt that the in vitro test is good enough for judging the efficacy of the vaccine. In view of these aspects, the company appealed to CDL for waiver of the in vivo test. In response to this, the company was asked to make proper justification for the waiver. Consequently, a formal study was carried out using statistical analysis of the past data to check the adequacy of in vitro test to judge the efficacy of the vaccine and to explore the possibility of developing an approval procedure to minimize the number of in-vivo tests. The study has resulted in dispensing with batch to batch in vivo testing and developing an approval procedure with periodic testing.

This chapter presents the statistical analysis part of establishing the justification for partial waiver of the in vivo test and the decision rule developed for deciding whether a batch of vaccine should go for in vivo test based on its in vitro test result. The organization of the chapter is as follows. The problem is described in Sect. 3.2. Section 3.3 discusses the statistical analysis part and presents the process of arriving at the results. Finally, Sect. 3.4 summarizes the results and conclusions.

3.2 Description of the Problem

Vaccine is produced in batches. A sample from the vaccine produced is subjected to two tests: (i) in vivo and (ii) in vitro. In the in vitro test, the response is measured through a chemical test conducted in the quality control laboratory and it represents the concentration of the antigen in the vaccine (World Health Organization TRS 786, Technical Series No. 786, 1989, pages 38–71). The unit of this response is in micrograms per milliliter ($\mu\text{g/mL}$) and the specification requires that this response must be at least $20\mu\text{g/ml}$. This response shall be denoted by X . WHO International Standard for Hepatitis B surface antigen 1985/65 is available for the calibration of their Internal Working Standard by the manufacturer.

In the case of in vivo test, the response is measured through a live pre-clinical trials by testing the vaccine in mice. The response variable here is called the relative potency which is computed on the basis of number of mice responding to the

test vaccine (the sample taken from the production batch) and the number of mice responding to standard vaccine. For the case of Hepatitis B vaccine, there is no specified reference standard (by WHO or any other organization). Companies producing recombinant Hepatitis B vaccine develop and maintain their own standards. The response variable mentioned here, the in vivo relative potency, shall be denoted by Y .

The vaccine produced in each batch is considered effective if the observed values of X and Y measured from a sample vaccine from the batch are at least 20 and 1 respectively.

As mentioned earlier, the practical problem in conducting the in vivo test is that it is very time consuming and results in loss of shelf life. Therefore, the problem at hand is to examine if the in vivo test result can be predicted on the basis of the in vitro test result. If this prediction can be made with high confidence, then the prediction procedure can be converted into a decision procedure.

3.3 Statistical Analysis

Consider the two variables X and Y defined earlier. For ease of understanding and analysis, it is convenient to treat both the variables on the same footing. This can be done by considering the new variable U defined by $U = X/20$ instead of the original variable X itself. Since this is only a scale transformation, the analyses, such as regression and distribution analyses, remain invariant and hence will not affect the inferences. Furthermore, with the new variable U , the pair of variables, U and Y , become directly comparable and have the same specification – that is, both should be at least one as the specification for X is 20.

Past data are available on 102 batches of vaccine (Table 3.1 in Appendix). For each batch, the in vitro measurement X and the in vivo relative potency Y are available. The problem at hand is essentially to infer about Y without measuring it but using the corresponding U (derived from X) value. Henceforth, we shall consider the derived variable U which shall be referred to as the **in vitro relative potency**. The problem at hand is to judge whether Y will be greater than 1 based on the observed value of U .

The analysis is carried out using two different approaches: (i) The regression approach, and (ii) Stochastic Comparison Approach.

3.3.1 Regression Approach

Since U and Y are the measurements corresponding to the same sample, the usual approach is to study the correlation between the two variables and if there is good correlation, regress Y on U to predict Y and make appropriate decision based on the observed value of U . One way of checking the efficacy of this approach is to

use cross validation by setting aside a part of the data for validation and use the remaining data for building the regression equation. Accordingly, data pertaining to 25 batches, picked randomly from the 102 batches, have been set aside for validation. We shall refer to the data of the 77 batches used for model building as model data and the data pertaining to the remaining 25 batches as validation data. In the

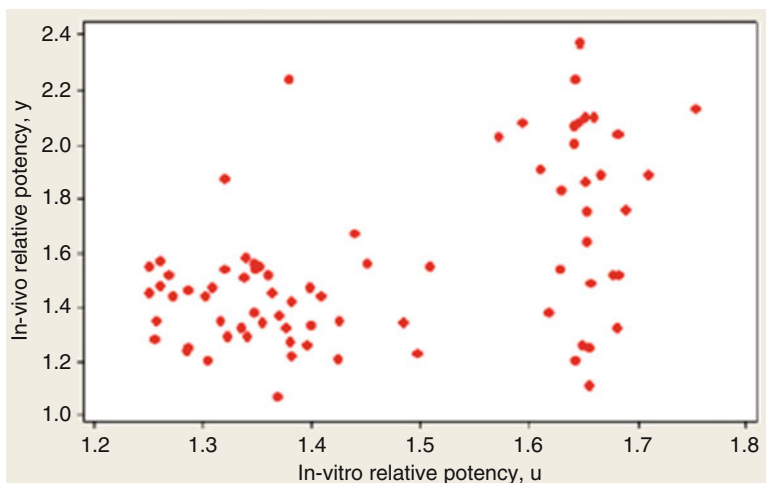


Fig. 3.2 Scatter plot of in vivo relative potency vs in vitro relative potency

regression approach, the in vivo relative potency, Y , is treated as response variable and the in vitro relative potency, U , is treated as the independent or predictor variable. The scatter diagram of U and Y is shown in Fig. 3.2. Though the scatter plot exhibits a bleak picture of the relationship, the square of the Pearson's correlation coefficient between U and Y is 0.2391 which is statistically significant (the p -value is less than 0.001). Normally, such low correlations are considered inadequate for prediction purposes. Further, from the scatter plot, a linear prediction model may not be a satisfactory one. However, since the correlation is significant, we shall proceed with the analysis and examine the efficacy of the prediction model by validating the same against the validation data.

The regression analysis [15] is carried out using the model data, that is, the data pertaining to the 77 batches of the 102 batches. The initial linear model with intercept showed that the intercept may be taken as zero. Consequently the model was rebuilt by dropping intercept. The summary of the results is presented in Fig. 3.3

The decision making rule in this approach is to use the prediction of Y for a new batch based on the observed U value of that batch. Consider batch 7 data (Appendix presents the data on 102 batches): $y = 1.69$, $x = 32.76$ and $u = 1.638$. The predicted value of Y at $U = 1.638$ is given by $\hat{y} = 1.7448$ which is more than the true value (1.69). In order to be confident that the true value will be more than the predicted value, we use the lower prediction limit for cut-off point. For this example, the

The regression equation is

$$y = 0.090 + 1.00 u$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	0.0902	0.2856	0.32	0.753	
u	1.0045	0.1934	5.20	0.000	1.000

S = 0.268998 R-Sq = 26.5% R-Sq(adj) = 25.5%

Regression Analysis (Without Intercept): y versus u

The regression equation is: $y = 1.07 u$

Predictor	Coef	SE Coef	T	P	VIF
Noconstant					
u	1.06524	0.02063	51.62	0.000	1.000
S = 0.267400					

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	190.55	190.55	2664.96	0.000
Residual Error	76	5.43	0.07		
Total	77	195.99			

Fig. 3.3 Regression analysis of y on u

95 % lower prediction limit (one-sided) is equal to 1.208. Therefore, we are 95 % confident that the true y will be more than 1.208 when the corresponding u is equal to 1.638. Since the predicted lower limit is more than 1, the decision will be not to go for in vivo test.

3.3.2 Stochastic Comparison Approach

In this approach we treat the problem as that of matched pairs [3, 15]. For each sample (arising from batches), we have the pair of observations (u, y) where u is the in vitro relative potency and y is the in vivo relative potency. Past data on (u, y) suggest that y tends to be larger than u . Therefore, one way of justifying the waiver of the in vivo test is to examine whether there is substantial statistical evidence to infer that Y is stochastically larger than U .

For two random variables S and T with same range, S is said to be stochastically larger than T if for every t in the range of S , $G(t) \leq F(t)$ with strict inequality holding for at least one t . Here, G and F are the cumulative distribution functions of S and T respectively. In other words, if t is any number, then probability of S being greater than t is greater than the probability of T being greater than t .

For the problem at hand, we can examine if Y is stochastically larger than U . Let F and G denote the cumulative probability distribution functions of U and Y respectively. We wish to test the hypothesis that Y is stochastically larger than U . In other words, we wish to carry out the following test of hypothesis:

$$H_0 : F(t) = G(t) \forall t \text{ Vs } H_1 : G(t) \leq F(t) \forall t \text{ and } G(t) < F(t) \text{ for at least one } t.$$

In the matched pairs problem [3, 15], the above hypothesis is tested by analyzing the pair wise differences. The difference for the i th sample is given by $v_i = y_i - u_i$. Define the random variable $V = Y - U$. Let K denote the cumulative distribution function of V . If the distribution of V is normal, then the above hypothesis can be tested by using parametric method and the t -distribution. When the distribution of V is unknown, one can use nonparametric method to test the hypothesis by applying Wilcoxon Signed Rank test. The assumption required for this test is that V has continuous distribution which is symmetrical about the median. Both parametric and nonparametric methods are examined for testing the above hypothesis.

3.3.3 Parametric Method

Firstly, we shall examine the distribution aspects of the random variables in question. Using Minitab, an attempt was made to identify the distributions of Y and U separately. The outcome is that none of the standard distributions listed in the software (normal, lognormal, gamma, weibul, exponential, extreme value distributions and so on) was suitable for these random variables. However, when the distribution of V is examined, it is found that there is no substantial evidence to reject normal distribution for the difference variable V which is of main interest to us. Before analyzing the distribution of V , we first scrutinize the data on V . It is found that there are four outliers in the data on V (examine the Boxplot in Fig. 3.4). The normal probability plots of before and after dropping outliers are shown in Fig. 3.5 (note the high p -values of Anderson-Darling tests). The test for normality using Kolmogorov-Smirnov test [8] has a p -value of more than 0.15, and that using Ryan-Joiner test [12] has a p -value of more than 0.1. Assuming normal distribution for V , the test for stochastic largeness of Y over U is equivalent to conducting one sample t -test to test that mean of V is equal to zero. The test is carried out on the data with outliers deleted. The summary of the test results is presented in Fig. 3.6.

A rule for deciding whether in vivo test for a new batch should be conducted or not can be derived using the distribution of t -statistic as follows. Let u and y be a pair of observations of in vitro and in vivo relative potencies for a new batch of vaccine, and let $v = y - u$. Define $t = v - \bar{v}$, where \bar{v} is the average of the observed differences and s is sample standard deviation of the observed differences. From the

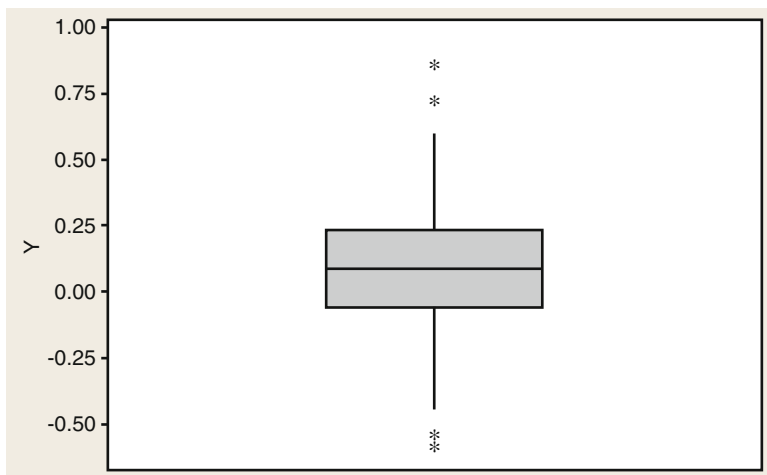


Fig. 3.4 Box plot for detecting outliers in data on V

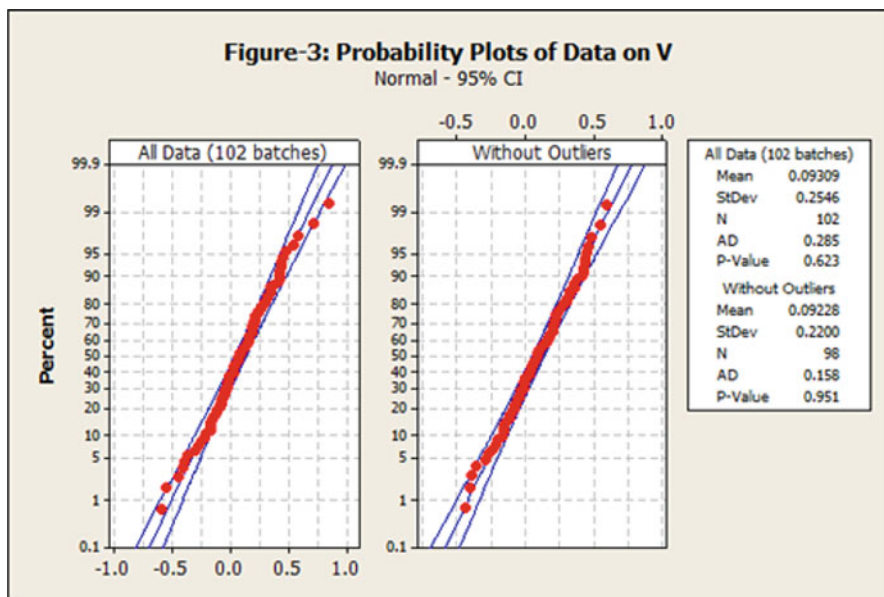


Fig. 3.5 Normal probability plots for before and after deleting outliers

normality assumption of V , it follows that $T = \sqrt{\frac{n}{n+1}} \frac{t}{s}$ has student's t -distribution with $(n-1)$ degrees of freedom, where n is the sample size. Therefore,

$$\text{Prob}\left(t = y - u - \bar{v} \geq t_{\alpha} \sqrt{\frac{n}{n+1}} s\right) = 1 - \alpha,$$

One-Sample T: Without Outliers
 Test of $\mu = 0$ vs > 0

95% Lower

Variable	N	Mean	StDev	SE Mean	Bound	T	P
V	98	0.0923	0.2200	0.0222	0.0554	4.15	0.000

Fig. 3.6 One-sample t -test on V for testing stochastic largeness of Y over V

which in turn is equivalent to saying that

$$\text{Prob}\left(y \geq u + \bar{v} + t_\alpha \sqrt{\frac{n}{n+1}}s\right) = 1 - \alpha. \quad (3.1)$$

We use (3.1) for the decision rule. For a new batch, the decision would be to find the in vitro relative potency u and if $u + \bar{v} + t_\alpha \sqrt{\frac{n}{n+1}}s$ is less than 1, then conduct the in vivo test, otherwise skip it.

Like in the regression approach, we use the model data for determining the decision rule and the validation data for evaluation. The summary of the basic statistics of V from the model data after deleting the outliers is presented in Fig. 3.7.

N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
74	0.0867	0.0273	0.2350	-0.4430	-0.0714	0.0895	0.2255	0.5970

Fig. 3.7 Summary statistics of model data on V after removing three outliers

Taking $\alpha = 0.05$, we have $t_\alpha \sqrt{\frac{n}{n+1}}s = (0.0867) + (-1.6657)(1.0067)(0.235) = -0.3074$. Hence, if the in vitro value is more than 1.3074, then skip the in vivo test. If this rule is applied to the 25 batches of the validation data, only 8 of them will require in vivo test and the remaining 17 get exemption of the in vivo test.

3.3.4 Nonparametric Method

The parametric method above assumes that the pair-wise difference random variable, V , follows normal distribution. The nonparametric method does not assume any specific distribution for V . The only assumption required in this case is that V has continuous distribution. The Wilcoxon signed rank test [3, 15] can be adopted for testing the hypothesis H_0 versus H_1 mentioned above. The signed rank of the pair-wise difference, v_i , is defined as, $-r_i$ if $v_i < 0$, else it is equal to r_i . Here, r_i is the rank of $|v_i|$ among $|v_1|, |v_2|, \dots, |v_n|$. The Wilcoxon signed rank statistic W is defined as the sum of the signed ranks of the positive v_i s. The possible values of W are $0, 1, \dots, m$, each having a probability of 2^{-n} , where $m = n(n+1)/2$. The hypothesis H_0 is rejected in favor of H_1 for large values of W . Statistical tables are available for

critical values of W for small sample sizes (see [15] for $n \leq 100$). When n is large, the normal approximation can be used. For large sample size, the α level critical value of W is given by

$$w_{1-\alpha} = \frac{n(n+1)}{4} + \left[\frac{n(n+1)(2n+1)}{24} \right]^{\frac{1}{2}} z_{1-\alpha}.$$

For the data of 102 batches, the critical value at $\alpha = 0.05$ is equal to 3119 and the observed value of W is equal to 3725. The p -value corresponding to 3725 using the normal approximation is equal to 0.0002. Thus, there is substantial evidence to infer that the in vivo relative potency is stochastically larger than in vitro relative potency.

3.4 Summary of Results and Conclusions

Each batch of vaccine produced is subjected to two tests, the in vitro measurement and the in vivo measurement. The responses measured are the concentration of antigen at ED50 (denoted by x) and the relative potency y measured through mice. Measurement of y takes long time which results in significant loss of shelf life of the vaccine. A study of the past data is carried out to explore the possibility of predicting y on the basis of x and thereby claim waiver on the in vivo test.

Data on (x, y) for 102 batches were analyzed to examine the scope for seeking the waiver on measuring y . For convenience of comparison and analysis, the variable x is replaced with $u = x/20$, a scale transformation of x , and the pairs of observations on (u, y) were considered instead of (x, y) .

The problem was studied using two different approaches: (i) the regression method and (ii) the matched-pair problem approach. Though the correlation between U and Y is statistically significant, the adequacy of regression model for prediction purposes is not satisfactory. The coefficient of determination for linear predictor of Y based on U is 0.24 which is statistically significant but not adequate for the purpose of good prediction. The scatter plot of (U, Y) not only exhibits poor correlation but also indicates absence of any functional relationship.

In the second approach, the problem is considered as matched pair observation problem. Considering the pair of observations (U, Y) for each batch, it was examined whether the in vivo relative potency Y is stochastically larger than the in vitro relative potency, U . Both parametric and nonparametric methods were used to test statistically the stochastic largeness of Y over U . In the parametric approach, the t -test was used and in the nonparametric approach, the Wilcoxon signed rank test was used. Both tests resulted in the conclusion that Y is stochastically larger than U . However, when the normality assumption holds for the differences (V), the t -test is more powerful compared to the Wilcoxon signed rank test [7, 9].

The results of the study were submitted to the Central Drugs Laboratory, Kasauli, which is the National Control Laboratory seeking waiver of in vivo test. Consequently a partial waiver has been granted in which it is required that one out of

every four batches be subjected to in vivo test. In addition to implementing this partial waiver, the company decided to apply the decision rule using t -distribution approach described in the study for every batch. This has resulted in significant savings in time and cost besides other benefits.

The approach used in this study may be useful in several other applications of evaluating vaccine potencies using both in vitro and in vivo tests. One such application is that of evaluating Anti Rabies vaccine potency. It is hoped that this study will help in reducing in vivo testing.

Acknowledgements The author thanks M/s Bharat Biotech International Limited, Hyderabad, India, for providing the opportunity to work on this project and giving permission to publish this work in this book.

Appendix

Table 3.1 In vivo and in vitro values for the 102 batches

Batch No	In vivo (y)	In vitro (x)	Batch No	In vivo (y)	In vitro (x)	Batch No	In vivo (y)	In vitro (x)	Batch No	In vivo (y)	In vitro (x)
1	1.11	33.12	27	2.03	34.02	53	1.34	27.10	79	1.42	26.52
2	1.83	32.60	28	1.89	33.32	54	1.36	26.00	80	1.31	26.81
3	2.07	32.84	29	1.52	33.55	55	1.35	26.31	81	1.32	26.40
4	2.08	32.90	30	1.26	32.98	56	1.23	29.96	82	1.42	25.86
5	2.37	32.94	31	1.58	33.20	57	1.55	27.04	83	1.51	26.75
6	1.54	32.58	32	1.25	33.12	58	1.44	28.18	84	1.52	25.36
7	1.69	32.76	33	1.49	33.15	59	1.35	28.51	85	1.58	26.78
8	1.86	33.04	34	1.52	33.65	60	1.38	26.95	86	1.28	25.10
9	2.24	32.86	35	1.38	32.38	61	1.27	27.60	87	1.48	25.22
10	2.00	32.84	36	1.07	27.37	62	1.26	27.55	88	1.37	27.39
11	1.64	33.06	37	1.20	26.10	63	1.26	27.92	89	1.46	25.73
12	1.75	33.06	38	2.24	27.59	64	1.47	27.97	90	1.47	26.17
13	2.10	33.04	39	1.21	28.48	65	1.44	26.04	91	1.58	25.15
14	2.08	31.87	40	1.29	26.44	66	1.33	28.00	92	1.36	25.28
15	2.13	35.08	41	1.29	26.81	67	1.32	27.52	93	1.68	25.21
16	1.91	32.21	42	1.32	27.80	68	1.41	27.26	94	1.45	25.00
17	2.04	33.62	43	1.52	27.20	69	1.54	26.96	95	1.54	25.21
18	2.04	33.66	44	1.87	26.40	70	1.22	27.63	96	1.47	25.22
19	2.03	31.45	45	1.28	29.66	71	1.42	27.63	97	1.57	25.22
20	1.77	32.72	46	1.67	28.80	72	1.45	27.28	98	1.55	25.01
21	1.89	34.20	47	1.55	30.19	73	1.33	26.82	99	1.46	25.18
22	1.10	33.70	48	1.56	29.02	74	1.42	26.95	100	1.35	25.14
23	1.20	32.86	49	1.34	29.70	75	1.32	26.70	101	1.44	25.45
24	1.32	33.63	50	1.68	28.19	76	1.77	26.82	102	1.25	25.74
25	1.76	33.78	51	1.56	26.95	77	1.30	26.12			
26	2.10	33.20	52	1.24	25.69	78	1.54	26.39			

References

1. Balcombe, J.: The Use of Animals in Higher Education, Problems, Alternatives & Recommendations. The Humane Society Press, Washington, DC (2000). http://www.hsus.org/web-files/PDF/ARI/ARIS_The_Use_Of_Animals_InHigher_Education.pdf. Cited 29 Oct 2000
2. Berry, D.S.: Statistical Methodology in the Pharmaceutical Sciences. Marcel Dekker, New York (1990)
3. Bickel, P.J., Doksum, K.A.: Mathematical Statistics. Holden-Day, San Francisco (1977)
4. Bolton, S.: Pharmaceutical Statistics: Practical and Clinical Applications, 3rd edn. Marcel Dekker, New York (1997)
5. Buncher, C.R., Tsay, J-Y.: Statistics in the Pharmaceutical Industry, 3rd edn. Chapman & Hall/CRC, Boca Raton (2006)
6. Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA) 2006 Amended
7. Conover, W.J.: Practical Nonparametric Statistics, 2nd edn. Wiley, New York (1980)
8. Lilliefors, H.W.: On the Kolmogorov Smirnov Test for Normality with Mean and Variance Unknown. J. Am. Stat. Assoc. **62**, 399–402 (1967)
9. Mood, A.M.: On the asymptotic efficiency of certain non-parametric two-sample tests. Ann. Math. Stat. **25**, 514–522 (1950)
10. Peace, K.E.: Biopharmaceutical Statistics for Drug Development. Marcel Dekker, New York (1987)
11. Peace, K.E.: Statistical Issues in Drug Research and Development. Marcel Dekker, New York (1990)
12. Ryan, T.A. Jr., Joiner, B.L.: Normal probability plots and tests for normality. Technical report, Statistics Department, The Pennsylvania State University (1976)
13. Wales J.: Animal testing. (2015). http://en.wikipedia.org/wiki/Animal_testing
14. WHO Technical Report Series 889. 48th Report, Amended (1999)
15. Zar, J.H.: Biostatistical Analysis, 3rd edn. Prentice Hall, Upper Saddle River (1996)

Chapter 4

Plant Loading and Dispatch Planning

Abstract This case study deals with a project carried out for a cement company. The company has six production plants and a large stockists base. The company produces different types of cement and sells it under a number of brands. The company wanted a logistics software that helps them in production planning and dispatches on a monthly basis. The main objective of the company was to do this planning in such a way that the contribution – a measure of profit – by way of determining optimal production and dispatch plans. But the problem is that the sale prices of cement are subject to fluctuations and cannot be known at the time of planning. An important discovery of this project is that the optimal solutions can be obtained even without knowing the sale prices. Further, it is shown that substantial reduction in the size of the problem can be achieved by exploiting the special structure of the problem. The problem involves a number and variety of constraints. This case study should help in handling large scale projects and building confidence in doing so.

4.1 Introduction

This project was carried out for a leading cement manufacturing company in India. The size and complexity of this project appears quite formidable for an Operations Research (OR) professional who is a novice or moderately experienced. For such consultants/professionals, this study should help in undertaking meaningful industrial projects confidently. To give an idea of the typical problems faced by the OR professionals while dealing with applied projects, some of the practical difficulties encountered during this project are pointed out in this presentation. The problem and the background process are discussed in the rest of this section.

The company has six production plants situated in two different states – three in each state. Based on the composition and quality, the cement is sold under 85 different products. The sales take place through stockists spread across seven different states every month. The movement of material takes place either directly from

the plants or through their depots. The company has 75 depots, called Stock Point Offices (SPOs). The company's production plants, SPOs, branch offices and their corporate office are all connected through an intranet and almost all of their routine business transactions are computerized.

One of the major decision making problems of the company is: How to plan the productions at various plants? This depends on the customer requirements. Majority of the requirements are received from the stockists through the intranet in the form of orders and a small negligible portion is received by other means. The orders for any given month are received in the previous month, latest by the last week of the previous month. These orders fine-tuned by the company's marketing department are taken as the basis for the monthly production plans. At the time of initiating this project, the decisions were being made by the management based on their experience and expertise. The size and complexity of the problem made the company realise the necessity and importance of having a decision support system. Actually a leading consulting firm was responsible for making the company realize the importance of having a decision support system for the management. Subsequently, this project was assigned to the consulting firm and the author (of this book) helped the two companies as an OR expert on a consulting assignment. The formulation and the solution methodology were developed by the author and the consulting firm developed and implemented the decision support system. This chapter presents the modeling aspects of this project.

4.2 The Decision Problem

This section introduces the background of the project, defines the scope and objectives of it. This will help in understanding the size and complexity of the large scale OR problems encountered in industrial applications.

4.2.1 Background

As mentioned earlier, the cement company has three production plants in one state and three in another state. These plants are capable of producing different types of products. Currently the company produces different types of products – all produced by first producing three types of Portland cement. Portland cement is the most common type of cement in the world used as a basic ingredient of concrete, mortar, stucco, and most non-specialty grout. For a detailed discussion on types of cement and their usages see http://en.wikipedia.org/wiki/Portland_cement (accessed last on October 04, 2014). The three types of cement produced by the company are: Ordinary Portland Cement (OPC), Pozzolan Portland Cement (PPC) and Sulphate Resisting Portland cement (SRC). Of these, SRC is produced only for one customer, that too in a very small quantity. Virtually it is OPC and PPC which account for almost 100 % of the sales.

Cement is sold in different *product combinations*. A product combination (PC) consists of cement type (i.e., OPC, PPC or SRC), brand, grade and packing type. *Brands* are created by the company based on certain marketing strategies. Presently there are nine brands such as SooperKing (SK), BrilliantWhite (BW), etc. Grade refers to average compressive strength of cement at 28 days. It is measured in Mega Pascals (MPa). Cement is classified as 33 grade if the average compressive strength at 28 days is between 33 to 43 MPa; as 43 grade if the strength is between 43 to 53 MPa; as 53 grade if the strength is above 53 MPa. Company produces 43 and 53 grade cements. Packing is done in two ways – paper packing and high-density polyethylene (HDPE) (see http://en.wikipedia.org/wiki/High-density_polyethylene, last accessed on October 04, 2014). Company sells cement under different PCs. For example (OPC, SK, 53, paper) is a PC. Though one can think of $108 (= 3 \times 9 \times 2 \times 2)$ PCs, the number of prevailing combinations at present is 85. Throughout this chapter the terms ‘product’ and ‘product combination’ are used interchangeably.

The company has a stockist base of about 2500 stockists spread across the seven states mentioned earlier, but in any given month about 1000 stockists, on an average, place orders for different product combinations with the company. There are two types of transactions, namely, the *trade* and the *non-trade*. Under trade, the customers (stockists) are allowed to purchase material on credit. Credit is not allowed in non-trade. Trade accounts for about 75 % of the sales volume, the balance being non-trade.

The orders are received from stockists daily. The branch offices of the company receive these orders and compile them. Stockists place orders for different PCs. For example, a stockist in a place called Guntur may ask for 20 tonnes of product combination (OPC, Mishu, 53, paper) and 10 tonnes of (PPC, Cardal, 43, HDPE); and a stockist from another place Bangalore may ask for 12 tonnes of (OPC, Mishu, 53, paper) only. Complete information regarding the stockists such as stockist code, location (state, district and location codes), point of delivery, freight etc., is maintained in the company’s database. To facilitate the monthly production plan at a broad level, the marketing department of the company prepares the list of district-wise requirements of OPC, PPC and SRC. This is so, because to produce any product combination, the first step is to produce the three basic material OPC, PPC and SRC.

The products to stockists are supplied either directly from the plants or from any of the SPOs. For the products to move from plants to SPOs, there are two types of transport modes – road and rail. Barring some exceptional cases, the supply of products from a plant or an SPO to the delivery point of any stockist is always by road. Shipment of material (products) from a plant or an SPO is restricted by respective transportation capacities. These and other related issues will be discussed in the section on Formulation (see Sect. 4.3).

When products are supplied to stockists, it involves: (i) cost of production, (ii) freight, (iii) handling charges, (iv) tax, (v) excise duty, (vi) discounts and (vii) price. Production cost consists of two components, namely, overhead costs and variable costs. If the material (products) is directly supplied from a plant to a stockist, then the freight involved in this transaction is called the *primary freight*. On the other hand, if the material is supplied to a stockist from an SPO, then the

freight for the movement of material from plant to SPO is known as primary freight, and the freight for movement of material from SPO to the stockist location is called the secondary freight. Sales tax is levied on the material supplied to stockists. Tax is calculated as a percentage of price minus effective freight. Effective freight is primary freight if the material is supplied from a plant to a stockist directly, else it is the secondary freight. Excise duty is levied on the manufacturing cost. For the purpose of this project it is taken as a percentage of the variable cost. Depending on the relationships with the stockists, the company offers discounts to their stockists; the discounts depend only on the PCs and the stockists' business history with the company.

Finally, price depends basically on the product combination, the location where the material is being supplied, and the date of supply. This should be understood carefully as it has an important consequence and cause for concern. Cement prices are subject to frequent fluctuations and a product price may be different on different days within a month even though it is supplied to the same stockist at the same location.

4.2.2 Management Decisions

The company faces the problem of deciding on where to produce what products and which plants should serve which orders. Also, the company has to decide on the mode of supplies. These decisions have to be made for every month. The large number of stockists, large number of product combinations and the number of transport options (modes of transport) make the problem extremely complex. Even to understand whether the decisions made are optimal or how far they are from optimality, one needs a formal OR approach to the problem. With this in view, the company decided to develop a decision support system that would help them in not only making optimal decisions but also in analyzing various decisions made from the view point of optimality. This would provide them a good management information system. For the purpose of making decisions optimally, the company decided to take the contribution as the objective function. Contribution is like profit. More precisely, it is defined as

$$\text{Contribution} = \text{Price} - Ft - \text{tax} - \text{VarCost} - Hc - Ed - \text{Discounts}, \quad (4.1)$$

where Ft is freight, VarCost is variable cost, Hc is handling (loading and unloading) charges and Ed is the excise duty. Contribution is not profit in full sense because it does not include the overheads component of the production cost. As the overhead costs are anyway unavoidable and more or less fixed for each plant, the management decided to use contribution for the objective function.

4.2.3 Objectives and Scope

After a series of discussions with the consulting firm and the author, and considering various practical issues, the top management of the company set the following as the objectives and scope of the project.

- (i) Planning Module: Given the monthly requirement of each stockist product combination-wise (specified by the marketing department), determine the optimal plant loading and dispatch plan (OPLDP) for the month. Here, optimality is with respect to maximizing overall contribution.
- (ii) MIS Module: Keep track of the historical and legacy data, analyse management actions and compare them with optimal decisions arrived at using the planning module.

It should be emphasized here that the scope of the project is limited to working out the plant loadings and dispatch plans. The scope does not cover the production scheduling or the dispatch scheduling. The decision support system will only determine the quantities of products to be produced at various plants and the quantities of material to be moved, product combination-wise, from plants/SPOs to stockists/SPOs along with the mode of transportation. The company has agreed to take the responsibility of converting these decisions in to production planning schedules and the dispatch schedules. The planning module is meant for determining the optimal plans. However, in the actual execution it may not be feasible to implement the decisions of this module in toto due to various factors affecting the actual planning and execution. Such factors include disturbances in the scheduled production plans, disturbances in the scheduled transportation, change requests from the stockists, etc. The MIS module is useful in understanding the impact of gap between the plan and actual execution. This chapter will present the details of the planning module. The MIS module was handled completely by the consulting firm in consultations with the cement company.

4.3 Formulation

This section presents a detailed description of the decision making problem and a formulation for solving it. Some of the practical difficulties encountered in understanding and the efforts in resolving them are also presented. This is done to highlight some practical aspects that OR professionals come across while dealing with OR applications. To make it simpler, the presentation is confined only to those products which are made from PPC and OPC. Recall that the products made from these types of cement account for almost all the products produced by the company (see Sect. 4.2.1).

4.3.1 Initial Discussions with Management

When the project was signed between the cement and consulting firm, there was a lack of clarity in understanding the inputs and outputs. As per the initial understanding, it was agreed that the freight charges will be made available between supply locations (plants and SPOs) and the district branch offices of the company. It took sometime to convince the company's management that these inputs on freights are not adequate and that the freight charges between the supply locations (plants and SPOs) and stockists were essential. The management agreed to provide these inputs and update their database systems in their intranet to provide these freights.

A Significant Contribution

Another important issue was about specifying prices. As the prices fluctuate even within a month, the question is: Which prices should we take for the contribution? This is a big question as the actual prices cannot be known at the time of planning. To overcome this predicament, the management agreed to go with some average prices. Actually this is an interesting aspect which has led to an important research finding. In fact, this is one point that the author wishes to emphasize as a significant contribution to this project. The details of this will be discussed later in Sect. 4.4.2.

4.3.2 Planning Horizon

The company wanted a decision support system for monthly planning. That is, based on the orders issued by the marketing department for any given month, the decision support system should roll out an optimal plan for production and dispatches to meet the orders for that month. The marketing department will issue the orders three working days before the first day of the month for which the optimal plan is to be determined (it is taken as the first day of the month and not the first working day of the month as the plants run on all days except for the shutdown days for maintenance). The lead time of three working days has been specified to take care of compiling and finalizing the orders for the subsequent month and for communicating the decisions to all plant managers for implementation. Therefore, the planning horizon is taken as month.

4.3.3 Inputs and Notation

In order to define the decision variables and formulate the problem, we first introduce the following notation.

Let M_1, M_2, \dots, M_a denote the plants and let D_1, D_2, \dots, D_b denote the SPOs. In this project, $a = 6$ and $b = 75$. Let L_1, L_2, \dots, L_n be the stockists who require products during the period of planning. The M_i s will be referred to as sources; L_k s will be referred to as destinations. Sources are the places where products are produced or supplied from and the destinations are the places where the products are received. SPOs will act as both sources and destinations. When material is received from the plants, they act as destinations and when the material is shipped from them to the stockists, they act as sources. For this reason D_j will be referred to as both source and destination. Let P_1, P_2, \dots, P_g be the PCs made from OPC, and let $P_{g+1}, P_{g+2}, \dots, P_{g+h}$ be the PCs made from PPC. Let q'_{kld} denote the quantity of P_l (in tonnes) ordered by L_k for day d of the month, $l = 1, 2, \dots, g + h$; $k = 1, 2, \dots, n$. Then $q_{kl} = \sum_d q'_{kld}$ is the total quantity of P_l ordered by L_k for the entire month. For example, L_k wants 500 tonnes of P_l during the entire month (that is, $q_{kl} = 500$) in the following fashion: wants 200 tonnes on day 7 ($q'_{kl7} = 200$), wants 100 tonnes on day 15 ($q'_{kl15} = 100$), and wants 200 tonnes on day 27 ($q'_{kl27} = 200$). For ease of understanding, Table 4.1 presents the list of indices and objects.

Production Capacities

The specification of production capacities of plants are somewhat special in their nature. The production capacities are the maximum quantities of cement that can be made available for production during the month. The production capacity of plant M_i is specified as S_i and T_i , where total quantity of all OPC products supplied during the entire month from M_i cannot exceed S_i , and that the total quantity of all products supplied during the entire month from M_i cannot exceed T_i . In other words, there is a limitation of the quantity of OPC available for the month, namely, S_i . This aspect will be made more explicit in the constraints specification later.

Cost Elements

The objective of the decision making problem is to maximize the contribution defined in Eq. (4.1). It involves the components price, variable cost, freight, tax, handling charges, excise duty and discount. The formulation of the problem depends on the structure of these components. From this, it is essential to understand the details of these cost elements. The price depends only on the product combination, the stockist and the day of the month on which the product is supplied. Let p_{kld} be the unit price (price per ton) of product P_l supplied to stockist L_k on day d . Unfortunately, these prices cannot be known at the time of determining the optimal plan. For this reason, the management has agreed to go with constant projected or forecasted prices. To simplify the complications, the management decided to specify a common price for all days in the month, such as an average price p_{kl} for all p_{kld} . Accordingly, the decision is to compute the contribution using these specified constant prices.

Table 4.1 Notation used for various objects such as plants, SPOs, stockists, products, quantities, etc.

Objects	Index	List/Definition
Plants	i	M_1, M_2, \dots, M_a
SPOs	j	D_1, D_2, \dots, D_b
Stockists	k	L_1, L_2, \dots, L_n
Products*	l	P_1, P_2, \dots, P_g are products made from OPC, and $P_{g+1}, P_{g+2}, \dots, P_{g+h}$ are products made from PPC
Routes	u $(i, j, 2)$ $(i, j, 3)$ $(i, j, 1)$	$u = 1$: Direct from plant to stockist by road, $u = 2$: plant to stockist via SPO all by road, $u = 3$: plant to SPO by rail and from SPO to stockist by road; Transport from M_i to SPO j by road, and from SPO j to stockist; Transport from M_i to SPO j by rail, and from SPO j to stockist; When route $(j, u) = (j, 1)$, j has no relevance as $u = 1$ stands for direct supply from plant to stockist, and hence ignore j in this case
Cement type	Ct	Ct is equal to either OPC or PPC
Product demands	q'_{kld} q_{kl}	q'_{kld} = quantity of P_l in tonnes ordered by L_k for day d total requirement of P_l by L_k for the month, $q_{kl} = \sum_{d=1}^{30} q'_{kld}$
Product prices	p_{kld} p_{kl}	Selling price of P_l for L_k on day d Assumed price for p_{kld} for all days d
Sales tax	t_{ikl}	Sales tax per ton of P_l shipped to L_k from M_i
Variable cost	v_{il} v_i^o v_i^p	v_{il} = variable cost per ton at M_i for product P_l ; $v_{il} = v_i^o$ for $l = 1, 2, \dots, g$, $v_{il} = v_i^p$ for $l = g + 1, g + 2, \dots, g + h$,
Handling charges	H_j	H_j = handling charges per ton at D_j
Expenditure	f_{ijk}	Sum (in rupees) of freight, tax on effective freight and handling charges per ton per ton shipped from M_i to L_k via route (j, u)
Plant capacities	S_i T_i	Total weight of all OPC products that plant M_i can produce during the month, Total weight of all products that plant M_i can produce during the month,
Transport capacities	W_{i1} W_{i2} W_{j3}	Maximum tons that can be shipped out from plant M_i by rail during the entire month, Maximum tons that can be shipped out from plant M_i by road during the entire month, Maximum tons that can be shipped out from SPO D_j during the entire month,
Brand	Br	SK, BW, Mishu, etc.
Grade	Gr	Gr , the strength of cement, is either 43 or 53
Packing type	Pt	Pt is equal to 'paper' or 'HDPE'

*Product means a product combination (Ct , Br , Gr , Pt)

Variable cost is a function of plant location and the type of cement. The cost and quality of limestone which is the main raw material in manufacturing of cement depend upon the lime quarries and their proximities to the manufacturing plants. The following notation will be used for the variable costs. For OPC, v_i^o is the variable cost per ton of any OPC product at plant M_i ; and for PPC, v_i^p is the variable cost per ton of any PPC product at plant M_i . Let v_{il} be the cost per ton of product P_l . Then, $v_{il} = v_i^o$ for $l = 1, 2, \dots, g$, and $v_{il} = v_i^p$ for $l = g + 1, g + 2, \dots, g + h$.

Excise duty depends only on the manufacturing cost of material supplied and nothing else. Excise duty being imposed by the central government, it does not vary with the state in which a plant lies. As it has been decided by the management that only variable cost will be considered for the manufacturing cost, the excise duty can be included in the variable cost. That is, the unit variable cost at a plant will be taken as the actual variable cost at that plant plus the excise duty as a percentage of the unit variable cost at the plant. Tax is a percentage of price minus the effective freight.

Freight depends on the source and destination, the distance between them or the contractual agreements with the transporters and the mode of transport. Strictly speaking, the freight also depends on the quantity of shipment and normally this will be a piece-wise linear function of the quantity of shipment. Since this optimization is a macro level planning and as the transportation is mostly based on contractual agreements, it has been proposed to the management to consider the freight as a linear function of the quantity, and the management has readily agreed for this. Consequently, the freight charges will be assumed to be proportional to the quantities of shipment. Two modes of transport are considered in this project – by rail or by road. By and large, road transport is available between any source and any destination, rail mode is available only between certain plants and SPOs. **Rail mode is not available between any plant and any stockist.**

Handling charges are incurred when the products are loaded or unloaded. These are inevitable at plants and stockists. Additional handling charges are incurred when the products are routed through SPOs. Company's management decided to consider the handling charges only at the SPOs. Therefore, the handling charges component is included in the contribution only if one of the source or destination is an SPO.

Discounts are offered to stockists who maintain good business history and it is fixed percentage of the sales price and the order quantity.

Transportation Capacities

Two modes of transport are used by the company – rail and road. Road transport is available from all sources to all destinations. But the rail mode is available only between certain plants and SPOs. This means that there are three ways in which material can be shipped from plants to stockists: (i) from Plant M_i to stockist L_k by road, (ii) from Plant M_i to stockist L_k via SPO D_j all by road, and (iii) from Plant M_i to SPO D_j by rail and from D_j to stockist L_k by road. These three options are shown in Fig. 4.1. The total transportation capacities of plants and SPOs are

specified as inputs. The maximum tonnage that can be shipped from plant M_i by rail during the month will be denoted by W_{i1} , $i = 1, 2, \dots, a$, the maximum tonnage that can be shipped from plant M_i by road during the month will be denoted by W_{i2} , $i = 1, 2, \dots, a$, and the maximum tonnage that can be shipped from SPO D_j (by road) during the month will be denoted by W_{j3} , $j = 1, 2, \dots, b$.

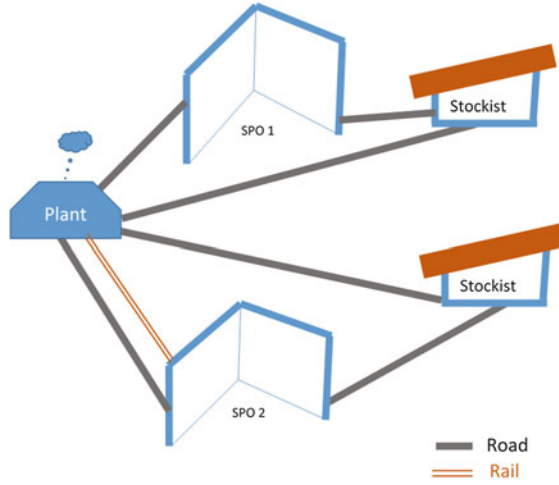


Fig. 4.1 Three modes of transport: (i) directly from plant to stockist by road, (ii) plant to SPO by rail and from SPO to stockist by road, and (iii) plant to SPO by road and from SPO to stockist by road

4.3.4 Decision Variables

The decision making problem of this project can be viewed as a multicommodity network flow problem [1, 4]. The size of the problem is too big to present its full details. For this reason, a miniature version of the problem is created in Example 1 for discussion and understanding.

Example 1: A Miniature Version

In this version, two states with three plants (M_1, M_2, M_3), four SPOs (D_1, D_2, D_3, D_4), 17 stockists (L_1, L_2, \dots, L_{17}) are considered. Rail transport is possible only for the following cases: (i) M_2 to D_1 , and (ii) M_3 to D_2 and M_3 to D_4 . The set up is shown in Fig. 4.2. Available rail transports are shown using *track lines*.

Six product combinations are considered – three of them, P_1, P_2, P_3 , made from OPC and the remaining three, P_4, P_5, P_6 , made from PPC.

The primary inputs are the stockist requirements. These are specified product combination wise and are shown in the form of a two-way table. Table 4.2 presents these requirements. All put together, there are 42 distinct demands for the 6 PCs from the 17 stockists in this miniature version (in the real version of the problem, this number is large). The figures in the body of the table are the quantities required by the stockists as weights (in tonnes) during the entire month.

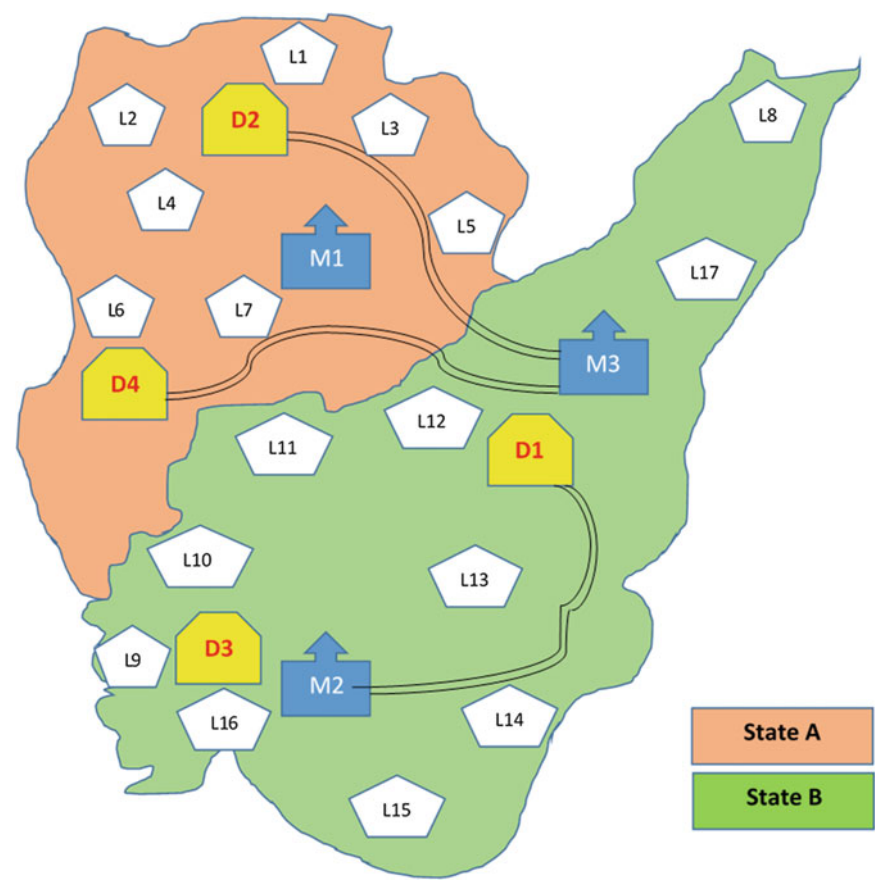


Fig. 4.2 Miniature version of the decision making problem with 2 states, 3 plants (M_1, M_2, M_3), 4 SPOs (D_1, D_2, D_3, D_4), 17 stockists (L_1, L_2, \dots, L_{17}) and 6 PCs

Next set of inputs is on the prices and costs necessary for computing the contributions (see Eq. (4.1)). Contribution has two parts: price and the expenditure. All components other than price in the right hand side of Eq. (4.1) come under expenditure. While expenditure components can be taken to be known at the time of planning, prices are subject to fluctuations, to a large extent, and cannot be known at the time of planning. For example, consider the 500 tonnes of P_3 required by stockist L_{13} (see Table 4.2). The stockist may have placed this order as follows: 200 tonnes in the first

Table 4.2 Stockists' requirements

Stockist	OPC products			PPC products		
	P_1	P_2	P_3	P_4	P_5	P_6
L_1		1000		600		1500
L_2			750	500		
L_3	1200				1400	
L_4	1000					
L_5			900			650
L_6		800			4000	
L_7	1750			1900		650
L_8		3000			750	
L_9			2500			
L_{10}	1000			2000	1000	
L_{11}	550					1300
L_{12}		1800		2200	3000	
L_{13}	950	500	500	600	1000	500
L_{14}		2000			1100	
L_{15}			3000	1300		
L_{16}	2000	1000	1000			
L_{17}			2000		500	900
Total	8450	10100	10650	9100	12750	5500

Note: Figures in the body of the table are the weights (in tonnes) of OPC/PPC required. $P_1 = (\text{OPC}, \text{SK}, 53, \text{HDPE})$, $P_2 = (\text{OPC}, \text{SK}, 53, \text{paper})$, $P_3 = (\text{OPC}, \text{BW}, 53, \text{paper})$, $P_4 = (\text{PPC}, \text{Mishu}, 53, \text{paper})$, $P_5 = (\text{PPC}, \text{BW}, 43, \text{paper})$ and $P_6 = (\text{PPC}, \text{SK}, 53, \text{HDPE})$

week, 100 tonnes in 3rd week and 200 tonnes in the last week of the month. Though the product is same, its prices may be different in these three different weeks. Therefore, to circumvent this issue, the management agreed to provide an expected common price as input for all such cases. Actually, there is an elegant solution to this problem which turned out to be a pleasant surprise to the management. This solution will be discussed later in this section.

Since there are 42 distinct orders (as mentioned above), each order can be served from any of the three plants, and from each plant it can be served in three different ways – rail-road combination through an SPO or road-road combination through an SPO or by road directly to the stockists. Some of these routes may be inadmissible due to non-availability of the transport modes or for other management decisions. For Example 1, the possible routes are shown in the *transportation network* diagram (see Fig. 4.3). Because of large number of routes, it is difficult to visualize the picture through this network diagram. Instead, it is much clearer if the same is presented in the form of a two-way table. This is shown in Fig. 4.4. This format also helps in presenting the costs in a lucid way. With the help of this, we can figure out the costs for different routes. The numbers in the body of the table in Fig. 4.4 are the sum of freight, tax on effective freight and handling charges per ton of material shipped in the respective route. Denote this sum by f_{ijk} . Consider the routes from M_3 to L_1 . In this case, there are three possible routes: (i) M_3 to L_1 directly by road shows the number 135 in Fig. 4.4 which is equal to f_{3*11} (* is used for the SPO as there is no

SPO in this case), (ii) M_3 to D_2 by rail and from D_2 to L_1 ; to find out the unit cost f_{3213} in this route add the figures 80 (M_3 to D_2) and 30 (from D_2 to L_1) which would yield $f_{3213} = 110$, (iii) M_3 to D_2 by road and D_2 to L_1 , to find $f_{3212} = 180$ add 150 and 30. In a similar fashion, other f_{ijkus} can be computed for all permissible routes (see Problem 4.4).

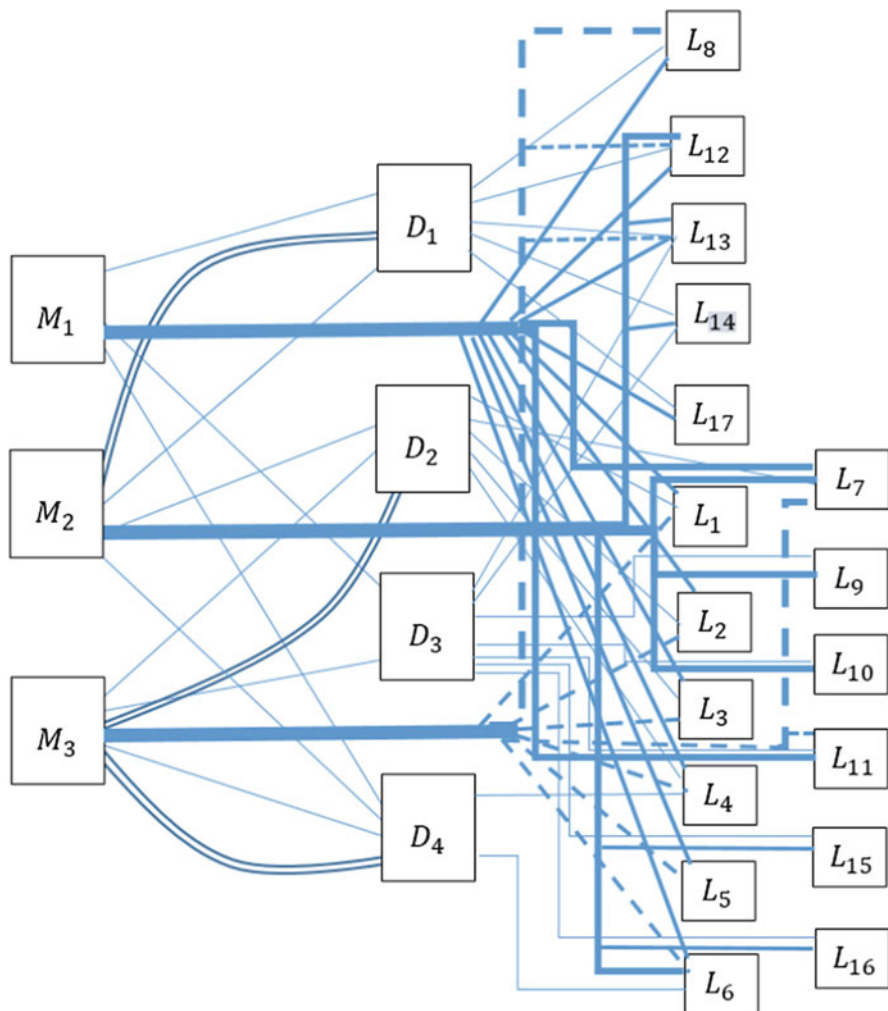


Fig. 4.3 Transportation network for Example 1 with admissible routes

SPO/ Stockist	Road			Rail			Road			
	M_1	M_2	M_3	M_1	M_2	M_3	D_1	D_2	D_3	D_4
D_1	90	110			70					
D_2		200	150			80				
D_3	150		190							
D_4	120	65	170			85				
L_1	90		135					30		
L_2	85		135					30		
L_3	65		115					30		
L_4	65		115					30		25
L_5	60		65					50		
L_6	80	125	125							40
L_7	60	115	85					40		40
L_8	125		115				120		225	
L_9		75							25	
L_{10}		75							25	
L_{11}	105	75	105						30	
L_{12}	65	95	75				25			
L_{13}	105	75	90				30		40	
L_{14}		65					40		35	
L_{15}		65							35	
L_{16}		60							25	
L_{17}	90						50			

Fig. 4.4 Permissible routes of Example 1. Shaded cells are inadmissible routes. For example, transportation is not possible by road from plant M_1 to D_2 or to L_k , $k = 9, 10, 14, 15, 16$ and similarly M_3 to D_1 or D_3 by rail is not possible, etc. Figures in the body of the table are f_{ijkus} per ton of material to be shipped between respective places (See page 63 for details)

Decision Variables

The network described in Fig. 4.3 is transportation network. Clearly, the problem is a multicommodity network flow problem with products as PCs. For each order (PC ordered by a stockist) one should consider all permissible routes. For example, consider 3000 tonnes of P_2 required by L_8 . This can be supplied in five different ways ($M_1 \xrightarrow{\text{Road}} D_1 \rightarrow L_8$, $M_1 \xrightarrow{\text{Rail}} D_1 \rightarrow L_8$, $M_2 \xrightarrow{\text{Road}} D_1 \rightarrow L_8$, $M_1 \xrightarrow{\text{Direct}} L_8$, $M_3 \xrightarrow{\text{Direct}} L_8$). Each of these options will give rise to a decision variable. It is possible to convert this multicommodity network flow problem into a single commodity network flow problem with additional constraints (see Problem 4.3). But for this, the network flow diagram will be very complex. This is particularly so because of

some special constraints of the problem which will be presented later in this section. In light of this, the problem is formulated as a general linear programming problem [2, 5, 6]. Before defining the decision variables, the following notation is necessary.

Definition 4.1. Routing Options. There are three different ways of shipping product from a plant to a stockist. These are: (i) plant to stockist directly, this option will be referred to as ‘Direct’ and the index u used for routes (see Table 4.1) has value 1 for this option, (ii) plant to SPO by road and from there to stockist, this option will be referred to as ‘Road’ and $u = 2$ stands for this option, and (iii) plant to SPO by rail and from there to stockist, this option will be referred to as ‘Rail’, and $u = 3$ stands for this option.

Let $x(i, j, k, l, u)$ denote the quantity of product P_l supplied from plant M_i to stockist L_k using route option u with the interpretation that if u is 2 or 3, then the supply is via SPO D_j . If $u = 1$, then ignore D_j . Note that $x(i, j, k, l, u)$ are defined for $i \in \{1, 2, \dots, a\}$, $j \in \{1, 2, \dots, b\}$, $k \in \{1, 2, \dots, n\}$, $u \in \{1, 2, 3\}$, and $l \in \{1, 2, \dots, g + h\}$. Recall that the product combination, PC, is specified as type of cement, brand, grade and type of packing. Note that P_2 is given by the product combination (*OPC, SK, 53, paper*). In Example 1, L_8 wants 3000 tonnes of P_2 . Observe that there are five different routes of shipping material from plants to L_8 discussed above (see Fig. 4.3). Therefore, there is one-to-one correspondence between the range of l and the product combinations (*Ct, Br, Gr, Pt*).

Permissible Options

If it is possible to ship P_l from plant M_i to stockist L_k using option (i, j, k, l, u) , then the 5-tuple (i, j, k, l, u) is called a permissible option. Note that $x(i, j, k, l, u)$ defined over all permissible options are the decision variables in this project. In order to solve the problem, a list of all permissible options is stored in the database.

Let us count the number of decision variables for Example 1. From Fig. 4.3, there are five ways of shipping product from the three plants. Therefore, for each product required by L_1 , there are five options or five decision variables. Since there are 3 products required by L_1 , there are 15 decision variables associated with products required by L_1 . In other words, there are 15 5-tuples (i, j, k, l, u) associated with orders of L_1 (each of these tuples will have $k = 1$). Likewise, there are ten variables associated with orders of L_2 , ten variables associated order of L_3 , and so on. In all, there are 239 decision variables associated with the orders of 17 stockists in Example 1 (see Problem 4.1). Each of these variables is associated with a unique 5-tuple (i, j, k, l, u) .

With 3 plants, 17 stockists and 6 products, the number of decision variables is 239. The real project has a large number of variables, that too by taking constant expected price for all days in the month (see Problem 4.2).

4.3.5 Objective Function

Each decision variable is associated with a contribution. Let $c(i, j, k, l, u)$ denote the contribution associated with the decision of shipping $x(i, j, k, l, u)$ tonnes of product P_l in the route (i, j, k, u) . These contributions are computed using the formula given by Eq. (4.1) and stored in the computer database. The objective function of the problem is given by

$$\sum c(i, j, k, l, u)x(i, j, k, l, u), \quad (4.2)$$

where the summation is taken over all permissible (i, j, k, l, u) .

4.3.6 Constraints

The decision variables have to satisfy a number of constraints in this project. These are presented below.

Permissible Route Constraints

Permissible options are the routes through which the products can be shipped to the stockists. These options are provided as inputs to the decision support system. The permissible routes and the freight charges associated with them are maintained in the company's data base. The consultant has to write computer codes, as part of the decision support system, to extract these inputs for formulating and solving the problem.

Plant Production Capacity Constraints

The production capacity constraints are derived from the availability of raw material necessary for producing products. It is understood that all products (the PCs) can be produced at all the six plants. The only constraint in this regard is the availability of raw material needed for the type of cement (that is, OPC or PPC) to produce the products. Recall the definitions of the production capacities S_i and T_i of plant M_i (see page 57). The total quantity of all OPC products supplied from M_i cannot exceed S_i , and the total quantity of all products supplied from M_i cannot exceed T_i .

$$\sum_{k=1}^n \sum_{l=1}^g \sum_{u=1}^3 x(i, j, k, l, u) \leq S_i, \quad i = 1, 2, \dots, a. \quad (4.3)$$

Note that the left hand side of the above inequality is the total quantity of OPC drawn from plant M_i . The other production capacity constraint states that the total supply from plant M_i cannot exceed T_i . This constraint is given by

$$\sum_{k=1}^n \sum_{l=1}^{g+h} \sum_{u=1}^3 x(i, j, k, l, u) \leq T_i, \quad i = 1, 2, \dots, a. \quad (4.4)$$

See Table 4.1 for the definition of g and h .

Transport Capacity Constraints

In this project, shipping of products occurs either at a plant or at an SPO. It is possible that a plant may have both rail and road facilities. In Example 1, M_2 and M_3 have this but not M_1 . The transport capacities are specified as the maximum tonnage that can be shipped during the entire month from a plant or an SPO, and these are specified for road and rail separately. Recall that W_{i1} is the maximum quantity that can be shipped out of plant M_i during the month by rail and that the corresponding route option is given by $u = 3$. Therefore, we must have

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^{g+h} x(i, j, k, l, 3) \leq W_{i1}, \quad i = 1, 2, \dots, a. \quad (4.5)$$

Similarly, as W_{i2} is the maximum quantity that can be shipped out of plant M_i during the month by road and that the corresponding route options are given by $u = 1$ and $u = 2$, we must have

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^{g+h} \sum_{u=1}^2 x(i, j, k, l, u) \leq W_{i2}, \quad i = 1, 2, \dots, a. \quad (4.6)$$

Recalling that W_{j3} is the maximum quantity that can be shipped from SPO D_j (by road) during the month, we must have

$$\sum_{i=1}^a \sum_{k=1}^n \sum_{l=1}^{g+h} \sum_{u=2}^3 x(i, j, k, l, u) \leq W_{j3}, \quad j = 1, 2, \dots, b. \quad (4.7)$$

New Plant Constraint

One of the six plants, M_5 , is a new plant (less than 10 years old). Owing to certain benefits that it has enjoyed, it must supply at least 50 % of its products to the stockists of its state. This constraint is termed as new plant constraint. This constraint is given by

$$2 \sum_{k \in LM_5} \sum_{j=1}^b \sum_{l=1}^{g+h} \sum_{u=1}^3 x(5, j, k, l, u) - \sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^{g+h} \sum_{u=1}^3 x(5, j, k, l, u) \geq 0, \quad (4.8)$$

where LM_5 is the set of all k such that L_k and M_5 are in the same state.

Product Demand Constraints

Lastly, the most important set of constraints is meeting the stockists' orders. The quantity of each product ordered by each stockist must be met. Recall that q_{kl} is the quantity of product P_l ordered by stockist L_k during the entire month. According to the decisions, $\sum_{i=1}^a \sum_{j=1}^b \sum_{u=1}^3 x(i, j, k, l, u)$ is the quantity of P_l supplied to L_k . Therefore we must have

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{u=1}^3 x(i, j, k, l, u) = q_{kl}, \quad k = 1, 2, \dots, n; \quad l = 1, 2, \dots, g + h. \quad (4.9)$$

Note that the largest number of constraints emerge from the demand constraints. With 17 stockists and 6 products, we have 42 product demand constraints for Example 1. In the project we have about 2500 stockists and 85 products. Each stockist may not order for all products. The average number of products per stockist is about 12. This means that we have about 84000 product demand constraints on an average.

4.3.7 Complete Formulation

Formulation 1: Putting all constraints together, the decision making problem is

$$\text{Maximize } \sum c(i, j, k, l, u) x(i, j, k, l, u)$$

subject to

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^g \sum_{u=1}^3 x(i, j, k, l, u) \leq S_i, \quad i = 1, 2, \dots, a,$$

(plant OPC production capacity constraints)

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^{g+h} \sum_{u=1}^3 x(i, j, k, l, u) \leq T_i, \quad i = 1, 2, \dots, a,$$

(plant total production capacity constraints)

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^{g+h} x(i, j, k, l, 3) \leq W_{i1}, \quad i = 1, 2, \dots, a,$$

(plant rail transport capacity constraints)

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^{g+h} \sum_{u=1}^2 x(i, j, k, l, u) \leq W_{i2}, \quad i = 1, 2, \dots, a,$$

(plant road transport capacity constraints)

$$\sum_{i=1}^a \sum_{k=1}^n \sum_{l=1}^{g+h} \sum_{u=2}^3 x(i, j, k, l, u) \leq W_{j3}, \quad j = 1, 2, \dots, b,$$

(SPO transport capacity constraints)

$$\sum_{j=1}^b \sum_{l=1}^{g+h} \sum_{u=1}^3 \left(2 \sum_{k \in LM_5} x(5, j, k, l, u) - \sum_{k=1}^n x(5, j, k, l, u) \right) \geq 0,$$

(new plant constraint)

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{u=1}^3 x(i, j, k, l, u) = q_{kl}, \quad k = 1, 2, \dots, n; l = 1, 2, \dots, g + h,$$

(product demand constraints)

 $(i, j, k, l, u) \in F$, where F is set of all permissible options.

4.4 Complexity, Model Simplification and Size Reduction

In this section we shall consider two cases of the decision making problem. In the first case, the contribution is computed based on the average figures recommended by the company's management. In the second case, we find optimal decisions even without knowing the actual contributions. The second case might sound somewhat puzzling but it can be done. A crucial observation of the structure of contribution renders this great advantage. Besides, it dramatically reduces the size of the problem. Again, Example 1 is used to explain these ideas.

Table 4.3 Constraint limits for Example 1

Source	Transport capacities		Production capacities	
	Rail	Road	OPC	Total
M1	15000	320000	25000	40000
M2	13000	35000	25000	38000
M3	9000	29000	30000	40000
D1		22000		
D2		27000		
D3		31000		
D4		28000		

4.4.1 The Case of Assumed Prices

Recall that the cement prices are subject to fluctuations within month and that the management suggested to take constant price for the entire month (that is, same price on all the days for a given outlet), and these prices are provided as inputs to the decision making problem. With this, the contributions $c(i, j, k, l, u)$ is computed for every permissible (i, j, k, l, u) . There are 239 permissible options for Example 1. Again all these can be presented to be viewed at a glance in the form of a two-way table. The contributions are computed for Example 1 and are shown in Table 4.4. This can be viewed as a single commodity network flow model [3, 4]. Here, the sources are identified as plants along with routing options. The last 15 columns of Table 4.4 represent these sources. Each nonempty cell will correspond to a demand node and the number in that cell is the contribution per ton of the respective product. For example, consider the cell containing the figure 2010 in the second column against the row containing L_1P_2 . For this cell, the source node is M_1 supplying product P_2 directly to L_1 (by road) (the notation for this route is $(1, *, 1, 2, 1)$), and the destination node is L_1P_2 . The network flow diagram for this problem will be a simple bipartite graph with an edge from a source to a demand node defined only if the corresponding route is admissible. Each cell in Table 4.4 with positive number (contribution) corresponds to a demand node and an edge (the flow arc) of the bipartite graph. For example, $((1, *, 1, 2, 1), L_1P_2)$ is an edge of the bipartite graph. Further, each of these demand nodes has a demand from a stockist and the demands are presented in Table 4.2. The transport and production constraint limits are presented in Table 4.3. See Sect. 4.7 for solution.

4.4.2 The Case of Unknown Prices

One of the most important contributions to this project is the content of this subsection. The major hurdle in this project is that the exact prices of products cannot be known at the time of planning. Without exact prices, contributions – the coefficients in the linear objective function – cannot be determined. Unlike the items such as freights, taxes, etc. which are fixed well in advance, prices are subject to frequent fluctuations. In this section we will show that optimal solution can be obtained even without knowing the actual prices. In other, words, whatever be the prices during the month, the solution obtained by Formulation 1 will be an optimal solution to the decision making problem provided a feasible solution exists.

We will now see how optimal solutions can be obtained even without knowing the exact prices. This happens because of the special structure of the objective coefficients and that the model is a network flow model. Before proceeding further, it is necessary to revisit the optimization model.

Table 4.4 Contributions for Example 1

Stockist-product	Direct			By road									By rail		
	$M_i \rightarrow L_k$			$M_i \xrightarrow{Road} D_j \rightarrow L_k$									$M_i \xrightarrow{Rail} D_j \rightarrow L_k$		
	M_1	M_2	M_3	M_1D_1	M_1D_3	M_1D_4	M_2D_1	M_2D_2	M_2D_4	M_3D_2	M_3D_3	M_3D_4	M_2D_1	M_3D_2	M_3D_4
L_1P_2	2010		2365					2170		2320				2390	
L_1P_4	2110		2165					1720		2120				2190	
L_1P_6	910		965					520		920				990	
L_2P_3	1215		1565					1370		1520				1590	
L_2P_4	1715		1765					1320		1720				1790	
L_3P_1	2435		2785					2570		2720				2790	
L_3P_3	1935		1985					1520		1920				1990	
L_4P_1	1835		2185			1755		1970	2110	2120		2105		2190	2190
L_5P_3	1840		2235					1950		2100				2170	
L_5P_6	2740		2835					2300		2700				2770	
L_6P_2	1420	1675	1775			1340			1695			1690			1775
L_6P_3	1920	1625	1975			1840			1645			1890			1975
L_7P_1	2440	2685	2815			2340		2560	2695	2710		2690		2780	2775
L_7P_4	2140	1835	2215			2040		1710	1845	2110		2090		2180	2175
L_7P_6	1140	835	1215			1040		710	845	1110		1090		1180	1175
L_8P_2	1175		1585	1090	925		1370				1285		1410		
L_8P_3	1275		1385	1190	1025		920				1085		960		
L_9P_3		1725			1325						1685				
$L_{10}P_1$		1925			1525						1885				
$L_{10}P_4$		2275			2425						2485				
$L_{10}P_3$		2075			2225						2285				
$L_{11}P_1$	1795	2125	2195		1720						2080				
$L_{11}P_6$	1095	875	1195		1020						1080				
$L_{12}P_2$	2235	2505	2625	2185			2465						2505		
$L_{12}P_4$	2535	2255	2625	2485			2215						2255		
$L_{12}P_3$	1735	1455	1825	1685			1415						1455		
$L_{13}P_1$	1795	2125	2210	1780	1710		2060				2070		2100		
$L_{13}P_2$	995	1325	1410	980	910		1260				1270		1300		
$L_{13}P_3$	1195	1525	1610	1180	1110		1460				1470		1500		
$L_{13}P_4$	2495	2275	2610	2480	2410		2210				2470		2250		
$L_{13}P_3$	1895	1675	2010	1880	1810		1610				1870		1650		
$L_{13}P_6$	1495	1275	1610	1480	1410		1210				1470		1250		
$L_{14}P_2$		1735		1370	1315		1650				1675		1690		
$L_{14}P_3$		2085		2270	2215		2000				2275		2040		
$L_{15}P_3$		2735			2315						2675				
$L_{15}P_4$		1485			1615						1675				
$L_{16}P_1$		2740			2325						2685				
$L_{16}P_2$		1740			1325						1685				
$L_{16}P_3$		1340			925						1285				
$L_{17}P_3$	2445			2360			2640						2680		
$L_{17}P_3$	1545			1460			1190						1230		
$L_{17}P_6$	2345			2260			1990						2030		

Extended Model

In our model, the decision variables are defined as functions of the permissible options represented by the 5-tuples (i, j, k, l, u) . Strictly speaking we should have one more variable in these options, that is, the day of the planning month. Let us assume that there are 30 days in the planning month. Recall that $q'_{kl d}$ is the quantity of product P_l ordered by stockist L_k for day d of the month ($q'_{kl d}$ is zero if there is no order for P_l by stockist L_k for day d). Further, q_{kl} defined earlier is the sum of $q'_{kl d}$ s. More precisely, $q_{kl} = \sum_{d=1}^{30} q'_{kl d}$.

Consider the price of one ton of P_l shipped to stockist L_k on day d from plant M_i via route (j, u) . The price of a product depends only the stockist location and the day of the month. In other words, the price depends only on k , l and d and does not depend on i (the source) or on (j, u) (the route). For this reason, the notation p_{kld} is used for the price of one ton of P_l shipped to stockist L_k on day d from plant M_i .

Consider the expenditure components of contribution: freight, variable cost, tax, handling charges, excise duty and discount. Recall that tax (the sales tax) is levied as a fixed percentage on the price minus effective freight. Therefore, the tax can be split into two parts – percentage on price and percentage on effective freight. For the purpose of modeling, we can redefine, without loss of generality, the price as price minus tax on the price part. Similarly, discount is also given as percentage of price for stockists with good history and does not depend on source of supply or the route. For these reasons, we may assume, without loss of generality, that p_{kld} is the price minus tax minus discount. The tax on the effective freight can be absorbed in the freight which depends on i , j , u and k . Variable cost depends only on i and the type of cement, and therefore the variable costs at plant M_i can be taken as v_i^o for all OPC products, and v_i^p for all PPC products (see Table 4.1); handling charges depend only on j (because handling charges at plants and at stockists are not considered in the model); excise duty depends only on i (plant) and type of cement (it is a percentage of variable cost and hence can be absorbed in variable cost).

Structure of Contribution

In view of the above observations, we reemphasize the following assumptions for the rest of this chapter:

- p_{kld} will stand for the sale price minus the percentage of tax on it minus the discount,
- f_{ijkul} will stand for the sum of freight (primary and secondary), handling charges and the tax on the effective freight, and
- v_{il} (v_i^o and v_i^p) will stand for the variable cost plus the excise duty.

All the three mentioned in the above list are unit price/costs, that is, per ton of cement. With these assumptions, the contribution of one ton of P_l shipped to stockist L_k on day d from plant M_i via route (j, u) can be written as

$$c'(i, j, k, l, u, d) = p_{kld} - f_{ijkul} - v_{il}. \quad (4.10)$$

Let us redefine the decision variables. Let $x'(i, j, k, l, u, d)$ be the quantity of product P_l shipped to stockist L_k from plant M_i via route (j, u) on day d . From the product demand constraints, we must have

$$\sum_{(i, j, u)} x'(i, j, k, l, u, d) = q'_{kld}. \quad (4.11)$$

From this we have

$$\begin{aligned} \sum_{(i,j,k,l,u,d)} p_{kld} x'(i,j,k,l,u,d) &= \sum_{(k,l,d)} p_{kld} \sum_{(i,j,u)} x'(i,j,k,l,u,d) \\ &= \sum_{(k,l,d)} p'_{kld} q'_{kld}. \end{aligned} \quad (4.12)$$

Thus, we see that $\sum_{(i,j,k,l,u,d)} p_{kld} x'(i,j,k,l,u,d)$ is a constant and does not depend on the choice of the decision variables $x'(i,j,k,l,u,d)$. Let this constant be K . Then the overall contribution for the month is given by

$$\begin{aligned} \text{Contribution} &= \sum_{(i,j,k,l,u,d)} c'(i,j,k,l,u,d) x'(i,j,k,l,u,d) \\ &= \sum_{(i,j,k,l,u,d)} (p_{kld} - f_{ijk u} - v_{il}) x'(i,j,k,l,u,d) \\ &= K - \sum_{(i,j,k,l,u,d)} (f_{ijk u} + v_{il}) x'(i,j,k,l,u,d) \quad (\text{from (4.12)}) \\ &= K - \sum_{(i,j,k,l,u)} (f_{ijk u} + v_{il}) \sum_d x'(i,j,k,l,u,d) \\ &= K - \sum_{(i,j,k,l,u)} (f_{ijk u} + v_{il}) x(i,j,k,l,u) \end{aligned} \quad (4.13)$$

The last equation above follows from the fact that

$$x(i,j,k,l,u) = \sum_d x'(i,j,k,l,u,d). \quad (4.14)$$

It may be noted that Eq. (4.13) holds only if there is a feasible solution. Therefore, an optimum solution, under the existence of a feasible solution, can be found by minimizing the overall *expenditure*

$$\sum_{(i,j,k,l,u)} e(i,j,k,l,u) x(i,j,k,l,u), \quad (4.15)$$

where the **expenditure coefficients** $e(i,j,k,l,u)$ are given by

$$e(i,j,k,l,u) = f_{ijk u} + v_{il}. \quad (4.16)$$

New Formulation

From the above deductions, it can be seen that the problem stated in Sect. 4.3.7 is equivalent to minimizing the overall expenditure given by (4.15) subject to the same set of constraints (see Problem 4.5).

4.5 Problem Size Reduction

In the above section it is seen that optimal decisions can be determined without knowing the exact sale prices. The problem is equivalent to minimizing the overall expenditure with the objective coefficients given by (4.16) and objective function given by (4.15). Consider a plant M_i and a stockist L_k . Since the variable cost component, v_{il} , of the objective coefficient $e(i, j, k, l, u)$ has the structure $v_{il} = v_i^o$ for $l = 1, 2, \dots, g$, and $v_{il} = v_i^p$ for $l = g + 1, g + 2, \dots, g + h$ (in the revised definitions, v_i^o is equal to variable cost of OPC product plus excise duty on it and v_i^p is equal to variable cost of PPC plus excise duty on it), it is possible to reduce the size of the problem by combining the orders of OPC into one and PPC orders into another for each stockist.

To understand the above, consider M_3 and L_{13} of Example 1. There are 12 variables $x(3, j, 13, l, u)$. These are $x(3, *, 13, l, 1)$, $l = 1, 2, \dots, 6$ (direct supplies from M_3 to L_{13}), and $x(3, 3, 13, l, 2)$, $l = 1, 2, \dots, 6$ (supplies from M_3 to L_{13} via SPO D_3 all by road). Note that $e(3, *, 13, l, 1) = 90 + v_3^o$ for $l = 1, 2, 3$, and $e(3, *, 13, l, 1) = 90 + v_3^p$ for $l = 4, 5, 6$; $e(3, 3, 13, l, 2) = 190 + 40 + v_3^o$ for $l = 1, 2, 3$, and $e(3, 3, 13, l, 2) = 190 + 40 + v_3^p$ for $l = 4, 5, 6$. These variables and their costs are shown in Fig. 4.5.

Define $\hat{e}_1(i, j, k, u) = f_{ijk u} + v_i^o$ and $\hat{e}_2(i, j, k, u) = f_{ijk u} + v_i^p$. Then the objective function (4.15) can be written as

$$\sum_{(i,j,k,u)} \hat{e}_1(i, j, k, u) \sum_{l=1}^g x(i, j, k, l, u) + \sum_{(i,j,k,u)} \hat{e}_2(i, j, k, u) \sum_{l=g+1}^{g+h} x(i, j, k, l, u) \quad (4.17)$$

Define the new variables

$$\hat{x}_1(i, j, k, u) = \sum_{l=1}^g x(i, j, k, l, u) \text{ and } \hat{x}_2(i, j, k, u) = \sum_{l=g+1}^{g+h} x(i, j, k, l, u), \quad (4.18)$$

for $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$, $k = 1, 2, \dots, n$ and $u = 1, 2, 3$. Notice that $\hat{x}_1(i, j, k, u)$ is the total quantity of all OPC products supplied to stockist L_k and $\hat{x}_2(i, j, k, u)$ is the total quantity of all PPC products supplied to stockist L_k . With the definition of these new variables, we can write the objective function (4.17) as

$$\sum_{(i,j,k,u)} \hat{e}_1(i, j, k, u) \hat{x}_1(i, j, k, u) + \sum_{(i,j,k,u)} \hat{e}_2(i, j, k, u) \hat{x}_2(i, j, k, u). \quad (4.19)$$

The above expression implies that the problem can be reduced to a problem with less number of variables, the new variables being the aggregation of the original variables stockist and cement type wise. From the above discussion of supplies of material from M_3 to L_{13} , the 12 variables boil down to 4 variables, and the scenario of Fig. 4.5 reduces to the one shown in Fig. 4.6.

All constraints of Sect. 4.3.7 can be written in terms of the new variables. The number of constraints remain same for all but the set of product demand constraints.

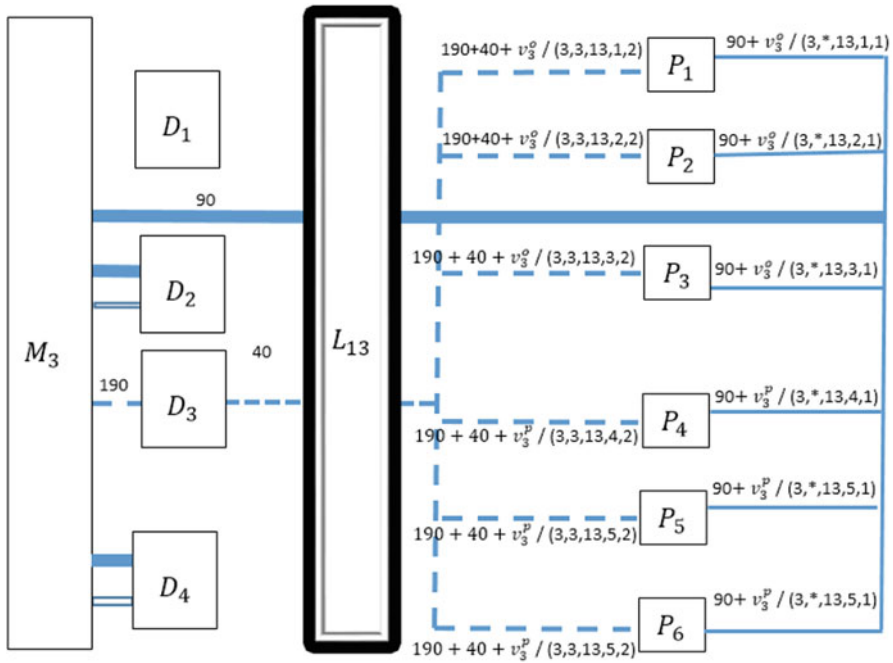


Fig. 4.5 Cost structure of supplies from plant M_3 to stockist L_{13} for Example 1

There will be at most two constraints with respect to each stockist demand in the reduced formulation. To write down the complete formulation let $Q_k^o = \sum_{l=1}^g q_{kl}$ and let $Q_k^p = \sum_{l=g+1}^{g+h} q_{kl}$. Then the product demand constraints can be written as

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{u=1}^3 \hat{x}_1(i, j, k, u) = Q_k^o, \quad k = 1, 2, \dots, n, \quad \text{and} \quad (4.20)$$

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{u=1}^3 \hat{x}_2(i, j, k, u) = Q_k^p, \quad k = 1, 2, \dots, n. \quad (4.21)$$

4.5.1 The Reduced Problem

Formulation 2: The complete formulation for the reduced problem is

$$\text{Minimize} \quad \sum_{(i,j,k,u)} \hat{e}_1(i, j, k, u) \hat{x}_1(i, j, k, u) + \sum_{(i,j,k,u)} \hat{e}_2(i, j, k, u) \hat{x}_2(i, j, k, u)$$

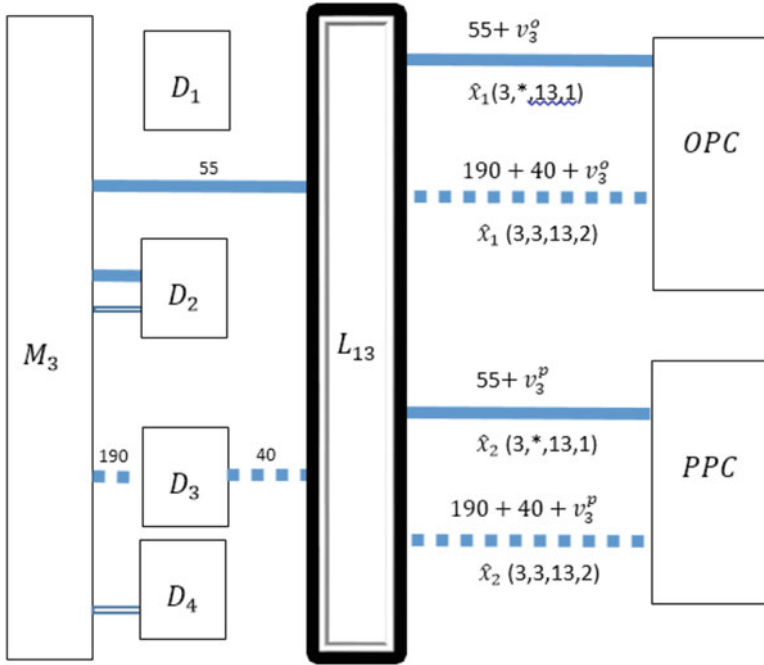


Fig. 4.6 Reduced structure of supplies from M_3 to L_{13} for Example 1

subject to

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{u=1}^3 \hat{x}_1(i, j, k, u) \leq S_i, \quad i = 1, 2, \dots, a,$$

(plant OPC production capacity constraints)

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{u=1}^3 (\hat{x}_1(i, j, k, u) + \hat{x}_2(i, j, k, u)) \leq T_i, \quad i = 1, 2, \dots, a,$$

(plant total production capacity constraints)

$$\sum_{j=1}^b \sum_{k=1}^n (\hat{x}_1(i, j, k, 3) + \hat{x}_2(i, j, k, 3)) \leq W_{i1}, \quad i = 1, 2, \dots, a,$$

(plant rail transport capacity constraints)

$$\sum_{j=1}^b \sum_{k=1}^n \sum_{u=1}^2 (\hat{x}_1(i, j, k, u) + \hat{x}_2(i, j, k, u)) \leq W_{i2}, \quad i = 1, 2, \dots, a,$$

(plant road transport capacity constraints)

$$\sum_{i=1}^a \sum_{k=1}^n \sum_{u=2}^3 (\hat{x}_1(i, j, k, u) + \hat{x}_2(i, j, k, u)) \leq W_{j3}, \quad j = 1, 2, \dots, b,$$

(SPO transport capacity constraints)

$$\sum_{j=1}^b \sum_{u=1}^3 \left(2 \sum_{k \in LM_5} (\hat{x}_1(i, j, k, u) + \hat{x}_2(i, j, k, u)) - \sum_{k=1}^n (\hat{x}_1(i, j, k, u) + \hat{x}_2(i, j, k, u)) \right) \geq 0,$$

(new plant constraint)

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{u=1}^3 \hat{x}_1(i, j, k, u) = Q_k^o, \quad k = 1, 2, \dots, n,$$

(OPC product demand constraints)

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{u=1}^3 \hat{x}_2(i, j, k, u) = Q_k^p, \quad k = 1, 2, \dots, n, \text{ and}$$

(PPC product demand constraints)

$$(i, j, k, u) \in \hat{F} = \{(i, j, k, u) : (i, j, k, l, u) \in F \text{ for some } l \in F\}$$

4.5.2 Effect of Reduction

It has been observed that under the given cost structure, the problem of maximizing the contribution is equivalent to minimizing the overall expenditure. The reduced model uses only the aggregated orders Q_k^o s and Q_k^p s. In this project, on an average 12 different products are ordered by each stockist making the product demand constraints about 84000 (see page 68). In the reduced model, the 12 product wise orders get reduced to 2 orders. This means that there is a 6-fold reduction in the number of constraints reducing the number of constraints to 14000 from 84000 on an average. The number of variables associated with each stockist will also have a 6-fold reduction. Assuming about 10 different options of serving a stockist, the number of decision variables will reduce from 8,40,000 to 1,40,000. Thus the impact of the reduced model has dramatic effect on size reduction.

Let us examine the reduction effect on the miniature Example 1. From Table 4.4, it can be observed that there are 239 variables and 42 product demand constraints. The reduced demands Q_k^o s and Q_k^p s for the example are given in Table 4.5. Thus, the 42 product demand constraints in the original problem reduces to 31 product demand constraints (the number is 31 and not 34 as L_4 , L_9 and L_{16} require only OPC products). And the number of variables reduces from 239 to 173. The reduced expenditure costs $(f_{ijk u} + v_{il})$ s are presented in Table 4.6.

Table 4.5 Product requirements (Q_k^o and Q_k^p s) in the reduced model

Stockist	L1	L2	L3	L4	L5	L6	L7	L8
OPC	1000	750	1200	1000	900	800	1750	3000
PPC	2100	500	1400	0	650	4000	2550	750

Stockist	L9	L10	L11	L12	L13	L14	L15	L16	L17
OPC	2500	1000	550	1800	1950	2000	3000	4000	2000
PPC	0	3000	1300	5200	2100	1100	1300	0	1400

4.6 Plant Loading

A solution to the reduced problem discussed above will provide a solution that will specify the sum of product quantities and the routes of transport for each of the two types of cement – OPC and PPC. It does not spell out the break up of product wise quantities at each plant. However, for the planning, the management needs these decisions. This problem leads another optimization problem which is a simple transportation problem without an objective function. This will be elaborated with the help of Example 1.

Table 4.6 Reduced costs for Example 1

Stockist-product	Direct			By road										By rail		
	$M_i \rightarrow L_k$			$M_i \xrightarrow{Road} D_j \rightarrow L_k$										$M_i \xrightarrow{Rail} D_j \rightarrow L_k$		
	M_1	M_2	M_3	M_1D_1	M_1D_3	M_1D_4	M_2D_1	M_2D_2	M_2D_4	M_3D_2	M_3D_3	M_3D_4	M_2D_1	M_3D_2	M_3D_4	
L_1 -OPC	1390		1035					1230		1080				1010		
L_1 -PPC	1290		1235					1680		1280				1210		
L_2 -OPC	1385		1035					1230		1080				1010		
L_2 -PPC	1285		1235					1680		1280				1210		
L_3 -OPC	1365		1015					1230		1080				1010		
L_3 -PPC	1265		1215					1680		1280				1210		
L_4 -OPC	1365		1015			1445		1230	1090	1080		1095		1010	1010	
L_5 -OPC	1360		965					1250		1100				1030		
L_5 -PPC	1260		1165					1700		1300				1230		
L_6 -OPC	1380	1125	1025			1460			1105			1110			1025	
L_6 -PPC	1280	1575	1225			1360			1555			1310			1225	
L_7 -OPC	1360	1115	985			1460		1240	1105	1090		1110		1020	1025	
L_7 -PPC	1260	1565	1185			1360		1690	1555	1290		1310		1220	1225	
L_8 -OPC	1425		1015	1510	1675		1230				1315		1190			
L_8 -PPC	1325		1215	1410	1575		1680				1515		1640			
L_9 -OPC		1075			1475						1115					
L_{10} -OPC		1075			1475						1115					
L_{10} -PPC		1525			1375						1315					
L_{11} -OPC	1405	1075	1005		1480						1120					
L_{11} -PPC	1305	1525	1205		1380						1320					
L_{12} -OPC	1365	1095	975	1415			1135						1095			
L_{12} -PPC	1265	1545	1175	1315			1585						1545			
L_{13} -OPC	1405	1075	990	1420	1490		1140				1130		1100			
L_{13} -PPC	1305	1525	1190	1320	1390		1590				1330		1550			
L_{14} -OPC		1065		1430	1485		1150				1125		1110			
L_{14} -PPC		1515		1330	1385		1600				1325		1560			
L_{15} -OPC		1065			1485						1125					
L_{15} -PPC		1515			1385						1325					
L_{16} -OPC		1060			1475						1115					
L_{17} -OPC	1355			1440			1160						1120			
L_{17} -PPC	1255			1340			1610						1570			

In Example 1, there are three OPC products P_1 , P_2 , P_3 and the total quantities of these three products required by all stockists put together are equal to 8450, 10100 and 10650 tonnes respectively (see the last row of Table 4.2). The sum of these quantities is equal to 29200. The solution to the reduced problem will distribute this total quantity to the three plants M_1 , M_2 and M_3 . For the purpose of understanding, assume that this distribution is as follows: M_1 supplies 9000 tonnes, M_2 supplies 13000 tonnes and M_3 supplies 7200 tonnes. Now the question is how much of 9000 tonnes supplied by M_1 is for P_1 , how much for P_2 and how much it is for P_3 . Similar questions for M_2 and M_3 need answers. These questions (decisions) can be answered by solving a transportation problem. To see this, let y_{il} denote the quantity of product P_l supplied by plant M_i for all stockists put together, $i = 1, 2, 3; l = 1, 2, 3$. Then, we must have $\sum_{l=1}^3 y_{1l} = 9000$, $\sum_{l=1}^3 y_{2l} = 13000$ and $\sum_{l=1}^3 y_{3l} = 7200$. Also, we must have $\sum_{i=1}^3 y_{i1} = 8450$, $\sum_{i=1}^3 y_{i2} = 10100$ and $\sum_{i=1}^3 y_{i3} = 10650$. This is nothing but a simple transportation model [6]. The transportation table is shown in Fig. 4.7.

		OPC Products			Availabilities
		P1	P2	P3	
Plants	M1	y_{11}	y_{12}	y_{13}	9000
	M2	y_{21}	y_{22}	y_{23}	13000
	M3	y_{31}	y_{32}	y_{33}	7200
Demands		8450	10100	10650	29200

Fig. 4.7 Transportation model for plant loading

Recall that the product is defined as the combination of type of cement (OPC, PPC), brand, grade and packing. In the actual problem, the company management decided to consider only the variable cost of raw material for the type of the cement (this is the reason for taking the manufacturing cost as a function of the cement type only). The implication of this is that within a plant the costs of making different products remains same. Under this situation, the above transportation problem will have no objective function, and any feasible solution can be used for loading the plants with different products.

Notice that there are two transportation problems that need to be solved – one for OPC and the other for PPC. These two problems will have no binding or common constraints as the plant capacity constraints which links the OPC and PPC production capacities are taken care of while solving the reduced problem itself.

Thus, we see that the management's decision making problem of production planning and dispatch planning is solved as a two stage problem. In the first stage, a reduced model of the problem is formulated and solved. Once the solution is obtained, the solution is used to decide the loading of plants with different products. This is the second stage problem and it is simple transportation problem without an objective function. Combining the solutions of the two stages will provide optimal decisions for the month.

4.7 Past Data Analysis and Results

In this section we shall apply the models discussed above to the miniature Example 1 and then present the results of application to one instance of the real problem. Any professional solver can be used to solve the instances of the problem. Here we will show a method of solving the problems using LINGO, a professional solver (www.lindo.com/products/lingo/). LINGO uses a powerful set-based language for model building and solving. It also provides a good support to import input data from excel and export output (results) to excel. The steps for model building are explained with the help of Example 1.

4.7.1 Basic Objects and the LINGO Model for Example 1

The basic objects in our problem are: (i) Plants, (ii) SPOs, (iii) Transport Modes, (iv) Stockists, (v) Cement Type, (vi) Brands, (vii) Grades and (viii) Packaging types. The decision variables and the input data (capacities, costs) are functions of these objects. To solve the problem, we first need to prepare the model file, supply the inputs and then run the LINGO solver to get a solution. This file for problem of Example 1 is shown in Exhibit 1. The file has three sections – a sets section, a data section and a model section. The sets section is used for setting up the framework of the problem in which the basic objects and some new objects are defined. These new objects are called the derived sets in LINGO's terminology. These definitions are like variable declarations in a computer programme such as a c-program. The data section is meant for passing the inputs and exporting the outputs. The objective function and the constraints are presented in the model section. The details of the model and solving the problem are given in the following paragraphs.

The file for LINGO model starts with the command “Model:” and ends with the command “End”. The LINGO language is not case sensitive. Any comment can be included between exclamatory symbol (!) and semicolon (;). In Exhibit 1 each line is numbered in comment. This is done for the purpose of referencing. Thus, the file starts with “Model:” command in Line 1 and ends with “END” command in Line 289. Lines from 42 to 277 are not shown in the exhibit.

The sets section starts in Line 2 and ends in Line 13 with the commands “Sets:” and “endsets”. Line 3 declares the ‘Plants’ object and states that ‘M1’, ‘M2’ and ‘M3’ are its elements. Note that S, T, MRoad and MRail *are declared* as the Plant variables. This means that each of these variables is a three dimensional vector. Recall that we used S_i as the OPC production capacity of plant i in Formulation 1. Since S is declared under “Plants”, we wish to use $S(M1)$, $S(M2)$ and $S(M3)$ as plant OPC production capacities of plants M1, M2 and M3 respectively. Note that in data section (Line 33), we have set “S = 25000 25000 30000;.” This amounts to setting $S(M1) = 25000$, $S(M2) = 25000$ and $S(M3) = 30000$ as inputs to the model. Interpret $T(M1)$, $T(M2)$ and $T(M3)$ in a similar fashion and observe their values in Line 34. The variable MRoad is used for the transportation capacities of the plants.

Exhibit 1: LINGO Model for the Problem

```

Model: ! Line 1;
sets: ! Line 2;
Plants/M1..M3/:S,T,MRoad,MRail; ! Defining plant objects; ! Line 3;
SPOs/D0..D4/:SRoad; !Defining SPO objective; ! Line 4;
TModes/RR TR/;; !Defining route objects; ! Line 5;
Stockists/L1..L17/;; !Defining Stockists; ! Line 6;
CTypes/OPC PPC/;; !Defining cement types; ! Line 7;
Brands/Br1..Br9/;; ! Defining brands.; ! Line 8;
Grades/G43 G53/;; ! Defining grades; ! Line 9;
Packing/Paper HPDE/;; ! Defining packing types; ! Line 10;
Options(Plants,SPOs,TModes,Stockists,CTypes,Brands,Grades,Packing):x, c; ! Line 11;
!Defining variables - x's are decision variables and c's are contributions; ! Line 12;
endsets ! Line 13;
! Maximize total contribution; ! Line 14;
max = @sum(Options:c*x); ! Line 15;
!Plant OPC Capacity constraints; ! Line 16;
@for(Plants(i):@sum(Options(i,j,u,l,m,b,g,k)| m # EQ# 1:x)<= S(i)); ! Line 17;
!Plant total Capacity constraints; ! Line 18;
@for(Plants(i):@sum(Options(i,j,u,l,m,b,g,k):x)<= T(i)); ! Line 19;
!Plant road transport capacity constraints; ! Line 20;
@for(Plants(i):@sum(Options(i,j,u,l,m,b,g,k)| u # EQ # 1:x)<= MRoad(i)); ! Line 21;
!Plant rail transport capacity constraints; ! Line 22;
@for(Plants(i):@sum(Options(i,j,u,l,m,b,g,k)| u # EQ # 2:x)<= MRail(i)); ! Line 23;
!SPO transport capacity constraints, ! Line 24;
D0 being direct supply no constraint for j=1; ! Line 25;
@for(SPOs(j):@sum(Options(i,j,u,l,m,b,g,k)| j# GT# 1:x)<= SRoad(j)); ! Line 26;
! Demand constraints; ! Line 27;
@for(PCs(l,m,b,g,k):@sum(Options(i,j,u,l,m,b,g,k):x) = q(l,m,b,g,k)); ! Line 28;
! New Plant constraint; ! Line 29;
@sum(Options(i,j,u,l,m,b,g,k)| i # EQ# 3 # AND# l # GE# 8:x) ! Line 30;
- @sum(Options(i,j,u,l,m,b,g,k)| i # EQ # 3:x)/2 >= 0; ! Line 31;
DATA: ! Line 32;
S = 25000 25000 30000; ! Line 33;
T = 40000 38000 40000; ! Line 34;
MRoad = 320000 35000 29000; ! Line 35;
MRail = 15000 13000 9000; ! Line 36;
SRoad = 29000 22000 27000 31000 28000; ! Line 37;
! There are 239 rows for options. Only a few are shown here.; ! Line 38;
Options = ! Line 39;
M1 D0 RR L1 OPC Br1 G43 Paper ! Line 40;
M1 D0 RR L1 PPC Br3 G43 HDPE ! Line 41;
... ! Line 42;
... ! Line 43;
... ! Line 44;
M3 D4 TR L7 PPC Br7 G43 HDPE; ! Line 278;
! Importing contributions from the excel file named 'Data.xlsx' and range ! Line 279;
named 'Contribution'; ! Line 280;
c = @OLE('C:\Path\Data.xlsx', 'Contribution'); ! Line 281;
! Importing demands from the excel file named 'Data.xlsx' and range ! Line 282;
named 'Demands'; ! Line 283;
q = @OLE('C:\Path\Data.xlsx', 'Demands'); ! Line 284;
! Exporting the solution to the range named 'Vol' in ! Line 285;
Excel file named 'Data.xlsx'; ! Line 286;
@OLE('C:\Path\Data.xlsx', 'Vol')=x; ! Line 287;
ENDDATA ! Line 288;
END ! Line 289;

```

That is, $MRoad(M1)$ is the maximum weight of the products that can shipped by road from plant M1 during the entire month. Similarly, $MRail(M3)$ is the maximum weight of the products that can shipped by rail from plant M3 during the entire month.

Line 4 defines SPOs with five elements D0 to D4 and declares SRoad as variable of this type. In Example 1, there are only four SPOs, call them D1 to D4. The SPO D0 is a dummy SPO. All direct supplies from plants to Stockists are imagined to be supplies through this dummy (artificial) SPO. Each SPO has transport capacity and since from SPO to the stockists the only transport is by road, SRoad is meant for these transport capacities. That is, SRoad(D2) is the maximum weight of products that can be shipped from SPO D2 in the entire month. Note that constraint in Line 26 states that transportation capacities of SPOs D1 to D4 cannot exceed the respective SRoad(D1) to SRoad(D4) (in this D0 is not included).

TModes defined with elements RR and TR in Line 5 stands for transportation modes. RR stands for ‘by road from plant to SPO and by road from SPO to stockist,’ and TR stands for ‘by rail from plant to SPO and by road from SPO to stockist.’

Line 6 defines the list of stockists as L1 to L17; Line 7 defines the type of cement as CType; Line 8 defines the Brands as Br1 to Br9; Line 9 defines the Grades as G43 and G53; Line 10 defines the Packing type as Paper and HDPE.

Line 11 defines the derived object ‘Options’ with arguments as Plants, SPOs, TMode, Stockists, CTypes, Brands, Grades and Packing. Further, it says that x and c are the variables of *Options type*. What this means is that x and c are 8-dimensional arrays or you can think of them as functions of the objects Plants, SPOs, TModes, Stockists, CTypes, Brands, Grades and Packing. In other words, we can talk about objects like $x(M1, D0, RR, L1, OPC, Br1, G43, Paper)$, $x(M1, D0, RR, L1, PPC, Br3, G43, HDPE)$, $c(M1, D0, RR, L1, OPC, Br1, G43, Paper)$, $c(M1, D0, RR, L1, PPC, Br3, G43, HDPE)$, and so on. Here, x is used to define our decision variables, and by $x(M1, D0, RR, L1, OPC, Br1, G43, Paper)$ we mean that our decision is to ship $x(M1, D0, RR, L1, OPC, Br1, G43, Paper)$ tons of product (OPC, Br1, G43, Paper) to stockist L1 from plant M1 directly (directly because of D0 and RR are the values of the arguments for SPOs and TModes); and $c(M1, D0, RR, L1, OPC, Br1, G43, Paper)$ stands for the contribution for this shipment per ton of the product. Observe that the two tuples (M1 D0 RR L1 OPC Br1 G43 Paper) and (M1 D0 RR L1 PPC Br3 G43 HDPE) used as arguments of x and c above are listed as possible options in Line 40 and Line 41 in Exhibit 1. Recall that there are 239 admissible options like this (see Problem 4.6). Note that (M1, D2, RR, L1, OPC, Br1, G43, Paper) is not an admissible option and hence this will not be listed in the lines Line 40 to Line 278.

In the model section (Line 14 to Line 31), the objective function and the constraints are specified. The objective function $@sum(Options:c*x)$ in Line 15 stands for $\sum c(i,j,u,l,m,b,g,k)x(i,j,u,l,m,b,g,k)$ where summation is taken over all

(i,j,u,l,m,b,g,k) listed in lines Line 40 to Line 278 of Exhibit 1. The statement $\text{max}=@\text{sum}(\text{Options:c*x})$ in Line 15 states that the objective function is to be maximized. Line 17 specifies the OPC capacities of plants. The function $@\text{for}(\text{Plant(i):...})$ stipulates the constraints for each plant. In the sum expression of this line, summation is taken over all options with $m = 1$ (this is achieved by the expression $m \# \text{EQ} \# 1$). Note that m stand for CTypes and 1 for this stands for OPC and 2 stands for PPC because the elements of CTypes are specified in the order OPC PPC in Line 7. The other constraints are self explanatory and are accompanied by comments. The terms 'EQ', 'GE' and 'GT' stand for 'equal to', 'greater than or equal to' and 'greater than' respectively.

In the model of Exhibit 1, the new plant constraint is included (Line 30 and Line 31) for the sake of completion. In this case, Plant M3 is taken as the new plant and stockists L8 to L17 are the ones which are in the same state of M3 (see Fig. 4.2).

Data can be imported or exported from external files such as ascii or excel files. Lines 281 and 284 import contribution (c) and demands (q) data in this fashion. For doing this we must first of all create an excel file named 'Data.xlsx' and create ranges named 'Contribution' and 'Demands' in the excel file. The statement in Line 287 exports the values of x to the range named 'Vol' in 'Data.xlsx' file. The number of cells in this range must be same as the number of values of variables imported or exported.

Once the model file is prepared in LINGO, it is run using LINGO solver. LINGO outputs the data and the solution and also passes on the solution to the excel file. It is possible to view the selected portions of the output using LINGO options. Using this, the nonzero values of x , the solution, are presented in Exhibit 2. The total contribution of the optimal solution is Rs.117779000/-.

It has been observed that the sale prices have no role in maximizing the contribution. If we replace the objective function by $\sum_{(i,j,k,l,u)} (f_{ijk} + v_{il})x(i,j,k,l,u)$ given in (4.15) and minimize it, we should still get an optimal solution. If there are multiple optimal solutions, then it is not necessary that we get the same optimal solutions for the two objective functions. However, in this case it so happened that we got the same optimal solution in both cases (see Problem 4.7).

4.7.2 Solution of the Reduced Problem for Example 1

In this subsection we shall solve the reduced problem of Example 1. For the reduced case, the product combinations are reduced to cement type. For each stockist, the demands are specified as total weights of OPC and PPC products. Table 4.5 presents these weights. The coefficients of objective function are given in Table 4.6. The options for this problem are specified as 5-tuples (Plant, SPO, TMode, Stockist, CType). There are 173 options for this reduced problem (as opposed to 239 options

Exhibit 2: Solution of Example 1 using LINGO Model

Variable	Value
X(M1, D0, RR, L6, PPC, BR4, G43, HFDE)	3750.000
X(M1, D0, RR, L17, PPC, BR4, G43, HFDE)	500.0000
X(M1, D0, RR, L17, PPC, BR7, G43, HFDE)	900.0000
X(M3, D0, RR, L5, OPC, BR8, G53, PAPER)	900.0000
X(M3, D0, RR, L5, PPC, BR7, G43, HFDE)	650.0000
X(M3, D0, RR, L7, OPC, BR2, G53, PAPER)	1750.000
X(M3, D0, RR, L7, PPC, BR3, G43, HFDE)	1900.000
X(M3, D0, RR, L7, PPC, BR7, G43, HFDE)	650.0000
X(M3, D0, RR, L8, OPC, BR1, G43, PAPER)	3000.000
X(M3, D0, RR, L8, PPC, BR4, G43, HFDE)	750.0000
X(M3, D0, RR, L11, OPC, BR2, G53, PAPER)	550.0000
X(M3, D0, RR, L11, PPC, BR7, G43, HFDE)	1300.000
X(M3, D0, RR, L12, OPC, BR1, G43, PAPER)	1800.000
X(M3, D0, RR, L12, PPC, BR3, G43, HFDE)	2200.000
X(M3, D0, RR, L12, PPC, BR4, G43, HFDE)	3000.000
X(M3, D0, RR, L13, OPC, BR2, G53, PAPER)	950.0000
X(M3, D0, RR, L13, OPC, BR1, G43, PAPER)	500.0000
X(M3, D0, RR, L13, OPC, BR8, G53, PAPER)	500.0000
X(M3, D0, RR, L13, PPC, BR3, G43, HFDE)	600.0000
X(M3, D0, RR, L13, PPC, BR4, G43, HFDE)	1000.000
X(M3, D0, RR, L13, PPC, BR7, G43, HFDE)	500.0000
X(M1, D1, RR, L14, OPC, BR1, G43, PAPER)	2000.000
X(M1, D1, RR, L14, PPC, BR4, G43, HFDE)	1100.000
X(M1, D1, RR, L17, OPC, BR8, G53, PAPER)	2000.000
X(M1, D3, RR, L15, OPC, BR8, G53, PAPER)	3000.000
X(M1, D3, RR, L15, PPC, BR3, G43, HFDE)	1300.000
X(M1, D3, RR, L16, OPC, BR2, G53, PAPER)	2000.000
X(M1, D3, RR, L16, OPC, BR1, G43, PAPER)	1000.000
X(M1, D3, RR, L16, OPC, BR8, G53, PAPER)	1000.000
X(M3, D3, RR, L9, OPC, BR8, G53, PAPER)	2500.000
X(M3, D3, RR, L10, OPC, BR2, G53, PAPER)	1000.000
X(M3, D3, RR, L10, PPC, BR3, G43, HFDE)	2000.000
X(M3, D3, RR, L10, PPC, BR4, G43, HFDE)	1000.000
X(M3, D2, TR, L1, OPC, BR1, G43, PAPER)	1000.000
X(M3, D2, TR, L1, PPC, BR3, G43, HFDE)	600.0000
X(M3, D2, TR, L1, PPC, BR7, G43, HFDE)	1500.000
X(M3, D2, TR, L2, OPC, BR8, G53, PAPER)	750.0000
X(M3, D2, TR, L2, PPC, BR3, G43, HFDE)	500.0000
X(M3, D2, TR, L3, OPC, BR2, G53, PAPER)	1200.000
X(M3, D2, TR, L3, PPC, BR4, G43, HFDE)	1400.000
X(M3, D2, TR, L4, OPC, BR2, G53, PAPER)	1000.000
X(M3, D4, TR, L6, OPC, BR1, G43, PAPER)	800.0000
X(M3, D4, TR, L6, PPC, BR4, G43, HFDE)	250.0000

for the original problem). The LINGO model for the reduced problem is very similar to the previous one and is given in Exhibit 3. The solution from LINGO is presented in Exhibit 4.

Exhibit 3: LINGO Model for the Reduced Problem

Model:

sets:

Plants/M1 M2 M3/:S,T,MRoad,MRail; ! Six Plants - 3 in AP and 3 in TN;

SPOs/D0 D1 D2 D3 D4/:SRoad; ! SPO is a stock point office (Depot).

For real problem, number of SPOs = 75. D0 stands for

Direct supply from plant;

CTypes/OPC PPC/;: ! Number of cement types 2, OPC, PPC and SRC, but SRC is not considered;

Stockists/L1..L17/;: ! Number of stockists 1000 per month out of 2500 base;

TModes/RR TR/;: ! Number of transport modes is 3: road direct (RD), (road, road), (rail(T), road(R));

PCs(Stockists,CTypes):q;

Routes(Plants,SPOs,TModes):;

! use the indices as i for plant, j for spo, u for TModes,

l for Stockists, m for CType, b for Brands, g for Grades, k for packing;

Options(Plants,SPOs,TModes,Stockists,CTypes):x,c;

endsets

! Minimize total contribution;

min = @sum(Options:c*x);

!Plant OPC Capacity constraints;

@for(Plants(i):@sum(Options(i,j,u,l,m)|m # EQ# 1:x)_i= S(i));

!Plant total Capacity constraints;

@for(Plants(i):@sum(Options(i,j,u,l,m):x)_i= T(i));

!Plant road transport capacity constraints;

@for(Plants(i):@sum(Options(i,j,u,l,m)|u # EQ# 1:x)_i= MRoad(i));

!Plant rail transport capacity constraints;

@for(Plants(i):@sum(Options(i,j,u,l,m)|u # EQ# 2:x)_i= MRail(i));

!SPO transport capacity constraints, D0 being direct supply no constraint for j=1;

@for(SPOs(j):@sum(Options(i,j,u,l,m)|j # GT# 1:x)_j= SRoad(j));

! Demand constraints;

@for(PCs(l,m):@sum(Options(i,j,u,l,m):x) = q(l,m));

! New Plant constraint;

@sum(Options(i,j,u,l,m)|i # EQ# 3 # AND# 1 # GE# 8:x)

- @sum(Options(i,j,u,l,m)|i # EQ# 3:x)/2 <= 0;

DATA:

! Stokist PC requirements;

PCs = @OLE('C:\ path\ EgF2.xlsx', 'PCs');

Options = @OLE('C:\ path\ EgF2.xlsx', 'Options');

c = @OLE('C:\ path\ EgF2.xlsx', 'Cost');

q = @OLE('C:\ path\ EgF2.xlsx', 'Demands');

S = 25000 25000 30000;

T = 40000 38000 40000;

MRoad = 320000 35000 29000;

MRail = 15000 13000 9000;

SRoad = 29000 22000 27000 31000 28000;

@OLE('C:\ path\ EgF2.xlsx', 'Vol')=x;

ENDDATA

END

Exhibit 4: Solution of the Reduced Problem

Variable	Value
X(M1, D0, RR, L6, PPC)	1550.000
X(M1, D0, RR, L17, PPC)	1400.000
X(M2, D0, RR, L9, OPC)	2500.000
X(M2, D0, RR, L10, OPC)	1000.000
X(M2, D0, RR, L14, OPC)	2000.000
X(M2, D0, RR, L15, OPC)	3000.000
X(M2, D0, RR, L16, OPC)	4000.000
X(M3, D0, RR, L5, OPC)	900.0000
X(M3, D0, RR, L5, PPC)	650.0000
X(M3, D0, RR, L6, PPC)	2200.000
X(M3, D0, RR, L7, OPC)	1750.000
X(M3, D0, RR, L7, PPC)	2550.000
X(M3, D0, RR, L8, OPC)	3000.000
X(M3, D0, RR, L8, PPC)	750.0000
X(M3, D0, RR, L11, OPC)	550.0000
X(M3, D0, RR, L11, PPC)	1300.000
X(M3, D0, RR, L12, OPC)	1800.000
X(M3, D0, RR, L12, PPC)	5200.000
X(M3, D0, RR, L13, OPC)	1950.000
X(M3, D0, RR, L13, PPC)	2100.000
X(M1, D1, RR, L14, PPC)	1100.000
X(M3, D3, RR, L10, PPC)	3000.000
X(M3, D3, RR, L15, PPC)	1300.000
X(M2, D1, TR, L17, OPC)	2000.000
X(M3, D2, TR, L1, OPC)	1000.000
X(M3, D2, TR, L1, PPC)	2100.000
X(M3, D2, TR, L2, OPC)	750.0000
X(M3, D2, TR, L2, PPC)	500.0000
X(M3, D2, TR, L3, OPC)	1200.000
X(M3, D2, TR, L3, PPC)	1400.000
X(M3, D2, TR, L4, OPC)	1000.000
X(M3, D4, TR, L6, OPC)	800.0000
X(M3, D4, TR, L6, PPC)	250.0000

4.7.3 Application to the Real Problem

In order to study the impact of applying the optimization tool, one month's past data were collected for which the actual contribution was known. This involved 8743 orders comprising 47 product combinations from 1092 stockists with total sales volume of 590101 tonnes. The total sales price was Rs.263,18,50,460/- and the contribution as per the actual distribution was Rs.868486641/-. The optimum plant loading and distribution were obtained by solving the problem using the formulations discussed in this chapter. The contribution increased to Rs.890476557/-. Thus, there was a savings of Rs.21989916/-.

4.8 Decision Support System

Development of methodology was only one part of the project. The major task of the project involved developing numerous codes for extracting the inputs from the company's database system, formatting the inputs to suit the OR solver and integrat-

ing the solver with the company's database system, developing codes for providing outputs in various formats as per the company's specifications. All this took about 6 months and the consultant was solely responsible for these tasks. A complete decision support system was developed by the consultant and implemented it for the company.

4.9 Summary

In this chapter we discussed a project from a process industry. The project is about macro level planning of production and dispatches from various plants of the company. An important contribution to this project is the observation that it is possible to achieve the company's objective of maximizing contribution even without knowing the fluctuating cement prices. At the first instance, the size of the problem appeared formidable. The size is not a real issue in the light of the present day's computing power available and the fact that the underlying optimization problem is a linear programming problem which can be solved very efficiently. Further, it is observed that the size of the problem can be reduced drastically by using the special structure of the problem. This project has helped in building confidence in handling large scale problems. This chapter also outlined some practical difficulties that arise in dealing with projects like the ones mentioned in this chapter.

Problems

4.1. For the miniature version Example 1, develop a systematic way of determining the number of decision variables and check that this number is 239 (hint: use the data structure of Fig. 4.4). Enlist all possible 5-tuples (i, j, k, l, u) associated with the decision variables.

4.2. In the real project, 47 of the 75 SPOs are connected to the six plants by rail. Consider the graph with 53 nodes (each node is either a plant or an SPO with rail connectivity). Join a plant-node and an SPO-node with an edge if there is a rail connection between the corresponding plant and SPO. In the real project, the number of edges in this graph is equal to 171. Compute the maximum number of possible 5-tuples (i, j, k, l, u) associated with the decision variables. State your assumption in your computation.

4.3. Figure 4.3 presents the transportation network for Example 1. However the full network for the problem should also include products. Perhaps one way of doing this is as follows. Add a node for each product and connect these nodes with stockists who require these products. Figure 4.8 shows how to do this for stockists L_1 and L_2 and products P_2, P_3, P_4, P_6 required by them. When you extend Fig. 4.3 in this fashion by including all six products you get the full network for Example 1. Take

a big sheet of paper and draw the full network or alternatively do it on computer. Compute the number of edges in this network. What will be the approximate number of edges in the real project?

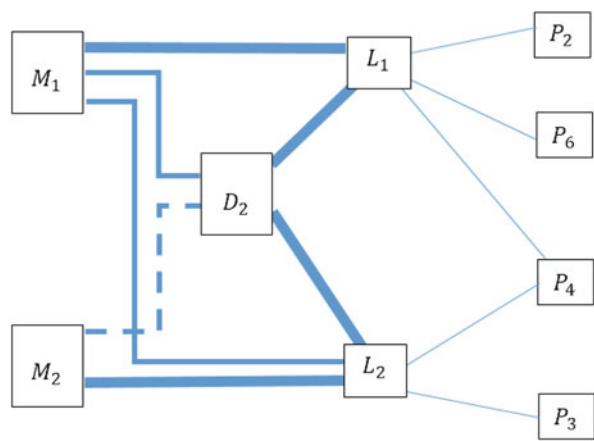


Fig. 4.8 Part of the network for Example 1 with products

- 4.4.** For Example 1, using the data of Fig. 4.4, write a programme to compute f_{ijkus} for all permissible routes.
- 4.5.** Consider the new variables $x'(i, j, k, l, u, d)$ introduced in Sect. 4.4.2. Show that the constraints with these new variables for the decision making problem of this project will be same as the constraints of the problem stated in Sect. 4.3.7.
- 4.6.** In Exhibit 3 on page 85, the complete list of options is not given. Using Fig. 4.4, prepare this complete list, and for each item in your list compute the unit transportation costs and the contributions (you have to use Fig. 4.4 and Table 4.4). Why (M1,D2,RR,L1,OPC,Br1,G43,Paper) is not a possible option?
- 4.7.** The variable costs for Example 1 are shown in Table 4.7. Using Fig. 4.4 and Table 4.7, compute the objective coefficients $f_{ijkus} + v_{il}$ given in (4.15) and verify that the solution given in Exhibit 2 is an optimal solution for minimizing problem in Formulation 1 for Example 1 with the objective function replaced by $\sum_{(i,j,k,l,u)} (f_{ijkus} + v_{il})x(i, j, k, l, u)$ (remember that the objective is to minimize this objective function).

Table 4.7 Variable costs for Example 1

Cement Type	M1	M2	M3
OPC	1300	1000	900
PPC	1200	1450	1100

References

1. Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: Network optimization. In: Pardalos, P.M., Resende, M.F.C. (eds.) *Handbook of Applied Optimization*. Oxford University Press, New York (2002)
2. Bazaraa, M.S., Jarvis, J.J., Sherali, H.D.: *Linear Programming and Network Flows*. Wiley, Hoboken (2010)
3. Broy, M.: Software engineering—from auxiliary to key technologies. In: Broy, M., Dener, E. (eds.) *Software Pioneers*, pp. 10–13. Springer, Heidelberg (2002)
4. Murty, K.G.: *Network Programming*. Prentice-Hall, Englewood Cliffs (1992)
5. Murty, K.G.: *Linear and Combinatorial Programming*. Wiley, New York (1976)
6. Taha, H.A.: *An Introduction to Operations Research*. Prentice Hall, Englewood Cliffs (2007)

Chapter 5

Deckle Optimization in Paper Industry

Abstract This chapter deals with a case study on deckle optimization problem encountered in paper and paperboard industries. The problem is crucial for the industry as it concerns huge financial implications. The decision making problems are too complex to handle manually and software solutions are indispensable. Deckle optimization problems are solved typically using the one-dimensional cutting stock problem and travelling salesmen problem models. Intelligent modeling is essential to solve the cutting stock problems arising in the context of deckle optimization. With a brief introduction to the products and processes relevant to this case study, the chapter presents various formulations to handle the challenges. Time and material costs are two conflicting components of the deckle optimization problem. The formulations proposed aid the user to explore the consequences of weighing these two conflicting factors while choosing the strategy for optimal decision making. The formulations and solutions are illustrated with examples and a practical instance of the problem picked up from the past data. A software solution has been developed for the client which uses excel as frontend as desired by the client.

5.1 Introduction

Paper industry is like a mine for applications of Operations Research (OR) and Statistics. OR plays a crucial role in production planning. This project deals with planning at the winder section of paper industry. The problem is that of planning production of child reels from jumbo reels for a given run of production based on customer orders. Trim losses and excessive production are critical parameters at this stage of production. Any savings here are usually very substantial and paper industries invest huge money in procuring software for this planning. This project was taken up at the request of a leading paper and paperboard manufacturing company

which has production establishment at multiple locations. In fact, the company has an expensive software solution for the problem at one of their establishments and wants an economically viable solution for their establishments at other locations.



Fig. 5.1 Jumbo roll

The problem that we are going to deal with in this chapter is known as deckle optimization in paper/paperboard industry. It pertains to the Winder Section of paper/paperboard manufacturing process [1] where the produced paper/paperboard in the form of jumbo rolls (Fig. 5.1) is cut into smaller size rolls as per customer orders. In this operation, a part of the jumbo rolls gets scrapped resulting in material losses. If planned properly, the extent of scrap can be reduced and even marginal savings will result in huge money, usually running into several lakhs of rupees. But the planning is a very challenging problem and involves intensive calculations and it is almost impossible to get the best solutions manually. Actually the problem requires mathematical modelling and efficient solution procedures. Efficient professional packages from big companies are available and they are generally very expensive. The company for which this project was taken up has four manufacturing units across India and has a professional deckle optimization package for one of their manufacturing units. The cost of extending this package to their other units is prohibitively large. Consequently, this project was taken up to provide an alternative

solution to the problem. A software solution is developed with Microsoft Excel as the frontend so that it is very user-friendly as the management staff is used to Excel for most of their work.

The problem is a nice application of cutting stock and travelling salesman problems [3–7, 9, 10]. Finding the best solutions to these problems is hard. In this chapter, we shall present the problem, different ways of solving it, the difficulties in various approaches and an efficient solution for the practical problem at hand [2, 8]. The software solution developed as a part of this project acts as a decision support system for the managers.

This chapter covers a detailed account of the process and the problem, mathematical modeling of the problem, proposed algorithms for solving the problem, the computational aspects and the results of applying the solutions to an instance of the problem from the past data. The presentation of the work starts in Sect. 5.2 with an introduction to the part of the manufacturing process relevant to this project. The cutting stock problem, one of the two basic models used for formulating the problem, is presented and discussed in Sect. 5.3. In order to understand the application of cutting stock problems to deckle optimization, one must understand the operations at winder section and the subsequent operations in sheeter section and the various calculations that go into the planning. These details are discussed in Sect. 5.4. The cutting stock problem is the key to material savings at the winder section and is discussed at length in Sect. 5.5. Different formulations are presented. This section sets up the framework for solving the deckle optimization as an operations research problem. The other important problem in deckle optimization pertains to reducing production time after determining the cutting plans. This involves sequencing the production activities and is formulated as a traveling salesman problem. This part is presented in Sect. 5.6. A brief description of the complete solution to the deckle optimization problem is presented at the end of this section. Section 5.7 provides the summary of the project.

5.2 Products and the Process

This section briefly describes the products and the process concerned with this project. This is to facilitate the understanding of the modeling and mathematical formulations for solving the deckle optimization problem. The process pertains to what is known as Finishing House in paper/paperboard industry.

5.2.1 Products

Typically customers place orders for paper or paperboard in two forms, namely, sheets and reels. The orders for reels are referred to as **reel orders** and the orders for sheets are referred to as **sheet orders**. The orders will specify the features of

the product and the quantity required. Features of a product (paper or paperboard) include grade, GSM (grams per square meter) and caliper (thickness of the paper or paperboard usually specified in microns).

Reel Orders

The reel orders are packed in the form of reels and they are specified by the product features, the dimension of the reel, and the weight of the quantity required. Reels being cylindrical in shape, their dimensions are specified by two parameters – the reel diameter and the width of the reel. As the paper or paperboard is wound on a *core* pipe, the diameter of the core pipe is also specified. Thus, a typical reel order will look like: 6837 kg of ABC grade with 230 GSM, 300 μ caliper paperboard in reels of dimension 95 cm diameter and 60 cm width wound on a 3-inch core pipe (it is customary to use microns for caliper and inches for core diameter in the paper/paperboard industry, at least in the company where this project was carried out). On the basis of this information, the number of reels for a particular order is calculated using GSM, caliper and the reel dimensions. The following formula gives the weight M in kilograms of one reel with GSM g (through out this chapter, ‘weight’ will stand for ‘mass’ as it is customary to call mass as weight), caliper c μ , diameter D cm, width w cm, core diameter d inches

$$M = \frac{\pi(D - 2.54d)^2 wg}{1,000c}. \quad (5.1)$$

Sheet Orders

Sheet orders are specified by product features, the sheet dimensions and the quantity by weight. In addition, there is one more specification for a sheet order. This specifies the direction of the grain or the fiber direction in the sheet. To describe this, we first need to introduce the direction in a roll. The direction in which a roll unfolds (tangential direction to the circle of cylinder) is called the long way (this is also referred to as machine direction (MD)) and the direction across the width of roll is called the short way. If you view jumbo or master roll as a cylinder, the side along the height of the cylinder is the short way and it is also called the cross direction (CD). See Fig. 5.2. For a sheet (rectangular in shape), call the longer edge as length and the shorter edge as its breadth. Each sheet order comes with direction specification as one of the following: grain long way (GLW), grain short way (GSW) and grain any way (GAW). If the specification is GLW, then the sheets should be cut in such a way that the length of the sheet is in the long way; if the specification is GSW, then the sheets should be cut in such a way that the length of the sheet is in the short way; and if the specification is GAW, then the sheets can be cut in any way. A typical sheet order will look like: 5000 kg of XYZ grade with 230 GSM, caliper 280 μ , sheet dimension 62.5 \times 86 cm, and the direction is GLW. For the purpose

of this project, the **width of a sheet order** is defined as the dimension of the sheet in the CD direction. Thus, width of a sheet order is equal to breadth of the sheet if the specification is GLW, and is equal to length if the specification is GSW. For sheet orders with specification GAW, width can be taken as length or breadth; both options are open in this case.

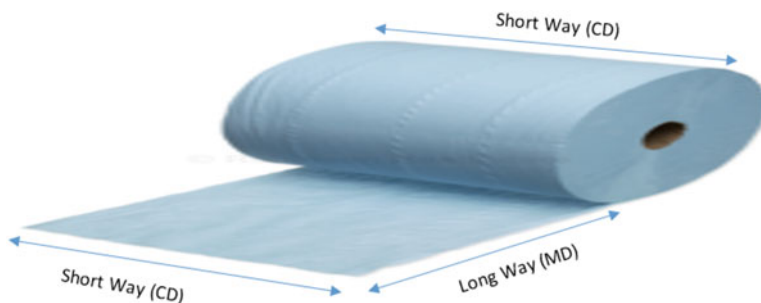


Fig. 5.2 Jumbo roll with directions specified

The first step in producing sheets for sheet orders is that of producing reels of suitable dimensions, and then convert them into sheets. Usually there is no restriction on the diameter of these reels but the width of the child reel has to be a multiple of the width of the sheet order plus the trim allowance.

5.2.2 Winder

We have just observed that irrespective of the type of orders, reel or sheet, the jumbos are to be reduced to smaller size reels some of which go directly as reel orders and the others will be processed to convert them into sheet orders. These small size reels are termed as **child reels**. Child reels are produced from jumbo reels (see Figs. 5.1 and 5.3) in the Winder section using a machine called winder. Jumbo reels are much bigger in size (about 2 m in diameter and 3.8 m in width) compared to child reels. Child reel diameters are usually around 1 m in diameter and the width varies according to the customer requirements and the sheet sizes required. Figure 5.3 shows the picture of a winder, a jumbo and child reels.

A winder is equipped with several knives which can be set at desired positions across the width of the winder to produce the child reels of desired widths. The winder unwinds the jumbo at the input end and rewinds the child reels at the other end. Between these two ends are the knives doing their job of slitting the wide paper or the paperboard. The winder unwinds the jumbo until the child reels attain their required diameter. This phase is called a single run. In a single run, it is only possible to produce child reels of the same diameter. Child reels produced in a single run is called a **set**. Thus, all the child reels in a set will necessarily have the same diameter

but may have different widths. Depending upon the size of the jumbo diameter, a jumbo can produce more than one set. Normally, a jumbo will produce two sets and occasionally the management may plan to produced three sets.

The number of knives on the winder depends on the make of the winder. Some winders have seven knives, some have ten knives and so on. If k is the number of



Fig. 5.3 Winder machine converts jumbo rolls to child reels

knives on a winder, then the maximum number of child reels that can be produced in a single run is at most $k - 1$ (see Fig. 5.4). The knives may be repositioned after a set is completed. Changing the knife positions is a time consuming activity and minimizing this time is crucial and is an important optimization problem.



Fig. 5.4 A winder with seven knives. This can cut at most six child reels in a single run. It is not necessary to use all the knives in a run but minimum two knives should be used. To cut p child reels $p + 1$ knives are used, $p = 1, 2, \dots, 6$

5.2.3 *Deckle*

Consider a jumbo whose width is 324 cm. Due to unevenness of the edges of jumbo, a small portion of the edges becomes unusable and is trimmed off. Usually it is 1 or 2 cm on either side. Assuming this to be 2 cm on both sides, the effective width of the jumbo available for the child reels is 320 cm. Suppose 4 child reels are required, two of them with width 80 cm and the other two of width 60 cm each. If these reels are cut from the jumbo, then there is a trim loss of a reel whose width is 40 cm. Note that the 4 cm mentioned above on the edges of jumbo is not considered as trim loss. The 320 cm of the jumbo will be called the effective deckle width or simply **deckle width**.

5.2.4 *Sheeters*

The operation of converting child reels into sheets is called sheeting. This operation is performed on a machine called **sheeter**. Sheeter to some extent is like a winder. It is also equipped with knives but the knives are classified into longitudinal and transversal knives. The longitudinal knives are positioned across the width of the child reel which slit the child reel into subreels. The purpose of transversal knife is to chop the unfolding reel into sheets. At the input end, a child reel is mounted and the sheeter unwinds this child reel, if necessary by slitting the child reel into subreels, and chops it periodically with the transversal knife to cut the reel into sheets. It is important to understand the process of sheeting because it also plays an important role in deckle optimization problem. This is best understood with the help of an example. Suppose a customer requires 40000 sheets of size 60×90 cm. To produce these sheets, child reels of 60 cm wide are used. When this child reel is mounted on a sheeter, the sheeter unwinds the reel and chops the reel at every 90 cm with the transversal knife, and the sheets (of dimension 60×90 cm) thus produced are piled into pallets. Assume that each reel of diameter 1 m with 60 cm width yields 10000 sheets of dimension 60×90 cm. Thus, to produce 40000 sheets mentioned above, one needs four 60 cm child reels of diameter 1 m. This is the case with some type of sheeters which do not use any longitudinal knives.

The sheeters which are equipped with longitudinal knives require trimming of the child reels just like trimming jumbos at winder. In such sheeters usually there is a 1 cm trim required at the two ends of a child reel. In this project we have two types of sheeters. One type of sheeters have no longitudinal knives and they can only chop the reel into sheets. On these sheeters there is no trim loss. These sheeters, limited in number, are available with outside vendors and their quality of sheeting is generally poor. The other type of sheeters are within the company and are equipped with four longitudinal knives. These are based on advanced technology and the quality of sheeting is good. But certain trim loss is inevitable in these sheeters and the extent of trim loss depends upon how they are utilized. Three ways of producing the 40000 sheets mentioned above is pictorially depicted in Fig. 5.5. In the first case only two

longitudinal knives are used (part (a)); in the second case three longitudinal knives are used (part (b)); and in the third case all the four longitudinal knives are used (part (c)). The trim loss is same in all the three cases (2 cm) but this is in absolute terms. If we compare the trim loss per sheet produced, then the relative losses are 2, 1 and $\frac{2}{3}$ for (a), (b) and (c) respectively. Moreover, the productivity in case (c) is thrice that of case (a), and the productivity in case (b) is twice that of case (a). To meet the order requirement of 40000 sheets, we can take two child reels, one of width 62 cm and one of width 182 cm, and cut them according to (a) and (c) respectively. The other option is, use four child reels each of width 62 cm and cut them according to part (a) to meet the order. Obviously, the former option is more effective and economical. Therefore, the deckle optimization has to explore the utility of these options for a given order.

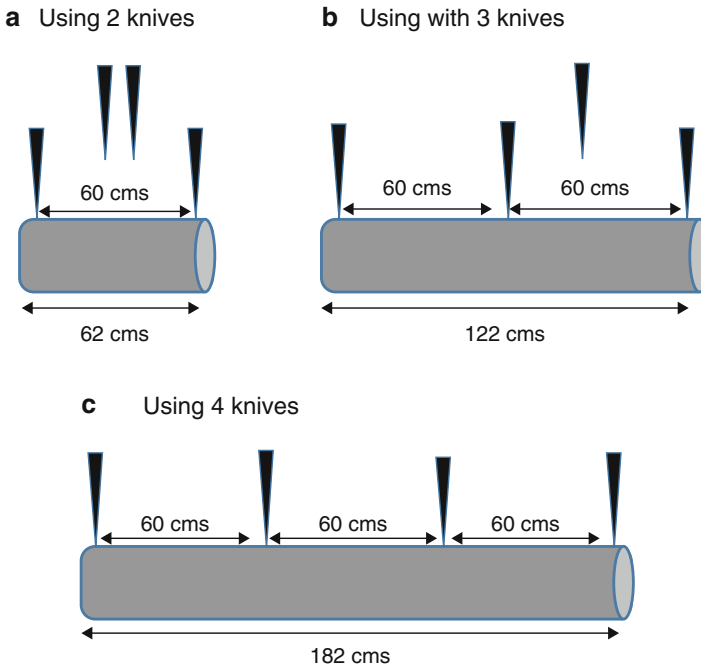


Fig. 5.5 Company's sheeter with four longitudinal knives. Knives not touching the roll are disentangled, and are not used in parts (a) and (b)

5.3 Cutting Stock Problem

Cutting stock model is one of the most widely applied OR model in real world applications [3, 4, 6, 7, 9, 10]. This model is used in material optimization studies. The cutting stock problem can be described best with the help of a simple example.

Consider the paper rolls or cello tape rolls. What we see in shops are usually have small widths. For example, we may find cello tapes of size 3, 4, 5, or 6 in. in width. These are produced from 12-in. wide cello tape rolls. Imagine a customer has placed an order as follows. Supply 3-in. tapes 200 numbers, 4-in. tapes 150 numbers, 5-in. tapes 275 numbers, and 6-in. tapes 100 numbers. The words ‘roll’ and ‘tape’ are used interchangeably. One optimization problem here is to find a way to supply the order using minimum number of 12-in. rolls. Another problem is to minimize the trim losses. When a 12-in. roll is cut into smaller widths, certain portion of the master roll goes as scrap. This scrap is termed as trim loss (see the next subsection for a clear definition of trim loss). This problem is known as one-dimensional cutting stock problem. For ease of presentation, we shall refer this problem as **cello tape problem**. This problem is only a subproblem of the deckle optimization as will be seen later.

In this section we will see how this problem is formulated, what are local and global optimal solutions, different formulations of solving this problem and some solution procedures used for this project. See [7] for an elaborate discussion the problem and connected literature.

5.3.1 Cutting Patterns

To meet the requirements in the cello tape problem, the 12-in. rolls, called the master rolls, are to be cut into the required widths first. The master roll can be slit into four 3-in. tapes, or it can be slit into three 4-in. tapes, or it can be slit into two 3-in. tapes and one 6-in. tape, and so on. In this problem, there are four different widths required, namely, 3, 4, 5 and 6 in. In a general problem, imagine rolls of k different widths, say, w_1, w_2, \dots, w_k are required in quantities q_1, q_2, \dots, q_k respectively. A cutting pattern for this problem is a k -dimensional integer vector (c_1, c_2, \dots, c_k) which specifies that cut c_j pieces of width w_j from a single master roll, $j = 1, 2, \dots, k$. According to this notation, the three cutting patterns cited above for the cello tape problem are $(4, 0, 0, 0)$, $(0, 3, 0, 0)$ and $(2, 0, 0, 1)$. The widths and quantities for this problem are $w_1 = 3, w_2 = 4, w_3 = 5$, and $w_4 = 6$ and $q_1 = 200, q_2 = 150, q_3 = 275$, and $q_4 = 100$. For a (c_1, c_2, \dots, c_k) to be a cutting pattern, the c_j s must satisfy the condition that $\sum_{j=1}^k w_j c_j$ is less than or equal to width of the master roll. Thus, $(1, 1, 1, 1)$ is not a cutting pattern. Let $C = \{C_1, C_2, \dots, C_N\}$ denote the set of all cutting patterns. The number of cutting patterns is finite but grows exponentially as the number of widths (k) and the width of the master roll, say W (for the cello tape problem $W = 12$ in.), increase. For a cutting pattern (c_1, c_2, \dots, c_k) , the difference $W - \sum_{j=1}^k w_j c_j$ is called the trim loss. Cutting patterns may be many but several of them may be redundant. For example, patterns such as $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, etc. are redundant in the cello tape problem. The meaningful cutting patterns for the cello tape problem are listed in Table 5.1.

Table 5.1 Cutting patterns and plans for cello tape problem

Meaningful cutting patterns			Cutting plans	
ID	Pattern	Trim loss (inches)	Plan 1	Plan 2
C_1	(4, 0, 0, 0)	0	50	12
C_2	(2, 1, 0, 0)	2	0	0
C_3	(2, 0, 1, 0)	1	0	1
C_4	(2, 0, 0, 1)	0	0	0
C_5	(1, 1, 1, 0)	0	0	150
C_6	(0, 3, 0, 0)	0	50	0
C_7	(0, 1, 0, 1)	2	0	0
C_8	(0, 0, 2, 0)	2	138	62
C_9	(0, 0, 1, 1)	1	0	0
C_{10}	(0, 0, 0, 2)	0	50	50

5.3.2 Cutting Plans

A solution to a cutting stock problem is the specification of cutting patterns along with their frequencies. Consider the four widths of the cello tape problem. A typical solution used by the shop floor managers is the following: cut 50 master rolls using (4, 0, 0, 0); cut 50 master rolls according to (0, 3, 0, 0); cut 138 master rolls according to (0, 0, 2, 0); and cut 50 master rolls according to (0, 0, 0, 2). Here 50, 50, 138 and 50 are the frequencies of the respective cutting patterns. This plan is shown as Plan 1 in Table 5.1. The specification of cutting patterns along with the respective frequencies, a solution to the cutting stock problem, is called a cutting plan. The cutting plan for the cello tape problem just specified (Plan 1) uses 288 master rolls and has a trim loss of 276 inches of roll. The cutting stock optimization problem seeks solution to one or both of the following two problems:

- Find a cutting plan with smallest number of master rolls
- Find a cutting plan with minimum trim loss

The problem of finding best cutting plan can be modelled as a mathematical problem. A cutting plan obtained using a mathematical model is shown as Plan 2 in Table 5.1. This plan uses 275 rolls and has much less trim loss (125 in. of roll) compared to Plan 1.

5.3.3 Formulation of One-Dimensional Cutting Stock Problem

We first present the one-dimensional cutting stock problem mathematically. We use the paper industry example of rolls of bigger size being reduced to smaller size to present the mathematical model.

One-Dimensional Cutting Stock Problem

We are given a positive real number W , called the width of the master roll, and two k -dimensional real vectors $w = (w_1, w_2, \dots, w_k)$ and $q = (q_1, q_2, \dots, q_k)$ with all w_i s positive and less than W , and all q_i s positive integers. A cutting pattern is a k -dimensional nonnegative integer vector (c_1, c_2, \dots, c_k) which satisfies $\sum_{i=1}^k w_i c_i \leq W$. Say that cutting pattern (c_1, c_2, \dots, c_k) is redundant or dominated if there exists another cutting pattern $(c'_1, c'_2, \dots, c'_k)$ such that $c'_i \geq c_i$ for all i with strict inequality for at least one i . Let $U = \{C_1, C_2, \dots, C_N\}$ be the collection of all undominated cutting patterns (note that the cutting patterns are denoted using upper case letter with subscripts (C_j)). Let $C_j = (c_{j1}, c_{j2}, \dots, c_{jk})$, $j = 1, 2, \dots, N$, be j th cutting pattern and let C be the $N \times k$ matrix whose j th row is C_j . Let $u = (u_1, u_2, \dots, u_N)$ be a nonnegative integer vector. Consider the functions $f(u) = \sum_{j=1}^N u_j$ and $TL(u) = \sum_{j=1}^N u_j (W - \sum_{i=1}^k w_i c_{ji})$. The cutting stock problem with minimum number of master rolls as the objective function is to find a u for the problem (5.2).

$$\begin{aligned} &\text{Minimize } f(u) \\ &\text{subject to } uC \geq q, \\ &u = (u_1, u_2, \dots, u_N) \text{ is a nonnegative integer vector.} \end{aligned} \tag{5.2}$$

The cutting stock problem with trim loss as the objective function is to find a u for the problem (5.3).

$$\begin{aligned} &\text{Minimize } TL(u) \\ &\text{subject to } uC \geq q, \\ &u = (u_1, u_2, \dots, u_N) \text{ is a nonnegative integer vector.} \end{aligned} \tag{5.3}$$

The constraint $uC \geq q$ is imposed to ensure that the order quantities of small rolls are met.

For the cello tape problem, $W = 12$, $w = (3, 4, 5, 6)$, $q = (200, 150, 275, 100)$ and the rows of the matrix C are listed in Table 5.1. The trim loss objective function for this problem is $TL(u) = 2(u_2 + u_7 + u_8) + u_3 + u_9$. A solution to problem (5.2) is $u = (12, 0, 1, 0, 150, 0, 0, 62, 0, 50)$. It uses 275 master rolls and has a trim loss of 125 in. of roll. On the other hand, a solution to problem (5.3) is $u' = (50, 0, 0, 100, 275, 50, 0, 0, 0, 50)$. It has 0 trim loss but uses 525 master rolls. Comparing the two solutions, it can be observed that minimizing trim loss has a tendency to use more master rolls and hence result in exceeding the order quantities. It is possible to think of a via media solution (see Problem 5.3) but practical problems have additional constraints to take care of this aspect.

5.3.4 Formulations for Finding Optimal Solutions

Any solution of cutting stock problem (5.2) or (5.3) will be called a global optimum solution to the corresponding problem. To solve these problems, first of all

one should have all the cutting patterns. Many cutting stock problems encountered in practice are big in size, that is, the number of cutting patterns in them will be too large, and in most cases it is difficult to enlist all cutting patterns (see Problem 5.2). It is possible to avoid listing all cutting patterns to solve the problem. We shall present two methods which have been used in this project. One of them is a heuristic and empirical experience shows that this method works well when applied to instances of deckle optimization problem. The other method is a formulation that asserts global optimal solution to problem (5.2). Both these methods do not require the cutting patterns to be listed, rather they generate the cutting patterns required. Furthermore, they accommodate an important constraint imposed by the number of knives in cutting a role. That is, if p is the number of knives, then any cutting pattern generated by these two methods will have at most $p - 1$ rolls (child rolls in the case of winder problem) cut. The problem and the two methods are presented below using the problem at winder as example.

Winder Problem

Consider the problem of reducing jumbo into child reels of widths w_1, w_2, \dots, w_k with the requirement q_i child reels of width w_i are needed. In addition, the following assumptions are made:

- Assumption 1. The diameter of all child reels is same.
- Assumption 2. Deckle width of the jumbo is W .
- Assumption 3. Number of knives on the winder is p .

Method 1 (Heuristic)

- Step 0. Set $\bar{q}_i = q$ and set $j = 0$.
- Step 1. If $\bar{q}_i = 0$ for $i = 1, 2, \dots, k$, then go to Step 6.
- Step 2. Solve the integer programming problem stated below.

$$\begin{aligned}
 &\text{Maximize } \sum_{i=1}^k w_i c_i \\
 &\text{subject to } \sum_{i=1}^k w_i c_i \leq W, \\
 &\quad \sum_{i=1}^k c_i \leq p - 1, \\
 &\quad c_i = 0 \text{ for } i \text{ such that } \bar{q}_i = 0, \\
 &\quad c_1, c_2, \dots, c_k \text{ are nonnegative integers.}
 \end{aligned} \tag{5.4}$$

- Step 3. If $c_i = 0$ for all i in the above problem, then go to Step 6.
- Step 4. Set $j \leftarrow j + 1$ and set $C_j = (c_1, c_2, \dots, c_k)$.

- Step 5. Set $\bar{q}_i \leftarrow \bar{q}_i - \lfloor \lambda \rfloor c_i$, where $\lambda = \min\{\frac{\bar{q}_i}{c_i} : c_i > 0\}$. Go to Step 1.
- Step 6. Set $N = j$ and let C be the matrix with its r th row as C_r , $r = 1, 2, \dots, N$.
- Step 7. Solve problem (5.2) with C and q .

The method above does the following. It starts with the requirement vector q , finds cutting pattern (c_1, c_2, \dots, c_k) with least trim loss to cut child reels for the unfilled widths, updates the requirement vector by removing as many child reels as possible using cutting pattern (c_1, c_2, \dots, c_k) , and records the cutting pattern as C_j . It repeats the process of generating cutting patterns with the updated requirements until all requirements are met. Using the cutting patterns C_j s generated in this process, problem (5.2) is solved. The constraint $\sum_{i=1}^k c_i \leq p - 1$ ensures that the cutting patterns will take care of the restriction imposed by the number of knives.

Problem Instance 1

This method is implemented in Excel with a professional solver at the back end to solve the integer programs stated in the method. An instance with order for 13 widths is shown in Fig. 5.6. The deckle width is 380 cm and the number of knives on the winder is 10. The solution obtained is shown in Fig. 5.7. The solution produced 13 cutting patterns and used them to produce the child reels. The resulting trim loss is 2.78 % which is below the benchmark of 4 %. The total number of master rolls used is 45. It took less than 1 min to produce the solution.

	A	B	C	D	E	F	G	H
1	Deckle	380	Excess / Shortage		OverAll	Total Deckle Width		17100
2	No. Orders	13	Min	Max	Efficiency	Total Deckle Width Used		16624
3	No. of CPs	40	0.00	2.00	97.22%	Trim Loss		476
4	NewCpNo.	41	Trim Loss		2.78%	1.90	<< Excess Production %	
5	LowerBoun	42.93	44	< Cur.Solu.	No.Patterns	13		
6	MaxCuts	10	0	< Balance				
7	Total Pieces Required >>		167		<< SUM(CpNew)			
8	Sno	Width	Requires	NewCp				
9	1	127	8	2				
10	2	82	4	0				
11	3	131	23	0				
12	4	78	29	0				
13	5	110	19	0				
14	6	120	7	0				
15	7	101	14	0				
16	8	59	4	0				
17	9	139	12	0				
18	10	65	9	0				
19	11	100	2	0				
20	12	66	21	0				
21	13	85	15	0				
22								

Fig. 5.6 Screen shot of the Excel program. Shows the input screen for an instance of a deckle optimization with requirement for child reels of 13 different widths to be produced from jumbos. Deckle width for this problem is 380 cm and the number of knives on the winder is 10

Problem Instance 2

Results of another instance of deckle optimization for 294 child reels with 17 different widths are as follows: Solution produced 17 cutting patterns, trim loss 0.56 %, total number of master rolls is 67, time taken is less than a minute. The deckle width and number of knives are same as in the first instance, that is, 380 cm and 10 knives. The input data for this problem are given in Table 5.2. It is observed that the number

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Trim Loss	LP Solution	Solution	Width	127	82	131	78	110	120	101	59	139	65	100	66	85
2				Req	8	4	23	29	19	7	14	4	12	9	2	21	15
3				Actual	8	4	23	30	19	7	15	4	12	9	2	23	15
4				Status	0	0	0	1	0	0	1	0	0	0	0	2	0
5				Jumbo No		1	2	3	4	5	6	7	8	9	10	11	12
6	0	4.00	4	1	0	1	0	1	2	0	0	0	0	0	0	0	0
7	0	4.00	4	2	0	0	0	0	2	0	1	1	0	0	0	0	0
8	0	3.00	3	3	0	0	1	0	1	0	0	0	1	0	0	0	0
9	0	1.00	1	4	0	0	0	0	0	1	0	0	0	4	0	0	0
10	0	2.00	2	5	0	0	0	1	0	0	2	0	0	0	1	0	0
11	0	9.00	9	6	0	0	0	2	0	0	0	0	1	0	0	0	1
12	0	3.00	3	7	0	0	0	1	0	0	0	0	0	0	0	2	2
13	1	6.00	6	8	1	0	0	0	0	1	0	0	0	0	0	2	0
14	3	0.00	1	9	0	0	0	1	0	0	1	0	0	0	0	3	0
15	4	2.00	2	10	0	0	1	1	0	0	1	0	0	0	0	1	0
16	17	4.00	4	11	0	0	2	0	0	0	1	0	0	0	0	0	0
17	53	5	5	12	0	0	2	0	0	0	0	0	0	1	0	0	0
18	126	1	1	13	2	0	0	0	0	0	0	0	0	0	0	0	0
19																	

Fig. 5.7 Solution to Problem Instance 1 with cutting patterns generated. Cutting patterns are shown as rows against the Jumbo No. The solution $u = (u_1, \dots, u_{13})$ is shown against the column heading ‘Solution.’ The numbers against the row titled ‘Status’ are the numbers of child reels in excess of what is required for the respective widths

Table 5.2 Second instance of deckle optimization problem with requirement for 17 child reels

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
w_i	136	100	74	56	51	61	50	40	64	121	73	95	109	116	77	62	114
q_i	21	15	25	3	15	29	9	20	29	25	21	18	26	8	2	12	16

of cutting patterns produced by this method is close to the number of child reels with different widths. In practice, it is desirable to have a control on the number of distinct cutting patterns as changing from one cutting pattern to another affects productivity. Method 2 described below uses a formulation that will control the number of cutting patterns.

Method 2 (Formulation to Control Number of Cutting Patterns)

In the above method, cutting patterns were generated sequentially one at a time. In Method 2, all the cutting patterns are generated in a single shot. The question

is: how many cutting patterns should be generated? Two approaches are proposed to address this problem. When the user specifies an upper limit on the number of cutting patterns, say N , then formulate the problem with N cutting patterns and explore if there is a feasible solution. But the problem with this approach is that if N is such that there is no feasible solution, then the solver keeps searching for a feasible solution and the process may continue for long. Instead it is better to use the second approach which is described below.

It is easy to get an upper bound on the number of master rolls required to meet the orders. One trivial upper bound can be taken as $N = k + \sum_{i=1}^k \lfloor q_i / \lambda_i \rfloor$, where $\lambda_i = \lfloor W / w_i \rfloor$ and $\lfloor a \rfloor$ is the largest integer less than or equal to a . For Problem Instance 1, this number is 62. Once this number is determined, then solve the problem (5.5).

$$\begin{aligned}
 & \text{Minimize } \sum_{j=1}^N u_j \\
 & \text{subject to } \sum_{i=1}^k w_i c_{ji} \leq W u_j, \quad j = 1, 2, \dots, N, \\
 & \sum_{j=1}^N c_{ji} \geq q_i, \quad i = 1, 2, \dots, k, \\
 & \sum_{i=1}^k c_{ji} \leq p - 1, \\
 & c_{ji} \text{ is nonnegative integers,} \\
 & u_j \in \{0, 1\}, \quad j = 1, 2, \dots, N.
 \end{aligned} \tag{5.5}$$

In this formulation, c_{ji} s and u_j s are the decision variables, all functions are linear with integer restrictions on the variables, $C_j = (c_{j1}, c_{j2}, \dots, c_{jk})$ forms the j th cutting pattern, and $\sum_{j=1}^N u_j$ is the total number of master rolls used. If $u_j = 0$ for any j , then the first set of constraints in (5.5) forces $c_{ji} = 0$ for $i = 1, 2, \dots, k$. Therefore, the cutting pattern C_j for which $u_j = 0$ is useless and hence unused. It is possible that some of the cutting patterns are replicated in the list of useful cutting patterns. Theoretically, this formulation produces a global optimal solution (to minimize the number of master rolls required to meet the order) but may take long time to produce solutions as it is an integer programming problem with large number of variables. However, if a lower bound constraint is imposed on the objective function, it has been observed that the solutions are obtained very fast. One simple lower bound for the objective function is $\lfloor \frac{\sum w_i q_i}{W} \rfloor$, where W is the deckle width. Using this, a satisfactory solution to (5.5) may be obtained in reasonable time. Before proceeding further on Method 2, the aspects discussed in this paragraph are illustrated with the help of Problem Instance 1 introduced earlier.

Taking $N = 60$, (5.5) is solved for Problem Instance 1. Using a professional solver, a solution could not be reached even after five minutes of solver running time. Using $\lfloor \frac{\sum w_i q_i}{W} \rfloor$, the minimum number of master rolls required for the problem is at least 42. Solving (5.5) with the additional constraint $\sum_{j=1}^N u_j \geq 50$, solution is

obtained in a second. Running the solver after reducing the lower bound from 50 to 45, no solution could be found even after 3 min of solver running. When the solver was aborted, it gave a message that a feasible solution was found with 46 master rolls. Solver was run again with lower bound as 46, and the solution is flashed within two seconds. Out of the 46 cutting patterns, three are replicated twice and two are replicated thrice. Thus, there are 39 distinct cutting patterns in the solution. These are shown in Table 5.3. Trim loss for this solution is 6.67 %. Recall that the solution in Method 1 used 13 cutting patterns, 45 master rolls and the trim loss was 2.78 %.

Table 5.3 Distinct cutting patterns for Problem Instance 1

0000000000004	0000000000140	0000000012010	0000000020100	0000000021000
0000001001030	0000001020000	0000001102001	0000002010000	2000100000000
0000010000012	0000030000000	0000100000012	0000120000000	0000200000020
1002000000001	0001000020000	0001001000002	0004000100000	0010000010001
0010000010010	0011000000020	0011100100000	0012000000010	0012000001000
1011000000000	0020000000001	0020000000010	2001000000000	1001100001000
0103000100000	0110000001010	0200010000010	1000100000001	
0000003000010	0020001000000	0020100000000	0000300000000	0013000000000

Note: In the last row, first 3 are replicated twice and last 2 are replicated thrice.

$w = (127, 82, 131, 78, 110, 120, 101, 59, 139, 65, 100, 66, 85)$

In deckle optimization problem it is very important to keep the number of cutting patterns low. The above solution with 39 distinct cutting patterns will be quite expensive. Even the solution obtained using Method 1 with 13 cutting patterns may be difficult to accept. A solution with 8 cutting patterns, 46 master rolls and with a trim loss of 2.42 % is presented in Fig. 5.8. This is quite appealing compared to the previous two solutions. In fact, there is another solution with 9 cutting patterns, 45 master rolls and a trim loss of 2.52 %. Both these solution are obtained by intelligent formulations. Method 1 as well as formulation (5.5) have no control on the number

Cutting Patterns	Widths (cms)														Trim Loss	Solution u_j
	127	82	131	78	110	120	101	59	139	65	100	66	85			
1	0	0	0	0	0	0	0	0	0	0	1	4	0	16	3	
2	0	0	0	0	0	1	0	0	0	0	0	1	2	24	8	
3	0	0	1	3	0	0	0	0	0	0	0	0	0	15	3	
4	0	0	0	0	0	0	1	0	2	0	0	0	0	1	6	
5	0	0	0	0	0	0	3	0	0	0	0	1	0	11	3	
6	0	0	2	0	1	0	0	0	0	0	0	0	0	8	10	
7	0	1	0	3	0	0	0	1	0	0	0	0	0	5	4	
8	1	0	0	1	1	0	0	0	0	1	0	0	0	0	9	
q_i	8	4	23	29	19	7	14	4	12	9	2	21	15	424	46	
Yeild	9	4	23	30	19	8	15	4	12	9	3	23	16			

Fig. 5.8 A solution to Problem Instance 1 with 8 cutting patterns, 46 master rolls and 2.42 % trim loss

of cutting patterns. The constraint $\sum_{i=1}^k w_i c_{ji} \leq W u_j$ in formulation (5.5) only insists that the corresponding cutting pattern be used ($u_j = 1$) or not used ($u_j = 0$); it does not provide for allowing the cutting pattern to be used more than once directly. However, formulation (5.5) has a tendency to pick up good cutting patterns as it tries to minimize the number of master rolls. If we can somehow use these cutting patterns to produce a solution to the problem by using only a few of them, then we can hope to do better. The trick to do this is built in the following formulation.

$$\begin{aligned}
 & \text{Minimize } \sum_{j=1}^N v_j \\
 & \text{subject to } \sum_{j=1}^N u_j c_{ji} \geq q_i, \quad i = 1, 2, \dots, k, \\
 & \quad u_j \leq N v_j, \quad j = 1, 2, \dots, N, \\
 & \quad \sum_{j=1}^N u_j \leq N_0, \\
 & \quad \sum_{j=1}^N (W - \sum_{i=1}^k w_i c_{ji}) \leq T_0, \\
 & \quad u_j \text{ s are nonnegative integers,} \\
 & \quad v_j \in \{0, 1\}, \quad j = 1, 2, \dots, N.
 \end{aligned} \tag{5.6}$$

In the formulation (5.6): the first constraint determines the number of rolls (u_j) to be cut according to cutting pattern $C_j = (c_{j1}, \dots, c_{jk})$; the variables v_j s are introduced so that $\sum_{j=1}^N v_j$, the number of cutting patterns, can be minimized; the third constraint $\sum_{j=1}^N u_j \leq N_0$ ensures that the number of master rolls does not exceed N_0 , where N_0 is the desirable number of master rolls; the fourth constraint ensures that the trim loss does not exceed the limit T_0 . As before, if $v_j = 0$ for any j , then $u_j = 0$; on the other hand, if $v_j > 0$, then u_j , the number of master rolls to be cut according to C_j , can be any nonnegative integer less than or equal to N (we do not need this to be more than N in any case). It may be noted that all the functions in this formulation are linear with integer restrictions on the variables.

Taking the 39 cutting patterns generated by (5.5) for the Problem Instance 1, $N_0 = 46$ and $T_0 = 1,000$, the problem (5.6) is solved. The solver took less than two seconds to flash the solution. The solution used 8 cutting patterns, 46 master rolls and the trim loss $T_0 = 638$ (3.65 %). The solver was run once again by changing only T_0 to 500. Again the solution was flashed within 2 s and the solution is what is shown in Fig. 5.8. Next, the solver was run with $N_0 = 45$ and $T_0 = 500$. Solution is obtained in 3 s and the solution used 45 master rolls and the trim loss was 432 (2.52 %). When the problem was run with $N_0 = 44$ and $T_0 = 1,000$, solver displayed the message that there was no feasible solution. We shall now present Method 2.

Steps for Method 2

- Step 0. Get the inputs: deckle width W , width vector $w = (w_1, \dots, w_k)$, quantity vector $q = (q_1, \dots, q_k)$, and the number of knives on the winder p .
- Step 1. Compute the lower bound $B_L = \lfloor \frac{\sum w_i q_i}{W} \rfloor$ and set $N = k + \sum_{i=1}^k \lfloor q_i / \lambda_i \rfloor$, where $\lambda_i = \lfloor W / w_i \rfloor$.
- Step 2. Solve the problem (5.5) with the additional constraint $\sum_{j=1}^N u_j \geq B_L$. If the solver does not produce solution in reasonable time, step up the value of B_L and solve it again. Keep trying this until you get a satisfactory solution to the problem.
- Step 3. Drop the useless cutting patterns (those with $c_{ji} = 0$ for all i) in the solution obtained in Step 2, redefine N to be the number of useful cutting patterns, and relabel the useful cutting patterns as C_1, C_2, \dots, C_N . Using appropriate values for N_0 and T_0 , solve problem (5.6) and take the solution as the required solution.

5.4 Planning at Winder and Sheetters

In this section we shall take a closer look at the deckle optimization problem and the process of planning. Planning starts with customer orders. Customer orders are divided into groups so that the orders within each group have the same product grade, GSM and caliper. Production planning of the orders is done separately for each group. Therefore, this project is confined to planning of production of products for a given group of orders. Henceforth, any reference to customer orders will only mean that the orders belong to the same group, that is, they all have the same grade, GSM and caliper. Within a group, there may be both reel and sheet orders, and within reel orders the diameter specifications may differ. One group of orders taken from the company's past data are presented in Table 5.4. We shall use these orders as an example to model the problem and analyze the results. The problem with these data shall be referred to as **sample problem**. In the sample problem, we have six sheet orders and 39 reel orders. Of the reel orders, the first 18 are with a diameter of 100 cm and the remaining 21 are with diameter 95 cm.

5.4.1 Production Restrictions

Planning starts with the requirement of child reels for a given group of orders. Number of jumbos and their diameters are decided based on the requirements of the child reels of the group. Normally, a jumbo will produce two sets; one of the sets will be used to cut child reels of diameter r_1 cm and the other set will be used to cut child reels of diameter r_2 cm, where r_1 and r_2 can be equal or different. Thus, it is possible to choose the jumbo diameter for the sample problem so that both

sets will be used to produce the child reels of same diameter (both 95 cm or both 100 cm), or one set will be used for producing child reels of 95 cm diameter and the other set for producing child reels 100 cm diameter.

Table 5.4 Orders of the sample problem

Order ID	Grade	GSM	Caliper (μ)	Width (cm)	Length (cm)	Dia (cm)	Order type	Grain core dia (in.)	Quantity (kg)
001	ABC	230	294	72.0		100	Reel	3	1510
002	ABC	230	294	76.0		100	Reel	3	1594
003	ABC	230	294	80.0		100	Reel	3	1677
004	ABC	230	294	89.0		100	Reel	3	1866
005	ABC	230	294	67.0		100	Reel	3	7026
006	ABC	230	294	52.0		100	Reel	3	1090
007	ABC	230	294	51.5		100	Reel	3	4320
008	ABC	230	294	52.5		100	Reel	3	4404
009	ABC	230	294	90.0		100	Reel	3	1887
010	ABC	230	294	83.0		100	Reel	3	1740
011	ABC	230	294	73.0		100	Reel	3	1531
012	ABC	230	294	70.0		100	Reel	3	1468
013	ABC	230	294	83.0		100	Reel	3	1740
014	ABC	230	294	66.0		100	Reel	3	2768
015	ABC	230	294	48.0		100	Reel	3	1006
016	ABC	230	294	57.0		100	Reel	3	1195
017	ABC	230	294	66.0		100	Reel	3	5537
018	ABC	230	294	61.0		100	Reel	3	2558
019	ABC	230	294	49.0		95	Reel	3	5516
020	ABC	230	294	91.5		95	Reel	3	5151
021	ABC	230	294	89.0		95	Reel	3	6680
022	ABC	230	294	87.0		95	Reel	3	6530
023	ABC	230	294	84.0		95	Reel	3	7881
024	ABC	230	294	76.0		95	Reel	3	8556
025	ABC	230	294	74.0		95	Reel	3	6943
026	ABC	230	294	71.0		95	Reel	3	6661
027	ABC	230	294	68.0		95	Reel	3	11484
028	ABC	230	294	66.0		95	Reel	3	4954
029	ABC	230	294	61.0		95	Reel	3	10302
030	ABC	230	294	60.0		95	Reel	3	6755
031	ABC	230	294	59.0		95	Reel	3	7750
032	ABC	230	294	58.5		95	Reel	3	4391
033	ABC	230	294	57.0		95	Reel	3	7487
034	ABC	230	294	56.0		95	Reel	3	7355
035	ABC	230	294	55.0		95	Reel	3	6192
036	ABC	230	294	54.0		95	Reel	3	6079
037	ABC	230	294	53.0		95	Reel	3	5967
038	ABC	230	294	51.0		95	Reel	3	5742
039	ABC	230	294	50.0		95	Reel	3	5629
040	ABC	230	294	62.5	86.0	95	Sheet GLW		19938
041	ABC	230	294	43.5	63.5	95	Sheet GLW		20407
042	ABC	230	294	60.0	77.1	95	Sheet GLW		25896
043	ABC	230	294	69.1	101.5	95	Sheet GLW		31120
044	ABC	230	294	76.4	57.3	95	Sheet GSW		28673
045	ABC	230	294	79.4	59.3	95	Sheet GSW		31289

Planning for Reel Orders

Winder planning (of cutting patterns and plans) of reel orders is done separately for each diameter. This is because, it is not possible to cut child reels with different diameters in a single run. However, sheet orders are mixed with reel orders to avoid or reduce scrap (in paper industry scrap is called broke) and excessive production of ordered quantities.

Child Reels for Sheet Orders

Different options are possible for producing child reels for sheet orders. This is illustrated with the help of the sample problem. There are three options for producing the sheet order child reels in this case.

- Option 1. Use jumbo sets only for sheet orders.
- Option 2. Combine sheet order reels with 100 cm diameter child reels or with 95 cm diameter child reels but not both (this is preferred to avoid confusion and better tracking).
- Option 3. Allow sheet order child reels to have any of the two diameters.

Management prefers Option 2 to minimize wastage and to avoid confusion.

The quantity of each order ID is specified in kilograms. For production planning, these quantities have to be converted into number of child reels. As far as the reel orders are concerned, quantities are determined based on the reel diameter, GSM and caliper. However, when it comes sheet order, one needs to weigh different options to optimize.

Consider a sheet order with size w cm \times l cm which is to be cut from a child reel of width w cm. Let c be the caliper (in microns) and g be the GSM of this sheet order. Let the quantity required be Q kg. The volume and weight of one sheet of this order are equal to $wlc \times 10^{-4}$ cm³ and $wlg \times 10^{-7}$ kg respectively. Then, the volume of the order quantity of Q kg in cubic centimeters is given by

$$V = \frac{Q}{wlg \times 10^{-7}} \times wlc \times 10^{-4} = 10^3 \times Q \times \beta, \quad (5.7)$$

where $\beta = \frac{c}{g}$ is the *bulk density* of the sheet.

Let the diameter be D cm, and let the corresponding core diameter be d inches. Then the volume in cubic centimeters of one reel is given by

$$v_1 = \pi(D - 2.54d)^2 w \text{ cm}^3 \quad (5.8)$$

Therefore, the number of reels required to meet the order quantity is equal to $\frac{V}{v_1}$. The formula for $\frac{V}{v_1}$ is given by

$$\frac{V}{v_1} = \frac{1,000Q\beta}{\pi(D - 2.54d)^2w} \quad (5.9)$$

If $\frac{V}{v_1}$ is not an integer, then there is an integer h such that $h \leq \frac{V}{v_1} < h + 1$. Hence, the number of reels to be produced for the sheet order is either h or $h + 1$. Producing $h + 1$ reels will result in excess production. On the other hand, producing h reels will result in shortages. For sheet order ID 045 of the sample problem with required order quantity of 7797 kg, $\frac{V}{v_1} = 20.93$ for reel diameter 95 cm, and $\frac{V}{v_1} = 18.72$ for reel diameter 100 cm. Therefore, of the two diameters, 95 cm reels may be preferred for the order 045.

One of the production constraints is, when sheet orders are clubbed with reel orders in a single run, the child reels produced for sheet orders in that run should have the same diameter as the diameter of the reel orders of that run. It is possible that for the same grade, there could be reel orders with different diameters as in the case of sample problem. In that case, we can mix the sheet orders with either one or more of such diameters as mentioned in the options above. One has to explore through OR modeling which way of mixing will yield better solution (see Problem 5.6).

Role of Sheeters in Planning

Sheeters play a significant role on the cutting stock problem at the winder. To understand this, consider the following example. Suppose we want to produce the sheets for order ID 045 of the sample problem using the child reels with diameter 95 cm. Because these sheets are direction oriented (GSW, see Table 5.4), the width of the child reels for this order must be 79.4 cm plus the cutting allowance. The cutting allowance depends upon the type of sheeter used and the planning (see Sect. 5.2.4). For simplicity, we shall assume that only company sheeters are used. Then the cutting allowance will be 2 cm (see Fig. 5.4) but there is an option of choosing child reel widths from one or more of 81.4 cm ($= 79.4 + 2$), 160.8 cm ($= 2 \times 79.4 + 2$) and 240.2 cm ($= 3 \times 79.4 + 2$). To minimize waste, we should use a combination of these three options. Recall that the number of 95 cm diameter child reels required for order 045 is 21. Cut x_1 child reels with width 81.4 cm, x_2 child reels with width 160.8 cm and x_3 child reels with width 240.2 cm so that $x_1 + 2x_2 + 3x_3$ is at least 21 and should be as close to 21 as possible. An optimal solution to this is $x_1 = x_2 = 0$ and $x_3 = 7$ and this will waste a cutting allowance of 14 cm. One can meet the order 045 by using 21 child reels of width 81.4 cm, but in that case one would be wasting a cutting allowance of 42 cm. Moreover, productivity is very slow in this and this will also result in high underutilization of sheeter capacity. When we try to minimize the cutting allowance, this will automatically force the optimal solutions to pick options which will utilize sheeter in an optimal way.

5.4.2 Pooling Orders to Suit Reel Diameters

Normally company receives a number of orders from the customers. Not all orders are picked for production. Priority will be given to those sheet orders whose quantities closely match with the quantities of reel orders. For example, if a sheet order requires a quantity which will be equal to quantity of 2.5 child reels of a reel order, then this will result in excessive waste. In such case, customers are negotiated to increase the order quantity to that of three reels quantity or decrease to quantity of two reels quantity. Also, there is an understanding with the customers that the actual quantity supplied may be at variance by 10 % of the ordered quantity. It is customary to plan the production using quantities (weight in kg) rather than the numbers (of child reels).

Summarizing this section, the deckle optimization problem should plan the cutting patterns and cutting plans at winder. Planning should be done in such a way that all the orders are met. A variation of 10 % of order quantity is allowed for each order in planning. Priority is given to minimizing trim losses and excess production. The number of cutting patterns should be controlled so that the time of knife setting is minimized.

5.5 Formulations of the Deckle Optimization Problem

The deckle optimization problem is solved in two stages. In the first stage, the cutting stock problem of the winder is solved. Once a solution to the cutting stock problem is obtained, the sequencing of cutting patterns is determined using the traveling salesman problem formulation to minimize the number of knives setting changes. When the number of diameters is more than one among the reel orders, there are different options for considering the deckle optimization problem. We shall illustrate this with the help of the sample problem. In the process we see different formulations for the deckle optimization problem.

5.5.1 Different Options

In the sample problem, reel orders have two different diameters – 95 and 100 cm. We shall consider four different ways of planning the orders of the sample problem with data in Table 5.4. These are listed below.

1. Mixing sheet orders with reel orders of 95 cm diameter.
2. Mixing sheet orders with reel orders of 100 cm diameter.
3. Mixing sheet orders with reel orders of both diameters.
4. Separate planning for sheet orders.

Before proceeding further, we need the reel weights and the number of reels for each order of the sample problem. These are computed from the data in Table 5.4 and Eqs. (5.1) and (5.9) and presented in Table 5.5.

Table 5.5 Number of reels and reel weights for orders

Orders with 100 cm diameter				Orders with 95 cm diameter				Sheet orders			
ID	Width	NR	RW	ID	Width	NR	RW				
001	72.0	1	1510.15	019	49.0	6	919.50	When clubbed with 100 cm diameter			
002	76.0	1	1594.04	020	91.5	3	1717.02	Order ID	Width	NR	RM
003	80.0	1	1677.94	021	89.0	4	1670.11	040	62.5	15	1310.89
004	89.0	1	1866.71	022	87.0	4	1632.58	041	43.5	22	912.38
005	67.0	5	1405.27	023	84.0	5	1576.28	042	60.0	20	1258.45
006	52.0	1	1090.66	024	76.0	6	1426.16	043	69.1	21	1449.32
007	51.5	4	1080.17	025	74.0	5	1388.63	044	76.4	18	1602.43
008	52.5	4	1101.15	026	71.0	5	1332.33	045	79.4	19	1665.35
009	90.0	1	1887.68	027	68.0	9	1276.04	When clubbed with 95 cm diameter			
010	83.0	1	1740.86	028	66.0	4	1238.51	Order ID	Width	NR	RM
011	73.0	1	1531.12	029	61.0	9	1144.68	040	62.5	17	1172.83
012	70.0	1	1468.20	030	60.0	6	1125.92	041	43.5	25	816.29
013	83.0	1	1740.86	031	59.0	7	1107.15	042	60.0	23	1125.92
014	66.0	2	1384.30	032	58.5	4	1097.77	043	69.1	24	1296.68
015	48.0	1	1006.76	033	57.0	7	1069.62	044	76.4	20	1433.67
016	57.0	1	1195.53	034	56.0	7	1050.85	045	79.4	21	1489.96
017	66.0	4	1384.30	035	55.0	6	1032.09				
018	61.0	2	1279.43	036	54.0	6	1013.32				
				037	53.0	6	994.56				
				038	51.0	6	957.03				
				039	50.0	6	938.26				

NR Number of reels; RW Reel weight in kilograms

5.5.2 Model for Sheet Orders with 95 cm Diameter

Under this model, child reels for sheet orders and reel orders of 95 cm diameter are produced from master rolls of 95 cm diameter. From Table 5.4, we see that there are 21 reel orders with diameter 95 cm and six sheet orders. Thus, we have a total of 27 widths and the width vector w has 27 coordinates. The i th coordinate of w , w_i , is the width of order ID $18 + i$, $i = 1, 2, \dots, 27$. For example, w_1 is the width of the reel of order ID 019, w_2 is width of reel order ID 020, w_{22} is width of sheet order ID 040, and so on. There may be a confusion with widths of the sheet orders. Take order ID 044 or 045. In Table 5.5, w_{26} is taken as 76.4 cm and that w_{27} is taken as 79.4 cm. From the dimensions of sheets of these two orders, 76.4 and 79.4 should be their

lengths (see Table 5.5). But as these two orders are of GSW type, the lengths will go as widths of the child reels. Let q_i denote the number of reels required for width w_i , $i = 1, 2, \dots, 27$, and let $q = (q_1, \dots, q_{27})$.

For simplicity, we shall formulate the problem under the assumption that only company sheeters are used. The general case is a simple extension of this. Note that w_i and q_i for $i = 22, 23, \dots, 27$, correspond to sheet orders. For each of these orders, there are three ways of cutting the child reels (see the discussion on role of sheeters on page 111) using widths $w_i + 2$, $w'_i = 2w_i + 2$ and $w''_i = 3w_i + 2$. For $i = 22, 23, \dots, 27$, let $\bar{w}_i = w_i + 2$, $\bar{w}'_i = 2w_i + 2$ and $\bar{w}''_i = 3w_i + 2$. If the decision is to cut c_i reels of width \bar{w}_i , c'_i reels of width \bar{w}'_i and c''_i reels of width \bar{w}''_i , then the total number of child reels cut for this order is equal to $c_i + 2c'_i + 3c''_i$. Thus, taking sheeter facilities into account, we have introduced three widths $\bar{w}_i = w_i + 2$, $\bar{w}'_i = 2w_i + 2$ and $\bar{w}''_i = 3w_i + 2$ for each sheet order $i = 22, 23, \dots, 27$.

We are now ready for formulating the problem. We shall use Method 2 for this formulation. In this approach, we first generate the cutting patterns using formulation (5.5) and then, choose the best solution from among them. For the sample problem, taking the sheeter facilities into account, there are $k = 39 (= 21 + 6 \times 3)$ widths in each cutting pattern. The first 21 of these, w_1, w_2, \dots, w_{21} , correspond to the reel orders with IDs $18 + i$, $i = 1, 2, \dots, 21$, and the remaining correspond to sheet orders with the widths $\bar{w}_i = w_i + 2$, $\bar{w}'_i = 2w_i + 2$ and $\bar{w}''_i = 3w_i + 2$ meant for sheet order ID $18 + i$, $i = 22, 23, \dots, 27$. Therefore, a cutting pattern here has to specify a nonnegative integer for each of the 39 widths. Let the cutting pattern be defined by c_i for width w_i for $i = 1, 2, \dots, 21$, and c_i, c'_i, c''_i for widths $\bar{w}_i, \bar{w}'_i, \bar{w}''_i$ respectively, $i = 22, 23, \dots, 27$. Since there will be more than one cutting pattern, denote the j th cutting pattern by C_j . The width vector and the j th cutting pattern vectors are given below.

$$w = (w_1, w_2, \dots, w_{21}, \bar{w}_{22}, \bar{w}'_{22}, \bar{w}''_{22}, \bar{w}_{23}, \bar{w}'_{23}, \bar{w}''_{23}, \dots, \bar{w}_{27}, \bar{w}'_{27}, \bar{w}''_{27})$$

$$C_j = (c_{j1}, c_{j2}, \dots, c_{j21}, c'_{j22}, c''_{j22}, c_{j23}, c'_{j23}, c''_{j23}, \dots, c_{j27}, c'_{j27}, c''_{j27})$$

For the sample problem, the number of knives $p = 10$ and deckle width $W = 380$. Formulation (5.5) for the sample problem is given in (5.10).

Using the formulae of Step 1 of Method 2, we need a minimum of 43 master rolls and a maximum of 61 rolls. Solving the problem with $B_L = 43$ and $N = 61$, a feasible solution was found after 102 s of solver running time with an objective value of 45 (master rolls). Beyond this there was no reduction in the objective function even after 5 min of solver running time. The solver was aborted after 5 min. The solver was run once more by increasing B_L to 45. In this case, a feasible solution was found in the ninth second and the optimal solution was found in 36 s. The solution is summarized in Table 5.6. We now proceed to Step 3 of Method 2.

In Step 3 of Method 2, problem (5.6) is solved using cutting patterns generated by problem (5.5). In this formulation, one constraint puts an upper limit on the number

of master rolls to be used and one constraint puts an upper limit on the trim loss. To start with, we drop both these constraints and solve the problem to find the minimum number of cutting patterns. The problem after dropping the two constraints is given

$$\begin{aligned}
 & \text{Minimize } \sum_{j=1}^N u_j \\
 & \text{subject to} \\
 & \sum_{i=1}^{21} w_i c_{ji} + \sum_{i=22}^{27} (\bar{w}_i c_{ji} + \bar{w}'_i c'_{ji} + \bar{w}''_i c''_{ji}) \\
 & \leq W u_j, \quad j = 1, 2, \dots, N, \\
 & \sum_{j=1}^N c_{ji} \geq q_i, \quad i = 1, 2, \dots, 21, \\
 & \sum_{i=1}^N (c_{ji} + 2c'_{ji} + 3c''_{ji}) \geq q_i, \quad i = 22, \dots, 27, \\
 & \sum_{i=1}^{21} c_{ji} + \sum_{i=22}^{27} (c_{ji} + c'_{ji} + c''_{ji}) \leq p - 1, \\
 & \sum_{j=1}^N u_j \geq B_L \\
 & c_{ji}s, c'_{ji}s, c''_{ji}s \text{ are nonnegative integers,} \\
 & u_j \in \{0, 1\}, \quad j = 1, 2, \dots, N.
 \end{aligned} \tag{5.10}$$

in (5.11). The cutting patterns C_j used in this formulation are the ones generated from solution of (5.10). The decision variables are v_j s and x_j s, where x_j is the number of master rolls cut according to C_j .

Solving the above problem, the solution has 11 cutting patterns but uses 70 master rolls. Since it is possible to meet the required orders with 45 master rolls, problem (5.11) was solved with the additional constraint that number of master rolls should be less than or equal to 45 ($\sum_{j=1}^N x_j \leq 45$). It took 81 s to solve this problem and the resulting solution used 32 cutting patterns and had a trim loss of 3.7 %. To reduce the number of cutting patterns, problem (5.11) was solved by increasing the upper limit on $\sum_{j=1}^N x_j$ from 45 to 50. The resulting solution had 16 cutting pattern and had a trim loss of 5.7 %.

$$\begin{aligned}
& \text{Minimize } \sum_{j=1}^{45} v_j \\
& \text{subject to } \sum_{j=1}^{45} x_j c_{ji} \geq q_i, \quad i = 1, 2, \dots, 21, \\
& \sum_{j=1}^{45} x_j (c_{ji} + 2c'_{ji} + 3c''_{ji}) \geq q_i, \quad i = 22, \dots, 27, \\
& x_j \leq 45v_j, \quad j = 1, 2, \dots, 45, \\
& x_j \text{ s are nonnegative integers,} \\
& v_j \in \{0, 1\}, \quad j = 1, 2, \dots, 45.
\end{aligned} \tag{5.11}$$

Another way to solve the problem is to minimize the number of master rolls with a bound on the number of cutting patterns. The formulation for this problem with *MCP* as the maximum number of cutting patterns is given by

$$\begin{aligned}
& \text{Minimize } \sum_{j=1}^{45} x_j \\
& \text{subject to } \sum_{j=1}^{45} x_j c_{ji} \geq q_i, \quad i = 1, 2, \dots, 21, \\
& \sum_{j=1}^{45} x_j (c_{ji} + 2c'_{ji} + 3c''_{ji}) \geq q_i, \quad i = 22, \dots, 27, \\
& x_j \leq 45v_j, \quad j = 1, 2, \dots, 45, \\
& \sum_{j=1}^{45} v_j \leq MCP \\
& x_j \text{ s are nonnegative integers,} \\
& v_j \in \{0, 1\}, \quad j = 1, 2, \dots, 45.
\end{aligned} \tag{5.12}$$

A solution to the above problem with $MCP = 13$ uses 53 master rolls. As the number of cutting patterns is reduced, there will be an increase in the number of master rolls used. One can prepare a table of solutions with different alternatives. Using such a table, the manager can decide which one will be more preferable. Table 5.7 is obtained by solving problem (5.12) for different values of *MCP*. In all the solutions listed in Table 5.7 the number of cutting patterns turned out to be equal to *MCP*.

5.5.3 Model for Sheet Orders with 100 cm Diameter

In this case, sheet orders are mixed with reel orders of 100 cm diameter. Problem is solved in the same way as it was done for the previous case with reel orders of 95 cm diameter. While minimizing the number of master rolls, a lower bound of 28

Table 5.7 Summary of results of different solutions to problem (5.12)

<i>MCP</i>	Master rolls			Trim loss			Solver time
	Number	Weight	Excess (%)	Width	Weight	Percentage	
11	61	434369	44	713	13378	4.44	1
12	56	398765	32	1124	21096	7.00	2
13	53	377402	25	952	17870	5.93	1
14	52	370282	23	994	18656	6.19	3
15	51	363161	21	865	16234	5.39	7
16	49	348919	16	830	15579	5.17	3
17	49	348919	16	830	15579	5.17	5
18	49	348919	16	830	15579	5.17	19
19	48	341798	13	931	17472	5.80	16
20	48	341798	13	893	16754	5.56	47
21	48	341798	13	674	12653	4.20	93
22	47	334678	11	687	12888	4.28	28
23	47	334678	11	729	13685	4.54	37
23	47	334678	11	729	13685	4.54	37
28	46	327557	9	747	14014	4.65	69
32	45	320436	6	593	11141	3.70	168

Note: *MCP* is the maximum number of cutting patterns allowed in (5.12) and all the solutions attained this limit. Units of width is cm, weight is kg, and time is seconds.
Total weight of all child reels required, 301338 kg, is taken as the basis for computation of trim loss and excess percentages

master rolls was observed but the best solution observed was 30 master rolls. The solver was aborted when it could not reduce the objective function below 30 even after 5 min of running time. When the problem was solved to minimize the number of cutting patterns with 29 as the upper bound on the number of master rolls, the solver displayed the message that the problem was infeasible. Different solutions for mixing sheet orders with reel orders of 100 cm diameter are summarized in Table 5.8.

Table 5.8 Results of mixing sheet orders with reel orders of 100 cm diameter

<i>MCP</i>	Master rolls			Trim loss			Solver time
	Number	Weight	Excess (%)	Width	Weight	Percentage	
12	36	286928	30	842	17656	8.00	1
14	34	270987	23	793	16643	7.54	2
16	32	255047	16	612	12830	5.81	4
17	31	247077	12	555	11645	5.27	2
18	30	239106	8	474	9946	4.51	1

Note: *MCP* is the maximum number of cutting patterns allowed in (5.12) and all the solutions attained this limit. Units of width is cm, weight is kg, and time is seconds.
Total weight of all child reels required, 220772 kg, is taken as the basis for computation of trim loss and excess percentages

5.5.4 Model for Sheet Orders Mixed with Both Diameters

In this case, we distribute the sheet orders among the reel orders of both diameters. To model this, we introduce two sets of cutting patterns. Model is presented for the sample problem. General case can be built along similar lines.

For the 95 cm diameter reel orders, there are 21 widths, and for 100 cm diameter reel orders there are 18 widths. For sheet orders there are six widths. In the new formulation for the problem of mixing sheet orders with reel orders of both diameters, we shall use the index i as follows: $i = 1, 2, \dots, 18$ will stand for reel orders of 100 cm diameter with IDs from 001 to 018; $i = 19, 20, \dots, 39$ will stand for reel orders of 95 cm diameter with IDs from 019 to 039, and $i = 40, 41, \dots, 45$ will stand for sheet orders with IDs 040–045. Keeping the sheeter facilities in mind, for each sheet order, we define the sheet order widths $\bar{w}_i, \bar{w}'_i, \bar{w}''_i$ as before but noting that the index i now will run from 40 to 45.

Let $c_{ji}, i = 19, 20, \dots, 39$ and $c'_{ji}, c''_{ji}, i = 40, \dots, 45$ be the variables defined for the reel and sheet orders with 95 cm diameter reels like in problem (5.10). Similarly, define variables $b_{ji}, i = 1, 2, \dots, 18$ and $b_{ji}, b'_{ji}, b''_{ji}, i = 40, \dots, 45$ for the reel and sheet orders with 100 cm diameter. Note that the variables $b_{ji}, b'_{ji}, b''_{ji}$ for $i = 40, \dots, 45$ correspond to sheet orders. Consider the first sheet order corresponding to $i = 40$ (with ID 040 in Table 5.4). Then, the number of reels produced for sheet order i will be equivalent to $\sum_j (c_{ji} + 2c'_{ji} + 3c''_{ji})$ reels of width w_i with diameter 95 cm, plus $\sum_j (b_{ji} + 2b'_{ji} + 3b''_{ji})$ reels of order i with width w_i and 100 cm diameter.

There is one difficulty in mixing the sheet order with both diameters. For any sheet order, the number of sheets that can be obtained from a child reel is not same for 95 cm diameter rolls and 100 cm diameter rolls. To overcome this problem, we use the weight of the child reels as the requirement. Recall that the orders are specified by the customers in terms of weight which we converted to number of child reels for a specified diameter. From Table 5.5, for order ID 040, weight of one child reel of diameter 95 cm is 1172.83 kg and weight of one child reel of diameter 100 cm is 1310.89 kg. As per the decision variables, the weight of the sheet order $i = 40$ (that is, order ID 040) produced is equal to $1,172.83 \sum_j (c_{ji} + 2c'_{ji} + 3c''_{ji}) + 1,310.89 \sum_j (b_{ji} + 2b'_{ji} + 3b''_{ji})$, and this should be at least the required quantity of 19938 kg (see Table 5.4). In this formulation, $Q_i, i = 40, 41, \dots, 45$, will stand for the order quantity of sheet order i .

There will be two sets of cutting patterns in the new formulation, each set having a maximum of N cutting patterns. One of these sets is meant for reels of diameter 100 cm and these cutting patterns use the variables $b_{ji}, i = 1, \dots, 18, b_{ji}, b'_{ji}, b''_{ji}, i = 40, \dots, 45$ with j running from 1 to N . The other set of cutting patterns is meant for reels with diameter 95 cm, and these cutting patterns use the variables $c_{ji}, i = 19, \dots, 39, c'_{ji}, c''_{ji}, i = 40, \dots, 45$ with j running from $N + 1$ to $2N$, where N is large enough to allow feasibility. The complete formulation is shown in (5.13). Based on solutions to the previous problems, the lower bound B_L can be taken as 75 ($= 45 + 30$) to start with.

$$\begin{aligned}
& \text{Minimize } \sum_{j=1}^{2N} u_j \\
& \text{subject to} \\
& \sum_{i=1}^{18} w_i b_{ji} + \sum_{i=40}^{45} (\bar{w}_i b_{ji} + \bar{w}'_i b'_{ji} + \bar{w}''_i b''_{ji}) \\
& \quad \leq W u_j, \quad j = 1, 2, \dots, N, \\
& \sum_{i=19}^{39} w_i c_{ji} + \sum_{i=40}^{45} (\bar{w}_i c_{ji} + \bar{w}'_i c'_{ji} + \bar{w}''_i c''_{ji}) \\
& \quad \leq W u_j, \quad j = N+1, 2, \dots, 2N, \\
& \sum_{j=1}^N b_{ji} \geq q_i, \quad i = 1, 2, \dots, 18, \\
& \sum_{j=N+1}^{2N} c_{ji} \geq q_i, \quad i = 19, 20, \dots, 39, \\
& 1,310.89 \sum_{j=1}^N (b_{ji} + 2b'_{ji} + 3b''_{ji}) \\
& \quad + 1,172.83 \sum_{j=N+1}^{2N} (c_{ji} + 2c'_{ji} + 3c''_{ji}) \\
& \quad \geq Q_i, \quad i = 40, 41, \dots, 45, \\
& \sum_{i=1}^{18} b_{ji} + \sum_{i=40}^{46} (b_{ji} + b'_{ji} + b''_{ji}) \leq p - 1, \\
& \sum_{i=19}^{39} c_{ji} + \sum_{i=40}^{46} (c_{ji} + c'_{ji} + c''_{ji}) \leq p - 1, \\
& \sum_{j=1}^{2N} u_j \geq B_L \\
& \quad c_{ji}s, c'_{ji}s, c''_{ji}s \text{ are nonnegative integers,} \\
& \quad b_{ji}s, b'_{ji}s, b''_{ji}s \text{ are nonnegative integers,} \\
& \quad u_j \in \{0, 1\}, \quad j = 1, 2, \dots, N.
\end{aligned} \tag{5.13}$$

Using the variables c s and b s defined above, we can formulate the problem of producing required child reels as in problem (5.10) to minimize the master rolls as stated in the formulation (5.13). However, as the master rolls are of two different diameters, the right objective function would be to minimize the overall weight of the master rolls. This function is $M_{95} \sum_{j=1}^N u_j + M_{100} \sum_{j=N+1}^{2N} u_j$, where M_D is the weight of one master roll of diameter D cm, $D = 95, 100$. For the sample problem, $M_{95} = 7,130.8$ kg and $M_{100} = 7,970.29$ kg. The results for problem (5.13) with different objective functions are as follows.

- When the objective function is taken as the number of master rolls, the solution used 63 master rolls (22 of diameter 95 cm and 41 of diameter 100 cm) with a total weight of 450,859 kg.
- When the objective function is changed to minimizing the number of cutting patterns, the solution used 393 master rolls (168 of diameter 95 cm and 225 of diameter 100 cm) with the number of cutting patterns 24 (obviously, in the process of minimizing the number of cutting patterns, the material consumption is increasing).
- When the problem is solved to minimize the number of cutting patterns by imposing an upper bound of 63 on the number of master rolls, the solution used 30 cutting patterns with 63 master rolls (25 of diameter 95 cm and 38 of diameter 100 cm).
- When the problem is solved with weight of the master rolls ($M_{95} \sum_{j=1}^N u_j + M_{100} \sum_{j=N+1}^{2N} u_j$) as the objective function, the solution used 58 master rolls (23 of diameter 95 cm and 35 of diameter 100 cm).

Among the four solutions mentioned above, the last one with weight as the objective function was the most efficient one with an efficiency of 83 % and it took 80 seconds of solver time to produce this solution.

Company's management was curious to compare different alternatives such as clubbing sheet orders with one of the diameter rolls only or using separate cutting patterns exclusively to sheet orders. These alternatives can be explored using the formulation (5.13) by adding additional constraints. But before doing this, we must first look into the problem of producing sheet orders separately.

5.5.5 Model for Sheet Orders Separately

Reel orders are placed with specific diameters. But this restriction is not there for sheet orders. Therefore, if sheet orders are produced separately, then we must first of all decide on the diameter. This problem can be formulated mathematically but it leads us to a nonlinear programming problem. The problem is formulated below. For simplicity, we shall ignore the sheeter options in this formulation. Problem with sheeter options can be formulated in a similar fashion using \bar{w}_i , \bar{w}'_i , \bar{w}''_i defined earlier (see Problem 5.7).

For ease of presentation, we use the sample problem for this purpose. In this problem, there are six widths for the sheet orders. Let w_i , $i = 40, 41, \dots, 45$, be the widths required for the six sheet orders with i th order requiring Q_i kg. The decision variables in this problem are the diameter, number of reels of each width and the cutting patterns. Diameter determines the number of reels of each width and the number of reels in turn determines the number of cutting patterns. From the view point of operational convenience, management prefers one single diameter for all the reels and that the diameter should be within the range of 80–110 cm. Let D be the diameter of the master rolls and let x_i be the number of reels of width w_i

that need to be produced. Let $C_j = (c_{j40}, c_{j41}, \dots, c_{j45})$ be the j th cutting pattern, $j = 1, 2, \dots, N$. The formulation is explained in the following steps.

- Recall that the weight of a single reel of diameter D is $\pi(D - 2.54d)^2 gw_i / 1000c$, where w_i is the width, g is GSM, d is the core diameter, and c is the caliper (see Eq. (5.1)). As the weight required is Q_i and the number of reels is x_i for sheet orders of width w_i , we must have $\frac{\pi(D - 2.54d)^2 gw_i x_i}{1000c} \geq Q_i$ or $(D - 2.54d)^2 x_i \geq \frac{1000Q_i c}{\pi g w_i}$.
- For C_j to be a cutting pattern, we must have $\sum_{i=40}^{45} c_{ji} w_i \leq W$, $j = 1, 2, \dots, N$, where W is the deckle width. To decide whether to use the j th cutting pattern or not, we introduce the binary variable u_j and the cutting pattern constraint becomes $\sum_{i=40}^{45} c_{ji} w_i \leq W u_j$, $j = 1, 2, \dots, N$.
- The number of reels of width w_i produced must be at least x_i . Hence, we must have $\sum_{j=1}^N c_{ji} \geq x_i$, $i = 40, 41, \dots, 45$.
- The diameter must be between 80 and 110 cm, hence we must have $80 \leq D \leq 110$.
- Since the diameter D is same for all master rolls, we can take $\sum_{j=1}^N u_j$, the number of master rolls, as the objective function to minimize the total quantity required to meet the orders.

The problem is formulated in (5.14). It may be noted that this turns out to be a problem with one of the constraints being nonlinear.

$$\begin{aligned}
 & \text{Minimize } \sum_{j=1}^N u_j \\
 & \text{subject to} \\
 & (D - 2.54d)^2 x_i \geq \frac{1000q_i w_i c}{\pi g}, \quad i = 40, 41, \dots, 45, \\
 & \sum_{i=1}^6 c_{ji} w_i \leq W u_j, \quad j = 1, 2, \dots, N, \\
 & \sum_{j=1}^N c_{ji} \geq x_i, \quad i = 40, 41, \dots, 45, \\
 & 80 \leq D \leq 110, \\
 & x_i \text{s and } c_{ji} \text{s are nonnegative integers,} \\
 & D \text{ is real and } u_j \in \{0, 1\}, \quad j = 1, 2, \dots, N.
 \end{aligned} \tag{5.14}$$

For the sample problem the input data are the widths (w) and the corresponding quantities (Q) required. These are given below. Deckle width is 380 cm.

$$w = (62.5, 43.5, 60.0, 69.1, 76.4, 79.4) \text{ and}$$

$$Q = (19938, 20407, 25896, 31120, 28673, 31289).$$

With these data, problem (5.14) was solved (with diameter in the limits 80–110). Solver had to be interrupted after 11 minutes as it was still searching for global Optimal solution. The best solution at the time of interruption, used 18 master rolls with 110 cm diameter and with a total weight of 176204 kg, the x -vector is (13, 19, 17, 18, 15, 16). This solution was found in 2 min and 14 s, and there was no improvement thereafter. To see the effect of diameter, the problem is solved by varying the upper limit on the diameter. Results are summarized in Table 5.9. From the results, the optimal diameter (to minimize the total material) is around 100 cm.

Table 5.9 Effect of diameter on planning only sheet orders

Diameter limits	Optimal diameter	Master rolls			Trim loss		Solver time
		Number	Total weight	Efficiency %	Weight	%	
80–95	95	24	171139	92	6480	3.94	123 ^a
80–100	100	21	167374	94	6617	4.12	645
80–105	105	19	168270	93	6081	3.75	80
80–110	110	18	176204	89	13416	8.24	134 ^b

Note: Weight of the child reels ordered is 157323; weight units are in kg and time in seconds

^aSolver interrupted at 123 s ^bSolver interrupted at 134 s

5.5.6 Comparison of Different Alternatives

As mentioned earlier, the management wanted to explore different alternatives. We can do this using the formulations (5.13) and (5.14). Since 100 cm is also the optimal diameter for the sheet orders separately, we can use only (5.13) for comparing different alternatives. The four options are listed at the beginning of Sect. 5.5. In order to compare the alternatives we must first select a basis. In this case, it makes sense to use the total weight of all master rolls as the basis for comparison.

In the formulation (5.13), we can add or delete constraints to get solutions for each of the alternatives. The procedure to do this is explained for each of the alternatives in the following steps.

1. **Combining sheet orders with reel orders of 95 cm diameter.** To find the total weight of minimum number of master rolls under this alternative, we first solve (5.13) by imposing an additional constraint that $\sum_j (b_{ji} + 2b'_{ji} + 3b''_{ji}) = 0$ for $i = 40, 41, \dots, 45$. This will ensure that all b -variables corresponding to sheet orders will be zero and hence rolls of diameter 100 cm are not used for sheet orders and only the rolls of 95 cm diameter are used.
2. **Combining sheet orders with reel orders of 100 cm diameter.** For this case, we add the constraint $\sum_j (c_{ji} + 2c'_{ji} + 3c''_{ji}) = 0$ for $i = 40, 41, \dots, 45$ to the problem (5.13).
3. **Combining Sheet Orders with both Reel Orders.** Solve problem (5.13) as it is.
4. **Plan Sheet orders separately.** By dropping all b -variables and setting $\sum_j (c_{ji} + 2c'_{ji} + 3c''_{ji}) = 0$ for $i = 40, 41, \dots, 45$ in problem (5.13), we can find the

minimum number of rolls to meet reel orders of 95 cm diameter; by dropping all c -variables and setting $\sum_j (b_{ji} + 2b'_{ji} + 3b''_{ji}) = 0$ for $i = 40, 41, \dots, 45$ in problem (5.13), we can find the minimum number of rolls to meet reel orders of 100 cm diameter; and by dropping all c -variables from problem (5.13) and setting $\sum_j b_{ji} = 0$ for $i = 1, 2, \dots, 39$ in problem (5.13), we can find the minimum number of rolls to meet sheet orders.

Under Alternative 4, the problem was solved with two different objective functions – number of master rolls (OBJ 1) and weight of master rolls (OBJ 2). The solution with respect to OBJ 1 was better compared to the one with OBJ 2. The solutions to the four alternatives are summarized in Table 5.10. All the solutions listed in the table are near optimal solutions and the solver was interrupted in all the cases as the solver was still searching for global optimal solutions. Theoretically, solution to alternative 4 with objective function OBJ 1 should be better than the solution to alternative 1 as the feasible region is bigger for the former. But the results in the table show otherwise. This is because the solutions are near optimal. A slightly modified solution obtained under alternative 2 is presented in Fig. 5.9. This modified solution reduces the number of cutting patterns from 49 to 36 by consuming two additional master rolls.

Table 5.10 Effect of diameter on planning only sheet orders

Alternative	Objective function	Master rolls				Efficiency percentage
		Number			Total weight	
		95 cm	100 cm	Total		
1	OBJ 1	45	6	51	368707	94
2	OBJ 1	21	27	48	364943	95
3	OBJ 1	22	27	49	372073	93
4	OBJ 1	21	28	49	372913	93
4	OBJ 2	30	22	52	389269	89

Note: Weight units are in kg; Weight of the child reels ordered is 346245;
Efficiency percentage = $100 \times 346245 / (\text{total weight of master rolls})$

5.6 Sequencing Cutting Patterns in Deckle Optimization

Since the paper machine producing papers and paperboards runs continuously, the inflow of jumbo rolls into winder section will be continuous and any downtime at winder will result in clogging of jumbo rolls which will very badly affect the production. For this reason, it is very important to minimize the downtime at winder. Each time a cutting pattern is changed, the winder has to be stopped, the knives have to be reset and the winder has to be restarted. Of these activities, stopping and restarting times take major portion and compared to this time, the knives resetting time is relatively small. Therefore, to reduce the winder downtime, one must focus

Reel Orders with 95 cms Diameter																			
49.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
91.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
89.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
87.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
84.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
76.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
74.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
71.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
68.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
66.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
55.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
53.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Reel Orders with 100 cms Diameter																			
72.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
76.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
89.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
67.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
83.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
73.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
70.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
83.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
66.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
66.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sheet Orders (clubbed with 100 cms diameter rolls)	64.5	127.0	189.5	45.5	89.0	132.5	69.0	122.0	182.0	71.1	140.2	209.3	78.4	154.8	231.2	81.4	160.8	240.2	Frequencies

Note: This solution clubs sheet orders with reel orders of 100 cms diameter. It uses 50 master rolls (21 for 95 cms diameter reel orders, and 29 for sheet orders and reel orders of 100 cms diameter). The number of cutting patterns used by this solution is 36 (21 for reel orders of 95 cms diameter and 15 for sheet orders and reel orders of 100 cms diameter).

Fig. 5.9 A slightly modified solution obtained from the solution obtained under alternative 2

more on reducing the number of cutting patterns and then try to minimize the knives resetting time. Reducing the number of cutting patterns has been addressed in the previous section. This section will consider the problem of minimizing the knives resetting time.

5.6.1 Adjustment of Knives on the Winder

Before we consider the problem of minimizing the knives setting time, we must first understand the operations of changing the knives settings. In the type of winder considered in this project, knives, say 10 in number, are loaded on a horizontal shaft. The knives are numbered from 1 to 10 from left to right. The ordering of knives cannot be changed. For example you cannot bring knife numbered 5 between the knives numbered 3 and 4 or for that matter between 1 and 2. All that can be done is: (i) disable one or more knives and (ii) increase or decrease the distance between any two adjacent knives. The two end-knives are always used to trim the unevenness of the roll edges.

Let us try to understand these operations with the help of an example. Consider the cello tape problem discussed in Sect. 5.3. Imagine a baby winder for cutting the 12-in. cello tapes into cello tapes of smaller widths. For brevity, we shall assume that this baby winder has seven knives on it. Knives 1 and 7 are always fixed at the extremes and the distance between these two knives is 12 in. The cutting pattern C_2 of Table 5.1 cuts two pieces of width 3 in. each and one piece of width 4 in. Figure 5.10 shows two different ways of placing knives to cut the same cutting pattern C_2 . In this figure, knives 5 and 6 (counting from left) are disabled.

Supposing that the production is planned in such a way that after cutting the tapes according to C_2 , the next cutting pattern to be used is C_1 which cuts 4 pieces of 3 in. each (see Table 5.1). If for C_2 , the arrangement is as in part (a) of Fig. 5.10, then knife number 4 is to be shifted to its left by 1 in. This requires releasing knife 4, moving it to its left by 1 in. and fixing it there. On the other hand, if the arrangement for C_2 is as in part (b) of Fig. 5.10, then it is necessary to change two knives' settings. Therefore, the number of knife changes from one cutting pattern to another depends on the way in which a cutting pattern is arranged (see Problem 5.5). We shall refer to the ordering of widths in a cutting pattern as the **design of that cutting pattern**. Thus, Fig. 5.10 shows two different designs of the same cutting pattern C_2 .

Consider another situation where it is required to change the cutting pattern from C_2 to C_6 (C_6 cuts one piece of 6 in. and one piece of 4 in.). If the design of C_2 is as in part (a), then this changeover can be accomplished by just disabling one knife (knife 2). On the other hand, if the design of C_2 is as in part (b), then it requires shifting one knife and disabling another. Disabling a knife takes relatively smaller time compared to shifting a knife.

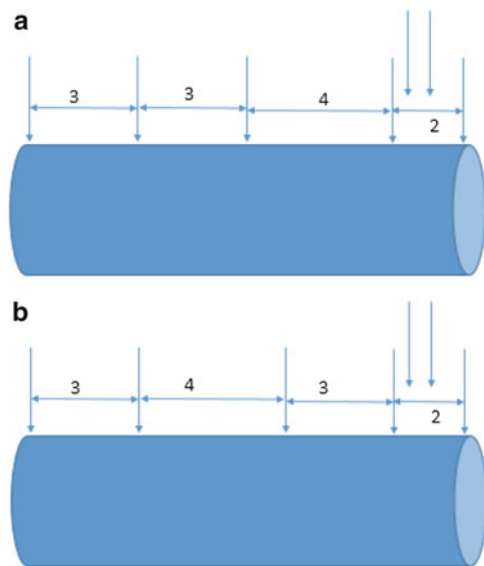


Fig. 5.10 Two different ways placing knives to cut the same cutting pattern C_2 . Arrows indicate the knife positions on the roll. Arrows not touching the roll indicate the disabled knives. Counting from left to right, knives 5 and 6 are disabled

5.6.2 Distance Between Two Cutting Patterns

Suppose A and B are two designs of two cutting patterns. Then the minimum number of knife changes required to change from A to B is same as that of changing from B to A . This number is unique and it is equal to 0 if A and B are same, else it is a positive integer. We shall define this number to be the distance between A and B . Remember that this distance is well defined only when the designs are fixed. If C_1, C_2, \dots, C_k , are the designs of k cutting patterns, then the square matrix $D = ((d_{ij}))$ of order k , where d_{ij} is the distance between C_i and C_j , is called the distance matrix of C_1, C_2, \dots, C_k . This matrix is a symmetric matrix.

5.6.3 Changeover Time Between Two Cutting Patterns

As observed in the previous subsections, the changeover time from one cutting pattern to another depends on how the two cutting patterns are designed. This means that the changeover time between two cutting patterns is not unique, rather it depends on how the knives are placed for the two cutting patterns. In order to make this distance well defined, it is necessary to adopt a convention of arranging widths in the cutting patterns (in other words, fix the designs of the cutting patterns). A convention that naturally arises is to put the widths in the ascending order. The cutting patterns

of cello tape problem put according to this convention are shown in Fig. 5.11. Also, we shall assume that the width of trim in each cutting pattern is less than the minimum of the widths in that cutting pattern (this can always be accomplished by producing an offcut of suitable width). With the convention just mentioned, the distance between two cutting patterns is well defined, and it is equal to the number of knife changes to move from one cutting pattern to the other. The distance matrix for the 10 cutting patterns listed in Table 5.1 is given in (5.15). Here C_0 is the cutting pattern which represents the starting point where all knives are free.

$$\begin{array}{c} \text{CPs} \end{array} \begin{array}{c} C_0 \quad C_1 \quad C_2 \quad C_3 \quad C_4 \quad C_5 \quad C_6 \quad C_7 \quad C_8 \quad C_9 \quad C_{10} \end{array} \begin{pmatrix} 0 & 5 & 5 & 5 & 4 & 4 & 4 & 4 & 4 & 4 & 3 \\ 5 & 0 & 3 & 3 & 1 & 2 & 3 & 2 & 3 & 2 & 2 \\ 5 & 3 & 0 & 3 & 3 & 3 & 3 & 2 & 2 & 3 & 3 \\ 5 & 3 & 3 & 0 & 3 & 2 & 3 & 3 & 2 & 3 & 3 \\ 4 & 1 & 3 & 3 & 0 & 1 & 2 & 1 & 2 & 1 & 1 \\ 4 & 2 & 3 & 2 & 1 & 0 & 2 & 2 & 1 & 2 & 2 \\ 4 & 3 & 3 & 3 & 2 & 2 & 0 & 2 & 2 & 2 & 2 \\ 4 & 2 & 2 & 3 & 1 & 2 & 2 & 0 & 1 & 1 & 1 \\ 4 & 3 & 2 & 2 & 2 & 1 & 2 & 1 & 0 & 2 & 2 \\ 4 & 2 & 3 & 3 & 1 & 2 & 2 & 1 & 2 & 0 & 1 \\ 3 & 2 & 3 & 3 & 1 & 2 & 2 & 1 & 2 & 1 & 0 \end{pmatrix} \quad (5.15)$$

5.6.4 Travelling Salesman Problem for Sequencing Cutting Patterns

Having defined the distance between any two cutting patterns, the problem of sequencing the cutting patterns can be formulated as a travelling salesman problem (TSP) [5, 11]. In simple terms, a TSP is that a salesperson has to visit n cities starting from one of the n cities, visit all cities, visiting each of the n cities exactly once, and come back to the city where salesperson started in the shortest distance possible. The inputs to this problem are the number of cities (n) and the distance matrix of these cities. In general, distance from city A to city B need not be same as the distance from city B to city A . Therefore, the distance matrix need not be symmetric in general. If the distance matrix is symmetric, then the TSP is called symmetric TSP, else it is called asymmetric TSP [12, 13]. Note that sequencing of cutting patterns is a symmetric TSP. The cutting patterns play the role of cities in this problem, and the distance between two cities (cutting patterns) is the minimum number of knife changes required to change one cutting pattern to the other. We shall first apply this to the cello tape problem and then to an instance of the deckle optimization problem.

Table 5.1 presents two cutting plans, Plan 1 and Plan 2, for the cello tape problem. Plan 2 uses five cutting patterns, namely, C_1 , C_3 , C_5 , C_8 and C_{10} . We shall add a dummy cutting pattern C_0 which has no widths (represents the starting city in the TSP). The distance matrix of the corresponding TSP is extracted from the

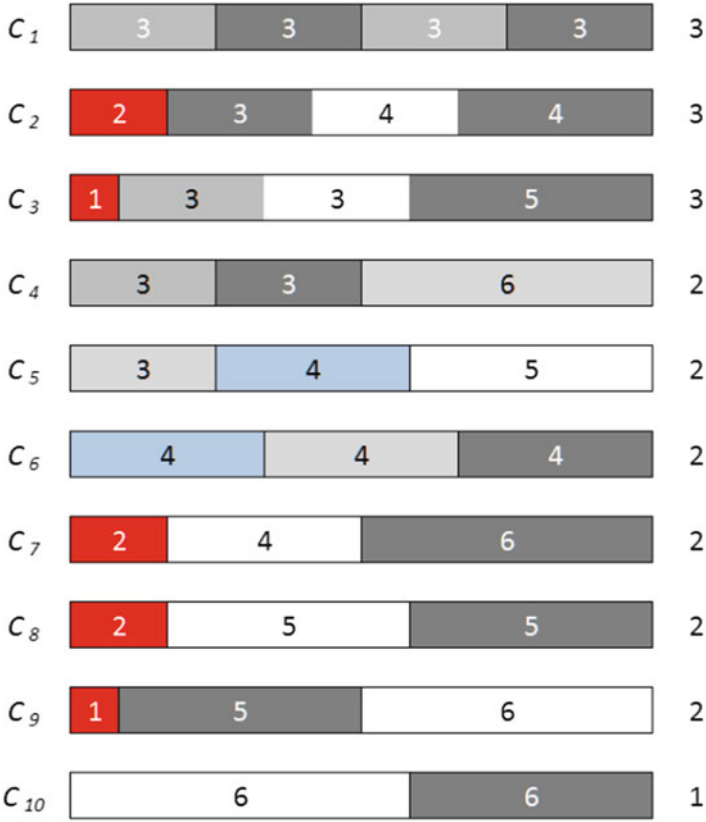


Fig. 5.11 Cutting patterns for the cello tape problem. The numbers on the right hand side of the cutting patterns are the number of knives set between the two extreme knives

matrix given in (5.15). It is the principal submatrix of the matrix in (5.15) corresponding to the columns labeled C_0 , C_1 , C_3 , C_5 , C_8 and C_{10} and is given in (5.16).

$$D = \begin{matrix} & \begin{matrix} \text{CPs} & C_0 & C_1 & C_3 & C_5 & C_8 & C_{10} \end{matrix} \\ \begin{matrix} C_0 \\ C_1 \\ C_3 \\ C_5 \\ C_8 \\ C_{10} \end{matrix} & \begin{pmatrix} 0 & 5 & 5 & 4 & 4 & 3 \\ 5 & 0 & 3 & 2 & 3 & 2 \\ 5 & 3 & 0 & 2 & 2 & 3 \\ 4 & 2 & 2 & 0 & 1 & 2 \\ 4 & 3 & 2 & 1 & 0 & 2 \\ 3 & 2 & 3 & 2 & 2 & 0 \end{pmatrix} \end{matrix} \quad (5.16)$$

Solving the problem, it is found that the optimal sequence is C_0 , C_3 , C_8 , C_5 , C_1 , C_{10} and C_0 , and the total number of knife changes is equal to 15. As an exercise, find the optimal sequence for Plan 1 of Table 5.1.

5.6.5 Optimal Sequence for Deckle Problem

We shall now apply the TSP model to sequence the optimal cutting patterns of a solution to the sample problem. Consider the solution presented in Fig. 5.9. In this solution, we have two sets of cutting patterns, one of them is the set of 21 cutting patterns used for reel orders of 95 cm diameter, and the other set has 15 cutting patterns used for sheet orders and reel orders of 100 cm diameter. Since the production at winder takes place diameter wise, we need to solve two TSPs – one for each set. The first TSP (TSP1) has 21 cutting patterns which correspond to reel orders of 95 cm diameter. Let us number these as C_1, C_2, \dots, C_{21} with C_i corresponding to the i th cutting pattern row under the cutting patterns of 95 cm diameter rolls in Fig. 5.9. The 15 cutting patterns of the second set shown in Fig. 5.9 are numbered as $C_{22}, C_{23}, \dots, C_{36}$ with C_i corresponding to the i th cutting pattern row under the cutting patterns of 100 cm diameter rolls in Fig. 5.9. Let us refer to the TSP corresponding to this as TSP2. The distance matrices and the optimal sequences for these two TSPs, TSP1 and TSP2, are presented in (5.17) and (5.18).

Distance Matrix and Optimal Sequence for TSP1

The distance matrix for TSP1 is given in (5.17). The optimal sequence is obtained by solving the TSP1 using a professional solver. It took 15 s of solver running time to get the optimal solution. The optimal sequence of the cutting patterns is in the following order: $C_{21}, C_{19}, C_{15}, C_9, C_6, C_{10}, C_{14}, C_{17}, C_{18}, C_{12}, C_7, C_{11}, C_2, C_1, C_4, C_3, C_5, C_{20}, C_8, C_{16}, C_{13}$. The starting knife position C_0 is omitted in the optimal sequence. The total number of knife changes for the optimal sequence is 215. If we assume that all knives are released and reset between any two cutting patterns, then any sequence will have the same number of knife changes, and this common number is an upper bound on the number of knife changes for any sequence of cutting patterns. This maximum number for TSP1 is 240. Thus, the optimal sequence reduces the number of knife changes by 25 from the worst case scenario.

$$\begin{array}{c} \text{CPs} \\ C_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ C_9 \\ C_{10} \\ C_{11} \\ C_{12} \\ C_{13} \\ C_{14} \\ C_{15} \\ C_{16} \\ C_{17} \\ C_{18} \\ C_{19} \\ C_{20} \\ C_{21} \end{array} \begin{pmatrix} \begin{array}{c} C_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ C_9 \\ C_{10} \\ C_{11} \\ C_{12} \\ C_{13} \\ C_{14} \\ C_{15} \\ C_{16} \\ C_{17} \\ C_{18} \\ C_{19} \\ C_{20} \\ C_{21} \end{array} \\ \begin{array}{cccccccccccccccccccccccccccc} 0 & 8 & 7 & 8 & 7 & 7 & 6 & 9 & 8 & 8 & 8 & 7 & 7 & 7 & 8 & 9 & 8 & 7 & 8 & 9 & 7 & 7 \\ 8 & 0 & 5 & 6 & 5 & 5 & 10 & 10 & 10 & 12 & 12 & 11 & 9 & 11 & 12 & 13 & 12 & 11 & 12 & 13 & 11 & 11 \\ 7 & 5 & 0 & 6 & 5 & 5 & 9 & 12 & 11 & 11 & 11 & 8 & 10 & 10 & 11 & 12 & 11 & 10 & 11 & 12 & 10 & 10 \\ 8 & 6 & 6 & 0 & 5 & 5 & 10 & 13 & 12 & 12 & 12 & 11 & 10 & 11 & 12 & 13 & 12 & 11 & 12 & 13 & 11 & 11 \\ 7 & 5 & 5 & 5 & 0 & 5 & 9 & 12 & 10 & 11 & 11 & 10 & 10 & 10 & 11 & 12 & 11 & 10 & 11 & 12 & 10 & 10 \\ 7 & 5 & 5 & 5 & 5 & 0 & 9 & 12 & 11 & 11 & 11 & 10 & 10 & 10 & 11 & 12 & 11 & 10 & 11 & 12 & 8 & 10 \\ 6 & 10 & 9 & 10 & 9 & 9 & 0 & 11 & 10 & 10 & 9 & 9 & 9 & 9 & 10 & 11 & 10 & 9 & 10 & 11 & 9 & 9 \\ 9 & 13 & 12 & 13 & 12 & 12 & 11 & 0 & 13 & 13 & 13 & 12 & 12 & 12 & 13 & 14 & 13 & 12 & 13 & 14 & 12 & 12 \\ 8 & 10 & 11 & 12 & 10 & 11 & 10 & 13 & 0 & 12 & 12 & 11 & 11 & 11 & 12 & 13 & 12 & 11 & 12 & 13 & 11 & 11 \\ 8 & 12 & 11 & 12 & 11 & 11 & 10 & 13 & 12 & 0 & 12 & 11 & 11 & 11 & 12 & 13 & 12 & 11 & 12 & 13 & 11 & 11 \\ 8 & 12 & 11 & 12 & 11 & 11 & 9 & 13 & 12 & 12 & 0 & 11 & 11 & 11 & 12 & 13 & 12 & 11 & 12 & 13 & 11 & 11 \\ 7 & 11 & 8 & 11 & 10 & 10 & 9 & 12 & 11 & 11 & 11 & 0 & 10 & 10 & 11 & 12 & 11 & 10 & 11 & 12 & 10 & 10 \\ 7 & 9 & 10 & 10 & 10 & 10 & 9 & 12 & 11 & 11 & 11 & 10 & 0 & 10 & 11 & 12 & 11 & 10 & 11 & 12 & 10 & 10 \\ 7 & 11 & 10 & 11 & 10 & 10 & 9 & 12 & 11 & 11 & 11 & 10 & 10 & 0 & 11 & 12 & 11 & 10 & 11 & 12 & 10 & 10 \\ 8 & 12 & 11 & 12 & 11 & 11 & 10 & 13 & 12 & 12 & 12 & 11 & 11 & 11 & 0 & 13 & 12 & 11 & 12 & 13 & 11 & 11 \\ 9 & 13 & 12 & 13 & 12 & 12 & 11 & 14 & 13 & 13 & 13 & 12 & 12 & 12 & 13 & 0 & 13 & 12 & 13 & 13 & 12 & 12 \\ 8 & 12 & 11 & 12 & 11 & 11 & 10 & 13 & 12 & 12 & 12 & 11 & 11 & 11 & 12 & 13 & 0 & 11 & 12 & 13 & 11 & 11 \\ 7 & 11 & 10 & 11 & 10 & 10 & 9 & 12 & 11 & 11 & 11 & 10 & 10 & 10 & 11 & 12 & 11 & 0 & 11 & 12 & 10 & 10 \\ 8 & 12 & 11 & 12 & 11 & 11 & 10 & 13 & 12 & 12 & 12 & 11 & 11 & 11 & 12 & 13 & 12 & 11 & 0 & 13 & 11 & 11 \\ 9 & 13 & 12 & 13 & 12 & 12 & 11 & 14 & 13 & 13 & 13 & 12 & 12 & 12 & 13 & 13 & 13 & 12 & 13 & 0 & 12 & 12 \\ 7 & 11 & 10 & 11 & 10 & 8 & 9 & 12 & 11 & 11 & 11 & 10 & 10 & 10 & 11 & 12 & 11 & 10 & 11 & 12 & 0 & 10 \\ 7 & 11 & 10 & 11 & 10 & 10 & 9 & 12 & 11 & 11 & 11 & 10 & 10 & 10 & 11 & 12 & 11 & 10 & 11 & 12 & 10 & 0 \end{array} \end{pmatrix} \quad (5.17)$$

Distance Matrix and Optimal Sequence for TSP2

The distance matrix for TSP2 is given in (5.18). It took 5 s for the solver to produce the optimal solution. The optimal sequence of the cutting patterns is in the following order: C_{12} , C_{14} , C_7 , C_8 , C_5 , C_{11} , C_{10} , C_6 , C_1 , C_3 , C_4 , C_2 , C_{13} , C_9 , C_5 . The number of knife changes for the optimal sequence is 76 and the maximum number is 108. The starting knife position C_0 is omitted in the optimal sequence.

$$\begin{array}{c} \text{CPs} \\ C_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ C_9 \\ C_{10} \\ C_{11} \\ C_{12} \\ C_{13} \\ C_{14} \\ C_{15} \end{array} \begin{pmatrix} \begin{array}{cccccccccccccccc} C_0 & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \end{array} \\ \begin{array}{cccccccccccccccc} 0 & 7 & 3 & 3 & 3 & 3 & 3 & 3 & 2 & 6 & 3 & 3 & 4 & 5 & 3 & 4 \end{array} \\ \begin{array}{cccccccccccccccc} 7 & 0 & 10 & 10 & 10 & 10 & 10 & 10 & 9 & 13 & 10 & 10 & 11 & 12 & 10 & 11 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 0 & 6 & 5 & 6 & 6 & 6 & 5 & 9 & 6 & 6 & 7 & 8 & 6 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 6 & 0 & 5 & 6 & 6 & 6 & 5 & 9 & 6 & 6 & 7 & 8 & 6 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 5 & 5 & 0 & 6 & 6 & 6 & 5 & 9 & 6 & 6 & 7 & 8 & 6 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 6 & 6 & 6 & 0 & 6 & 6 & 5 & 9 & 5 & 5 & 7 & 8 & 6 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 6 & 6 & 6 & 6 & 0 & 6 & 5 & 9 & 6 & 6 & 7 & 8 & 6 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 6 & 6 & 6 & 6 & 0 & 5 & 9 & 6 & 6 & 7 & 8 & 5 & 7 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 2 & 9 & 5 & 5 & 5 & 5 & 5 & 5 & 0 & 8 & 5 & 5 & 6 & 7 & 5 & 6 \end{array} \\ \begin{array}{cccccccccccccccc} 6 & 13 & 9 & 9 & 9 & 9 & 9 & 9 & 8 & 0 & 9 & 9 & 10 & 11 & 9 & 10 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 6 & 6 & 6 & 5 & 6 & 6 & 5 & 9 & 0 & 5 & 7 & 8 & 6 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 6 & 6 & 6 & 5 & 6 & 6 & 5 & 9 & 5 & 0 & 7 & 8 & 6 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 4 & 11 & 7 & 7 & 7 & 7 & 7 & 7 & 6 & 10 & 7 & 7 & 0 & 9 & 6 & 8 \end{array} \\ \begin{array}{cccccccccccccccc} 5 & 12 & 8 & 8 & 8 & 8 & 8 & 8 & 7 & 11 & 8 & 8 & 9 & 0 & 8 & 9 \end{array} \\ \begin{array}{cccccccccccccccc} 3 & 10 & 6 & 6 & 6 & 6 & 6 & 5 & 5 & 9 & 6 & 6 & 6 & 8 & 0 & 7 \end{array} \\ \begin{array}{cccccccccccccccc} 4 & 11 & 7 & 7 & 7 & 7 & 7 & 7 & 6 & 10 & 7 & 7 & 8 & 9 & 7 & 0 \end{array} \end{array} \end{pmatrix} \quad (5.18)$$

5.6.6 Complete Solution to Deckle Optimization Problem

In the previous section and this section we have seen how the deckle optimization problem is formulated using the models one-dimensional cutting stock problem and travelling salesman problem. The formulations used mixed integer linear programming problems. The formulations presented can be used to either minimize the material consumption or minimize the cutting patterns. Since production time is a major constraint at winder section, minimizing cutting patterns is important. But minimizing cutting patterns has an adverse effect on the material consumption. That is, as one tries to minimize the number of cutting patterns, the number of master rolls goes up. Thus, number of cutting patterns and number of master rolls are two conflicting factors in the deckle optimization problem. Formulations and solutions presented in this chapter can be used to minimizing one of the two conflicting factors while imposing an upper limit constraint on the other to explore optimal decisions. This provides a free hand to the user to decide on weighing between the number of cutting patterns and the number of master rolls.

Once the problems are formulated, they can be solved effectively using professional solvers such as COIN-OR, LINGO, CPLEX. As for the orders are concerned, industry is used to considering orders by weights and not by reel numbers. Therefore, to solve an instance of a problem, one needs to convert the weights into reel numbers and vice versa to formulate and solve the problem. The inputs from the user are usually provided in an excel file and these inputs should be converted into

inputs for the solver in a format that solver uses. Additional inputs such as time limits for the solver to terminate if optimal solution is not reached within the set time are to be captured each time an instance of a problem is solved. All these require a user-friendly software. For this project a software solution has been developed which is under testing with the company.

5.7 Summary

In this chapter we have presented a case study that provides a solution to the deckle optimization problem of paper and paperboard manufacturing industries. The project was taken up at the request of a leading paper and paperboard manufacturing company in India. The deckle optimization problem is critical to paper industry and results in potential savings. Starting with the relevant areas of production, the chapter presented the details of products (reels and sheets) and the process (winder and sheeter) and the decision making problems involved. The problem is modeled using the one-dimensional cutting problem and the travelling salesman problem. Various formulations of handling the cutting stock problems in the context of deckle optimization are presented. Reducing the number of cutting patterns is crucial to the deckle optimization problem. This problem of reducing the number of cutting patterns has been formulated intelligently. We believe that this is an important and significant contribution to the deckle optimization problem as it provides a free hand to the user to weigh between the two conflicting factors – the number of cutting patterns and the number of master rolls. A software solution has been developed for the client which is under testing at present. The software uses excel as the front-end as desired by the user.

Acknowledgements The author wishes to thank the client for providing the opportunity to work on this problem. This case study is a project undertaken by Aditya Gudimella as a part of the curriculum of the courses on Statistical Quality Control and Operations Research at the Indian Statistical Institute. With minimal guidance, he has been able to contribute significantly to this project. Two of his major contributions to this project are the formulation for minimizing the cutting patterns (5.6) and the development of the software solution.

Problems

5.1. Consider the reel orders for ABC product with 230 GSM, caliper 294μ with reel diameter 95 cm and core pipe diameter 3 in. The widths and quantities for these orders are given in Table 5.11. For this problem, take deckle width as 320 cm (that is, the width of the jumbo after taking away the end trim of 2 cm on each of the two ends).

Table 5.11 Widths and order quantities

Width (cm)	49.0	91.5	89.0	60.0	51.0	50.0
Quantity (kg)	1300	1500	1500	1685	1500	1500

- Compute the minimum number of reels required for each width.
- List 5 cutting patterns. Putting together the five cutting patterns, are all widths covered? If not, modify your list so that this happens. After that, find out the best cutting plan, using only the 5 cutting patterns, to minimize the number of sets (see page 95 for definition of set).
- Consider a cutting pattern $u = (u_1, u_2, \dots, u_6)$. Say that u is dominated by another cutting pattern $v = (v_1, v_2, \dots, v_6)$ if $u \leq v$ and $u \neq v$, that is, $u_i \leq v_i$ for $i = 1, 2, \dots, 6$ and $u_i < v_i$ for at least one i . Say that a cutting pattern is *undominated* if there is no cutting pattern that dominates u . Let P be the set of all cutting patterns and let U be the set of all undominated cutting patterns. In some sense, the dominated cutting patterns are redundant. Articulate this concept, formulate and prove that it is enough to consider the elements of U for the relevant optimization problem.
- Write a computer program to list all cutting patterns. What is the total number of cutting patterns?
- Write a computer program to list all undominated cutting patterns only. What is the difference in size between P and U ?

5.2. Complexity of Generating Cutting Patterns

Consider the one-dimensional cutting stock problem with inputs $W = 380$, $w = (127, 82, 131, 78, 110, 120, 100, 66, 85, 77, 105, 58, 136)$ and $q = (23, 29, 19, 7, 14, 4, 12, 9, 2, 21, 15, 10, 15)$. Try and see if you can write a program to enlist all undominated cutting patterns for this problem. If this becomes too much, generate your own cutting patterns, as many as you wish, that you think will be useful cutting patterns. Using your cutting patterns, solve this cutting stock problem. Can you guarantee that your solution is the best?

5.3. How to optimize both objectives?

Suppose we want to explore the solutions with respect to both objective functions, the number of master rolls and the trim loss. Solve the problem to minimize the number of rolls first. Suppose this number is m_0 . Now solve the problem with trim loss as objective function but with an additional constraint that the number of master rolls has to be less than or equal to m . Let the trim loss be $TL(m)$. Solve for $TL(m)$ for $m = m_0, m_0 + 1, m_0 + 2, \dots$. If you plot $TL(m)$, you must get a convex function. Do this exercise for the cello tape problem.

5.4. Get a better efficiency measure if possible.

Consider a cutting stock problem with inputs: W as the master roll width, width vector (w_1, w_2, \dots, w_k) and quantity vector (q_1, q_2, \dots, q_k) (that is, q_i pieces of width w_i are required). Ideally, if you have a master roll, an imaginary master roll, with width $\sum_{i=1}^k w_i q_i$, then you can produce all required pieces with zero trim loss. Therefore,

we use this $\sum_{i=1}^k w_i q_i$ for defining a measure of efficiency as follows. If a solution to this problem consumes m master rolls, then a measure of efficiency of the solution may be taken as $100 \sum_{i=1}^k w_i q_i / mW$ (equivalently 100 times weight of an imaginary master roll of width $\sum_{i=1}^k w_i q_i$ divided by weight of an imaginary master roll of width mW). It is possible that a global optimal solution to the problem may have this efficiency percentage strictly less than 100. Cite one instance for this case. I wish to improve this measure of efficiency if possible. Explore and formulate this as an optimization problem.

5.5. Develop computer programs for:

Develop a computer program to compute the distance matrix for sequencing cutting patterns (see Sect. 5.6). The problem of finding the best sequence along with ordering of widths in the cutting patterns is the same problem that we saw in Chap. 2. The problem requires solving travelling salesman and assignment problems sequentially until a local optimal solution is obtained. Write code for that problem.

5.6. Compute number of reels for sheet order.

For the sample problem, compute the number of reels required for each of the sheet orders if only 95 cm diameter reels are used. Do this only if 100 cm diameter reels are used. Consider the problem of distributing the a sheet order between the two diameters. That is, some of the sheets for this sheet order are taken from 95 cm diameter rolls, and the rest from the 100 cm diameter rolls. The objective is to minimize the total quantity (by weight). Formulate this problem as an optimization problem and solve it for the sheet order ID 040 of Table 5.4 using excel solver.

5.7. Determine the diameter for sheet orders with sheeter facilities.

In Sect. 5.5.5, the problem of determining the optimal diameter for sheet orders exclusively was considered. In the formulation given there, it is assumed that exact sheet widths are only considered. However, by allowing pieces of widths \bar{w}_i , \bar{w}'_i , \bar{w}''_i used in the formulation (5.10), there is scope for better optimization. State this problem clearly and formulate it. This formulation (5.14) used a nonlinear constraint. Can you find good formulation or approach for this problem with only a mixed integer linear programming problem?

References

1. Biermann, C.J.: Handbook of Pulping and Papermaking. Academic, San Diego (1993)
2. Bovet, D., Pierluigi, C.: Introduction to the Theory of Complexity. Prentice-Hall, New York (1994)
3. Farley, A.: Mathematical programming models for cutting stock problems in the clothing industry. J. Oper. Res. Soc. **39**, 41–53 (1988)
4. Farley, A.: The cutting stock problem in the canvas industry. Eur. J. Oper. Res. **44**, 247–255 (1990)
5. Gutin, G., Punnen, A.P.: The Traveling Salesman Problem and Its Variations. Springer, Heidelberg (2006)

6. Haessler, R.W.: A new generation of paper machine trim programs. *TAPPI J.* **71**, 127–130 (1988)
7. Haessler, R.W., Sweeny, P.E.: Cutting stock problems and solution procedures. *Eur. J. Oper. Res.* **54**, 141–150 (1991)
8. Papadimitriou, C.H., Steiglitz, K.: *Combinatorial Optimization: Algorithms and Complexity*. Dover, Mineola (1998)
9. Sweeney, P.E., Haessler, R.W.: One-dimensional cutting stock decisions for rolls with multiple quality grades. *Eur. J. Oper. Res.* **44**, 224–231 (1990)
10. Sweeney, P.E., Paternoster, E.R.: Cutting and packing problems: a categorized, application-oriented research bibliography. *J. Oper. Res. Soc.* **43**, 691–706 (1992)
11. Taha, H.A.: *An Introduction to Operations Research*. Prentice Hall, New Jersey (2007)
12. Volgenant, A., Jonker, R.: A branch and bound algorithm for the symmetric traveling salesman problem based on the 1-tree relaxation. *Eur. J. Oper. Res.* **9**, 83–89 (1982)
13. Volgenant, A., Jonker, R.: The symmetric traveling salesman problem and edge exchanges in minimal 1-trees. *Eur. J. Oper. Res.* **12**, 394–403 (1983)

Chapter 6

A Scientific Tool for Workforce Management in BPO Companies

Abstract There are many industries like the Business Process Outsourcing companies or industries where the operations involve receiving transaction/requests from clients and addressing them within a stipulated time. A common problem that is faced by these industries is that of planning and managing the agents/associates who receive the calls and address them. It turns out that this problem is too complex to handle manually without the aid of a software solution. A software tool that acts like decision support systems is badly required for routine decision making for this kind of problems. This work has evolved from one such requirement from an Information Technology Enabling Services (ITES) industry where the management found it very difficult to plan and manage effectively their human resources. Following a scientific approach, the company's problem is formulated as an integer-linear programming problem and a software tool is developed to aid the management. The tool is Excel based, simple to use and can be effectively deployed as a decision support system. The tool is applied to various projects using past data to evaluate and compare different types of solutions. Based on the analysis, a useful metric is proposed to assess the management decisions for managing the human resources. The primary objective of this work is to promote business excellence through scientifically designed low cost solutions for the industries related to BPO and ITES.

6.1 Introduction

Optimization and management of resources are crucial for business excellence in any organization [1–3]. Scientific approach coupled with software technology can provide amazing solutions in this regard. Many of the decision making problems in practice are very complex in nature and deal with hundreds or thousands of variables and constraints [5]. Such problems can be solved effectively only with the help of Decision support systems (DSSs). DSS, usually a software package, is a tool that

will take all necessary inputs (preferably minimal set of inputs) from the decision maker and present her/him the optimal or near optimal solution to the problem at hand in pre-designed user-friendly format(s) [7, 8]. DSS does lot of activities in the background, which the user or the decision maker does not get to see or understand, to derive the solution. DSS takes the necessary inputs, uses a pre-built backend model taking all objectives, constraints, formulation and solution methods into account, carries out all necessary computations in the background to obtain the solution each time it is run or executed, and finally presents the solution in formats designed for the end user.

DSSs are indeed available for many industrial problems. However, the awareness of DSSs and their utilities is mostly lacking among decision makers/industries. Even where there is awareness, people are not in a position to adapt due to prohibitive and unaffordable costs of the DSSs. In light of this, offering DSSs which are easily affordable and easy to implement will make significant contribution to business excellence. The work presented in this chapter is an effort towards one such contribution. It presents an excel based DSS to one of the most widely encountered problems in Information Technology Enabled Services (ITES)/Business Process Outsourcing (BPO) industries.

ITES/BPO companies across various industries are facing a common challenge of increasing gap between demand and supply. Typically, the operations involve receiving calls from the client base and addressing them within the stipulated time. The stipulated time is known as the Turn Around Time (TAT). The basic decision making problem here is to plan for the resources, the agents/associates, who would address the calls and utilize them efficiently. At present, most of the ITES/BPO industries plan these activities using experience based human intelligence. The difficulty in doing this and its bearing on ineffectiveness of the solutions will be highlighted through a live example.

The work reported in this chapter has evolved from the requirement of an ITES industry. It comprises the development of an excel based software tool that acts as a DSS to plan and monitor the resource management and service level agreement (SLA). It discusses various issues involved in the problem, and how the problem is formulated as a mathematical programming problem. Though the problem is formulated as mathematical programming problem which can be solved using excel solver or other professional operations research packages, an excel based macro is developed to solve it more efficiently than the solvers. Solutions of the live examples are presented in the chapter to emphasize the solution methodology, its simplicity and effectiveness.

The organization of this chapter is as follows. Section 6.2 presents the background of the problem and discusses various issues involved in it with the help of a live example from the ITES industry. Section 6.3 spells out the objectives of this project. Section 6.4 presents the approach and the formulation of the problem. Section 6.5 presents the results of the study with reference to the live example, the effectiveness of applying the software tool, and a comparison between the actual solutions implemented for the past data and what would be the impact of applying the tool on the same. Section 6.6 presents the details of the software tool developed for this project. Section 6.7 presents the results of the application of the tool to live data. Section 6.8 summarizes the work.

6.2 Problem Background

The ITES company in question receives calls from its international clients in Health-care industry on a daily basis. In this case, the calls are typically health insurance claims to be checked for their validity and the payments to be released for the valid claims within the stipulated time. The calls are divided into various categories and the calls in each category are processed independently by different teams. In view of the SLAs entered with the clients, each call that arrives at the company must be addressed within TAT – the time period committed in the SLA. It is very important to meet TAT, that is, address the calls within the committed period of time, failing which will result in penalties and more than penalties, loss of company's image with the clients. In view of this, the company deploys adequate staff to answer the calls within TAT. The major decision making problem here is to determine the optimal size of the team, that is, the number of agents/associates who address the calls.

At present, many ITES/BPO industries are making the decision of staffing using experience and intelligence. This task of manual and subjective decision making is extremely difficult and mostly results in improper and inefficient planning. This will be demonstrated using a live example from the company's recent past data. But, with a scientific approach, the problem can be modeled as a mathematical problem and effective solutions can be derived. The awareness of this fact is very much lacking in the industry. In fact, the company was very much impressed and got excited about the solution provided to the problem through this work.

6.2.1 Problem Description

Depending on the type, calls are divided into categories and each category is processed independently. It suffices to describe the process for any one type of calls. The analysis and solution methodology can be extended to all other categories in a similar fashion. Henceforth, the discussion will be confined to one category (type/group) of calls. Live data from one category of calls from the company are used for this study. A group of agents is dedicated to the category. The calls arrive through web round the clock. On arrival, each call gets assigned to an agent in the group provided an agent is free. Otherwise, it will be waiting in queue until an agent gets free. Calls are picked on first-come-first-serve basis. Normally, agents work during office hours. The company in question works on 5-day week basis (Monday to Friday) in a single shift. However, the work presented in this chapter can be extended to a more general set up.

The main decision making problem here is how many resource persons need to be deployed to serve the calls. This decision depends heavily on two aspects: (i) call arrivals and (ii) calls serviced. Typically, both these are random variables. However, in practice, the planning is done always on certain predictions. For the problem in

question, the planning is done on a monthly basis. That is, the number of agents to be deployed into the category is decided at the beginning of each month. Since the exact number of calls for the future is not known in advance, the planning is based on the expected call arrivals. The call arrivals are random but exhibit a fairly stable pattern over days in each month. The number of calls arrived against the working day number are presented in Table 6.1 and Fig. 6.6 for the months January to October 2010. It can be seen that the call volumes exhibit a stable decreasing trend. The call arrivals on week ends and holidays are negligible.

Once a decision is made about the number of agents/associates in the team which addresses the calls, it is not possible to drop any member of the team, at least not within the month. As the volumes decrease over days in the month, the idle time of the members keeps increasing as time passes towards the month end. If the extent of idle time can be assessed before hand, then the idle times of the agents can be planned effectively for activities such as improving their skills, availing their services during peak volumes by offering lay offs during the slack period and so on.

One of the important metrics in this problem is the distribution of TAT. The number of days that a call takes to get addressed is a random variable. If this is denoted by T , then the possible values of T are 1,2,3,... It is desired that the distribution of T is positively skewed and that the probability of T taking a value above TAT should be zero. The TAT distribution for the month of March 2010 as per the actual execution of calls is shown in Fig. 6.1. The distribution clearly indicates the deficiency in the planning and management. The distribution is negatively skewed with 36 % of the calls exceeding the TAT of 8 days.

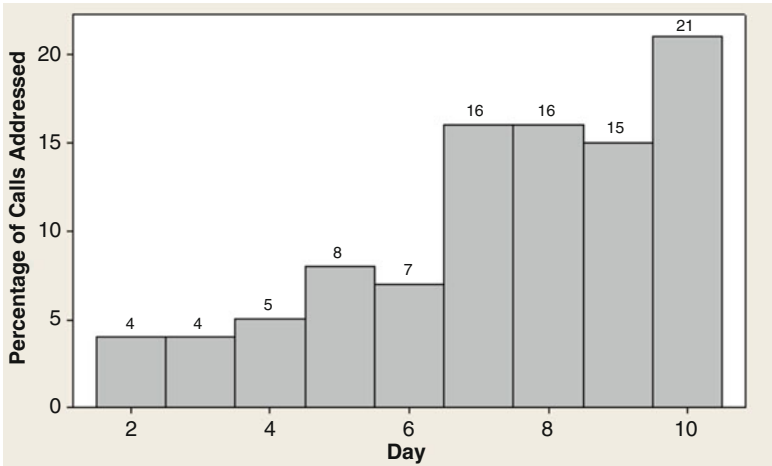


Fig. 6.1 Distribution of turn around time (TAT), the number of days to address a call

6.3 Objectives

The primary objective of this work is to provide a software tool to determine the optimum number of resource persons (agents/associates) to address the calls without violating the TAT condition. The set up of this problem is generic to a wide range of applications. A call can be viewed in different ways. For example, calls can be insurance claims (health, accidents, warranties, etc.) made, calls may be orders placed on a printing press, etc. Therefore, this work can be extended to a variety of applications. Since time taken to address a call depends upon the skill levels of the operating personnel, and since the variation in the skill levels is inevitable, it has been decided to consider the problem in terms of optimal number of calls to be addressed each day during the planning period so as to fulfil the requirements such as TAT conditions. Once the number of calls is determined, the resource allocation can be made based on the resource skills available, if necessary by marginally increasing the work loads by paying overtime etc.

Another important objective is to build a component in the software tool that will enable the manager or the decision maker to analyze and understand the consequences of a given scenario. This will be extremely useful to the manager as it will help in evaluating various options practically available and choosing one among them for implementation. It will also help in understanding utilization level of the resources. Yet another important objective is to make the tool dynamic so that manager can study the scenario on any day and can take mid course corrective actions. This will be particularly useful as the predictions/forecasts used at the beginning of the planning horizon will turn into partial realities as the call volumes get unfolded each day.

Further, it is intended, as a part of the study, to analyze the statistical aspects of the call arrivals and calls processed. It is also aimed that the tool will be used to study the efficacy of the scientific approach using past data. To sum up, the broad objectives of the study are listed in the following points:

1. To provide scientific models for the problem using mathematical formulations which will help in determining optimal solutions,
2. To build an excel based software tool which will act as a decision support system to managers in industries with operations that are similar to operations of BPO organizations,
3. Build the tool in such a way that it is user-friendly, easy to adapt and is available at low cost or no cost,
4. Build the tool with an optimizer module which will help in determining *call planner*,
5. Build the tool with an evaluation module which will help in understanding the effect and implications of a given solution or a plan,
6. Build the tool in such a way that it is dynamic in the sense that it should be useful in evaluating the status of a plan on any day after replacing the predictions/forecasts with actual calls arrived and processed known thus far,

7. Study the nature of call arrival distribution and the distribution of the calls processed for the live problem selected for the study using past data,
8. Assess the efficacy of the decisions made in the past based on experience and human intelligence,
9. Assess the efficacy of the solutions offered by the scientific approach on the same past data and compare it with the solutions actually implemented.

6.4 Approach

In order to model the problem it is important to introduce the necessary notation and terminology unambiguously. This is done by taking the present problem as an example for model building. For any similar problems, the model can be adapted with suitable definitions and modifications. The present problem involves receiving calls on a daily basis and the planning is done for a month in advance, usually at the beginning of each month. Planning horizon is the period for which the staffing is to be planned. For the purpose of modeling, staffing is defined as the number of calls, x_i , planned to be answered on Day i . Let t_0 denote the value of TAT. By definition of TAT, every call that arrives on a day should be answered within t_0 working days (the count of t_0 working days includes the day of arrival and/or the day on which the call is answered provided it is a working day). Assuming that the work starts each day at 8 am, a day is defined as 8 am of that day to 8 am of the next day.

To fix ideas and introduce further terminology, the following example is used. Imagine that staffing was to be planned for the month of March 2010 on March 01, 2010. For ease of understanding, required data are presented in Fig. 6.2 partially. In the third row of the figure, “H” stands for a holiday and “W” stands for a working day. Due to space constraints, some dates are skipped in the figure. The stipulated TAT is given by $t_0 = 8$. Since $t_0 = 8$ and the planning is done on March 01, 2010, calls that arrived on or before February 17, 2010 will not be relevant to the problem because the deadline for such calls is February 26, 2010 or before (see Fig. 6.2). Therefore, it is necessary and sufficient to consider the data for the period starting from February 18, 2010 to March 31, 2010. *In other words, for the purpose of modeling, it is assumed that calls that are valid according to TAT are only taken into account for the purpose of planning.*

Sno	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	...	42
Date	18-Feb	19-Feb	20-Feb	21-Feb	22-Feb	23-Feb	24-Feb	25-Feb	26-Feb	27-Feb	28-Feb	01-Mar	02-Mar	03-Mar	04-Mar	05-Mar	06-Mar	07-Mar	08-Mar	...	31-Mar
Day	W	W	H	H	W	W	W	W	W	H	H	W	W	H	W	W	H	H	W	...	W
V_i	19	34	0	0	48	28	30	0	0	0	0	388	287	200	165	154	9	6	127	...	64

Fig. 6.2 Sample data for modeling

The earliest date whose calls fall due on the day of planning will be called the **earliest due date** (EDD). The day of planning will be called the **first planning date** (FPD). The last date in the planning horizon will be called the **last date of planning** (LPD). For the example in question, EDD is February 18, 2010, FPD is March 01, 2010, and LPD is March 31, 2010. For the ease of modeling and notation, the days are numbered from EDD to LPD. Thus, day 1 refers to EDD, day 2 to February 19, 2010, and so on, and day 42 refers to LPD (see the first row of Fig. 6.2). Throughout this chapter n will stand for the day number of LPD and m will stand for the day number of FPD. Thus, for the example, $n = 42$ and $m = 12$.

6.4.1 Call Volumes

By definition, call arrivals or volumes of day i are the calls that arrive during the period of 8 am of day $(i - 1)$ to 8 am of day i and will be denoted by v_i . And calls processed during day i are defined as the calls processed during the period of 8 am of day i to 8 am of day $(i + 1)$. The interpretation of v_i depends upon day i . If $i \leq m$, then v_i stands for the number of calls that arrived on day i and are still pending at 8 am of FPD. If $i > m$, then v_i is the predicted number of call arrivals of day i .

6.4.2 The Decision Variables

Let x_i denote the number of calls planned to be processed on day i , $i = m, m + 1, \dots, n$. For $i < m$, set $x_i = 0$.

6.4.3 Basic Assumptions

Clearly the inputs to the staffing problem, the call volumes, are partly deterministic and partly stochastic in nature. Stochastic Programming techniques [6, 9] may be used for dealing with problems like the one at hand. However, these techniques ultimately use expectations and variances of the uncertain inputs. In order to minimize complications, this work assumes that call volumes and the calls processed are deterministic. In view of the stable patterns in the data, these assumptions may not hamper the solutions badly.

Another important assumption that is made is that the calls that arrive during a day are not processed during the same day. The rationale behind this assumption is that the calls that arrive during a day may arrive any time during the day and there may not be adequate time for processing them on the same day. While this assumption is necessary for formulating the problem, it may be violated in actual implementation. But that would not seriously affect the model solutions as the

numbers can be suitably modified. This aspect will be better clarified later under the discussion of model solutions to the live example that will be analyzed.

The staffing problem can be formulated as integer linear programming problem (see Sect. 6.5). Since this requires call volumes, part of the approach is devoted to understanding the call volume patterns from the past data and arriving at a forecasting method. Another issue that needs to be studied is the capacity to address the calls. This is necessary because the solution to the staffing problem will be in the form of number of calls to be processed and that needs to be translated into number of resource persons.

The objective of the staffing problem is to seek answers to two specific questions which will help in planning. The first question seeks the number of resource persons to be deployed, and the second question seeks the answer to “How to utilize the planned resources?”. The problem is solved using a two-stage approach. The answers can be obtained by solving two optimization problems (integer linear programming problems). The first problem leads to a minimax problem and the second problem uses the answer to the first problem.

6.5 Formulation

This section will present the mathematical formulations of the staffing problem. The problem is solved in two stages. The first stage problem involves solving a minimax problem and the second stage problem takes the answer to the first stage problem and formulates a new problem with additional constraints.

6.5.1 Constraints of the Staffing Problem

In order to develop constraints, it is necessary to introduce some more terminology. For any i , $1 \leq i \leq n$, let d_i denote the number of calls that must be completed on day i . Then we must have $x_i \geq d_i$. In the example under consideration we have: $d_i = 0$ for $i = 1, 2, \dots, 11$, and $d_{12} = 19$ as the deadline for the calls that arrived on February 18, 2010 is end of March 01, 2010 and 19 of such calls ($v_1 = 19$) are still pending at the beginning of March 01, 2010 (see Fig. 6.2). Similarly, the deadline for calls that arrived on February 19, 2010 is March 02, 2010. Note that the deadline for calls that arrived during February 20–23, 2010 is March 04, 2010 as holidays do not count under TAT condition (see Fig. 6.2).

Define p_i as the cumulative dues, that is, $p_i = \sum_{j=1}^i d_j$. Define c_i as the cumulative volumes, that is, $c_i = \sum_{j=1}^i v_j$. Next, define q_i as the **opening calls**, that is, the number of calls that are available for processing during day i . As before, set $q_i = 0$ for $i < m$ as these are concerned with past. Note that there is a recursive formula for the q_i which is given by

$$q_m = c_m, \text{ and } q_{(i+1)} = c_i - x_i + v_{i+1} = c_{i+1} - x_i, \quad i = m, m+1, \dots, n-1. \quad (6.1)$$

In order to make the model more flexible, the manager will be given the option of specifying upper limits on the number of calls that can be processed during day i . Let s_i be the maximum number of calls that can be processed on day i . Again, set $s_i = 0$ for $i < m$. To set the stage for formulating the constraints, define u_i , the cumulative calls processed, that is, $u_i = \sum_{j=1}^i x_j$. The constraints are listed below.

Constraint 1: The number of calls processed during a day cannot exceed the number of opening calls and the maximum number of calls that can be processed on that day. Thus, we must have: $x_i \leq q_i$ and $x_i \leq s_i$ for $i \geq m$. Note that $x_i \leq q_i$ is equivalent to $u_i \leq c_i$.

Constraint 2: Maintaining the TAT condition. This is ensured by setting $u_i \geq p_i$ for $i \geq m$. To see this, note that $u_m \geq p_m$, which is same as $x_m \geq d_m$, ensures that calls which are due on FPD are cleared on FPD. Next, $u_{(m+1)} \geq p_{(m+1)}$, implies, $x_m + x_{(m+1)} - d_m \geq d_{(m+1)}$. The left hand side of this last inequality is the number of calls processed during FPD and the next day after clearing the FPD dues. Therefore, the inequality ensures that day $(m+1)$ dues will be cleared by end of day $(m+1)$. This way the constraints ensure that the TAT conditions are met.

Constraint 3: Finally, it is obvious that all x_i s must be nonnegative integers.

6.5.2 Two-Stage Optimization and the Objective Functions

When planning the number of calls (equivalently the resource persons) over the working days of the planning horizon, the management has to plan for the same number of calls on each of the working days due to payment and other regulations. However, the management will have the option of utilizing the planned hours flexibly. If x is the constant number of hours planned over each of the working days of the planning horizon, then the management may utilize x_i hours on i th working day so that $x_i \leq x$ for any i . Therefore, the objective would be to choose x_i s and x so that all transactions are completed with in TAT. This objective will be achieved by solving two optimization problems in two stages. The first stage problem will find x and the second stage problem finds the x_i s. The formulations of the two stages are given below.

6.5.3 Formulation of Stage-1 Problem

In this stage we find out the common x . The objective function is

$$f(x_1, x_2, \dots, x_n) = \max_{1 \leq i \leq n} x_i$$

With this objective function the problem is the following integer linear programming problem.

$$\begin{array}{ll}
 \text{Minimize } & x \\
 \text{Subject to} & \\
 & \sum_{j=m}^i x_j \geq p_i, \quad i = m, m+1, \dots, n, \\
 & \sum_{j=m}^i x_j \leq c_i, \quad i = m, m+1, \dots, n, \\
 & x_i \leq s_i, \quad i = m, m+1, \dots, n, \\
 & x_i \leq x, \\
 & x_i\text{s are nonnegative integers.}
 \end{array}$$

6.5.4 Formulation of Stage-2 Problem

Having obtained the minimum x in Stage-1, the problem is to find the minimal x_i s which will meet the TAT requirements. Let x_0 be the optimal value of the objective function of the Stage-1 problem. In Stage-2 problem the objective function is taken as $\sum x_i$ which will be minimized with constraints $x_i \leq x_0$ for each i along with the constraints of the Stage-1 problem.

$$\begin{array}{ll}
 \text{Minimize } & \sum x_i \\
 \text{Subject to} & \\
 & \sum_{j=m}^i x_j \geq p_i, \quad i = m, m+1, \dots, n, \\
 & \sum_{j=m}^i x_j \leq c_i, \quad i = m, m+1, \dots, n, \\
 & x_i \leq s_i, \quad i = m, m+1, \dots, n, \\
 & x_i \leq x_0 \quad \text{for all } i \\
 & x_i\text{s are nonnegative integers.}
 \end{array}$$

The above problems can be solved using the solver that comes with Microsoft Excel or any other suitable solver such as LINGO, CPLEX, COIN-OR, etc. However, the Stage-1 problem can be solved very efficiently by writing a simple program. As a part of the software tool developed under this work, a macro is developed to solve this problem.

6.6 The Software Tool

The main objective of this work is to develop an Excel based software tool that will help as a decision support system for the industrial organizations dealing with problems that fit into the framework of the problem for which the solution is provided in this work. This section will describe the details of the software tool, its design, the input and output formats, and information that will be useful for the managers/decision makers.

6.6.1 Features

The software tool is aimed at providing multiple features so that it can be effectively used not only to determine the optimal solutions but also make it useful in evaluation of various options and make it dynamic so that situations can be reviewed on any day of the planning horizon taking into account the realizations of the forecasted call volumes. The intended features are listed below.

- Feature 1: Make the tool simple and user friendly with minimum input requirements.
- Feature 2: Provide a comprehensive screen in which all important inputs and outputs can be seen.
- Feature 3: Provide the option to choose the planning horizon flexibly so that it is not necessarily a rigid 1 month period.
- Feature 4: Provide the option to change the TAT as desired.
- Feature 5: Allow changes in s_i s, the upper limits on calls that can be processed, flexibly.
- Feature 6: Allow calls to be processed on holidays if desired.
- Feature 7: Obtain solutions to problems of Formulation 1 and Formulation 2.
- Feature 8: Allow mid-course reviews.
- Feature 9: Display utilization metric and the TAT distribution.

6.6.2 Main Screen

As mentioned before, the software tool is developed in Microsoft Excel. The present version of the tool comprises of two interactive sheets. One of them, titled “Volumes,” is used for providing inputs, and the other, titled “Direct,” is used for seeking solutions and providing partial inputs for analyzing various options. There are

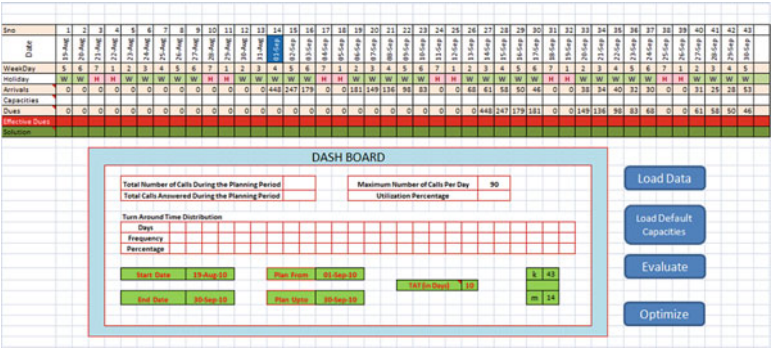


Fig. 6.3 Snapshot of the screen “Direct” of the software tool

four tabs in the screen “Direct”, named “Load Data”, “Load Default Capacities,” “Evaluate” and “Optimize.” In the screen, there is a row that presents the relevant dates, the row below that indicates whether the dates are holidays (H) or working days (W). The next two rows present the call volumes v_i s and the call upper limits (s_i s). The row titled “Calls Processed” presents the solution (x_i s). The row above this, titled “Dues,” presents the *status* of the solution, that is the number of calls missing the TAT. If all the entries in this are equal to zero, it means that the TAT condition is met by the solution. Of these rows, the user has permission to modify the entries in the row titled “Capacities.” The entries in this row are the s_i s, the maximum number of calls that can be addressed (Fig. 6.3).

In the box titled “Dash Board” there are few inputs that are expected from the user. These are Plan From date (to specify FPD) and Plan Up to date (to specify LPD). After entering these two dates, the user is expected to click the Load Data button. Up on clicking, the first four rows (sno, Date, Holiday and Call Volume rows) get filled with relevant data by picking the same from the Sheet “Volumes” where the basic data are previously loaded. The Dash Board also provides information on the total call volumes and the total number of calls answered by the solution, the utilization percentage and the TAT distribution. The formula for the utilization percentage is given by:

$$\text{Utilization Percentage} = \frac{100 \times \text{Average of } x_i\text{s}}{\text{Maximum of } x_i\text{s}} \quad (6.2)$$

The rationale behind this formula is the assumption that the deployment of resource persons is on monthly basis. If a certain number of people are deployed on a day, then the same number of people have to be deployed on any other day during the planning horizon. This metric may be little conservative but can be used as a useful indicator. More meaningful metrics may be computed using ground realities.

The button “Load Default Capacities” is used for setting uniform upper limit on the calls that can be processed (s_i s). To do this, the user has to first enter a number for the “Maximum Number of Calls Per Day” and then click the “Load Default Capacities” button. Upon clicking this button, all the s_i s (for $i \geq m$) will be set equal to the number specified for the “Maximum Number of Calls Per Day.” After doing this, if the “Evaluate” button is clicked, then a solution will be produced in which all x_i s are less than or equal to the specified maximum number of calls per day.

Finally, the “Optimize” button is used for getting an optimal solution for the given problem. To obtain optimal solution, the user needs to specify the FPD and LPD, click the “Load Data” button and after the data are loaded, press the “Optimize” button. On pressing this button, the tool will find an optimal solution, using a built-in macro, and post it in the screen. In the present version of the tool, only Formulation 1 is implemented for obtaining the optimal solution.

6.6.3 Inputs Screen

The main inputs for the problem should be provided in the sheet named “Volumes.” The main inputs required here are the dates, whether the dates are working days or holidays, and the call volumes. A snapshot of the format of the sheet “Volumes” is shown in Fig. 6.4. Since the call volumes change dynamically depending upon the call arrivals and calls processed, the user must ensure that right inputs are fed in the sheet “Volumes.” In actual practice, the user maintains a system file of call arrivals and calls processed which is updated, say, at 8 am every morning. The call volumes can be extracted from that file. For analyzing the past data, a file of the live data for the year 2010 is used in this study.

6.7 Application of Tool to Live Data

In order to understand the utility of the software tool, live data pertaining to one category of calls are used. The data are collected from January 01, 2010 to August 30, 2010. For the purpose of analysis, the data are exported to an excel sheet, and Fig. 6.5 presents a snapshot of the sheet. Basically there are five columns in this data. The first column is the date and all dates are present irrespective of whether it is a holiday or a working day. The column next to date shows the day of the date. Third column indicates whether the corresponding date is a holiday or a working day. The number 1 in this column stands for holiday, and 0 stands for a working day. The next column titled “Arrivals” presents the number of calls *received on the corresponding day*. The last column shows the actual number of calls addressed on the corresponding day. For the purpose of analysis, data were considered from 30-12-2009 to 30-08-2010. This is because the calls received prior to 30-12-2009 were all cleared before 01-01-2010. The arrivals shown against the dates 30-12-2009 and 31-12-2009 are the calls of respective dates pending as on January 01, 2010. For the rest of the dates (from 1st January onwards), the arrivals are the actual arrivals. Based on the call volumes the company has deployed three resource persons with an expectation of 40 calls per day to be addressed by each one of them. The first query from the company is whether the deployment of three persons is correct, if not, what is the optimal number. As this would depend on call arrivals and call departures (calls addressed), the first step towards answering the query is to examine the call arrival and departure patterns. It is observed from the data that calls do arrive on holidays and calls are processed sometimes on holidays due to exigencies of meeting TAT. But such volumes are considerably low. In any case, there is no harm in attributing the arrivals of holidays to the first working day that follows them. For instance, 1st, 2nd and 3rd of January are holidays and the first working day that follows them is 4th January. The arrivals of 1st, 2nd and 3rd, namely 63, 29 and 21 are added to 356 calls of 4th January, making call arrivals for 4th January as 469 (see Table 6.1). This way the call arrivals are calculated for each working day. The call

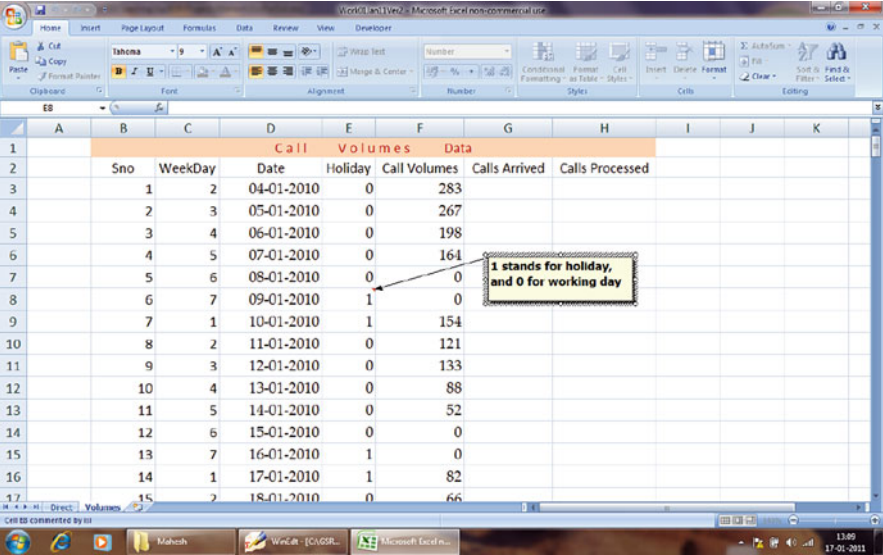


Fig. 6.4 Snapshot of the screen “Volumes” of the software tool

arrivals of the working days thus calculated, say r_i , against the working day (W-day) number i are presented in Table 6.1 for the 8 months, January to August.

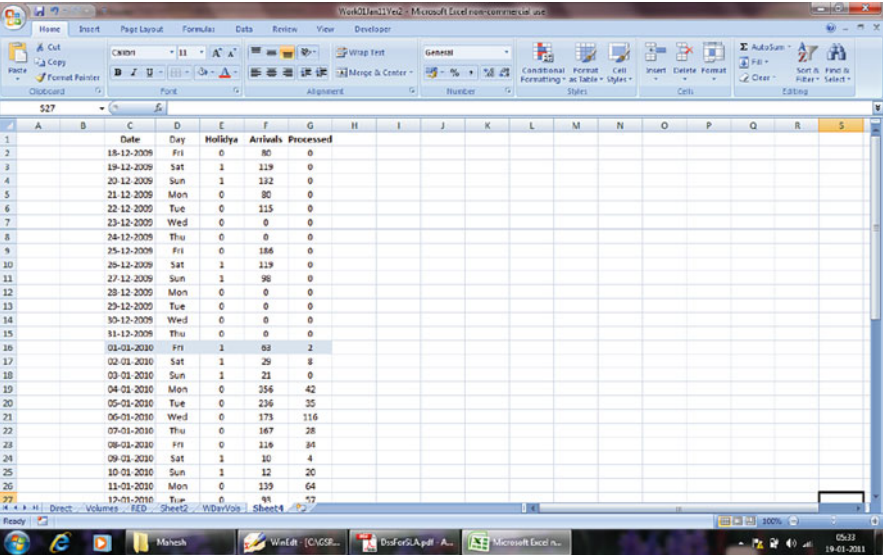


Fig. 6.5 Live data in excel

Table 6.1 Working day number wise call arrivals

W-day(<i>i</i>)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	\bar{r}_i	$\hat{\sigma}_i$
1	469	438	402	387	567	425	428	469	448	56
2	236	240	287	210	262	255	242	245	247	22
3	173	184	200	226	188	222	80	157	179	46
4	167	162	165	203	148	173	212	215	181	26
5	116	145	154	134	145	221	158	115	149	33
6	161	154	142	88	160	103	143	134	136	27
7	93	114	123	91	110	88	81	87	98	15
8	61	61	85	98	87	93	106	75	83	16
9	80	60	66	58	72	66	75	63	68	8
10	59	77	53	56	52	75	66	51	61	10
11	60	82	57	47	71	44	33	67	58	16
12	52	55	60	55	48	57	31	42	50	10
13	43	55	42	68	33	39	52	38	46	11
14	34	39	29	45	39	46	36	35	38	6
15	39	33	24	30	38	38	41	29	34	6
16	49	49	42	37	40	42	27	35	40	7
17	34	33	22	23	47	25	39	33	32	9
18	16	30	39	64	21	17	44	10	30	18
19	30	36	27	39	28	25	45	16	31	9
20	30	36	26	27	25	28	26	3	25	10
21			39	40	22	36	26	2	28	14
22			36	90		59	27		53	28
23			64						64	

Exhibit: Correlation matrix of the monthly volumes of different months

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Jan	1.00							
Feb	0.99	1.00						
Mar	0.97	0.98	1.00					
Apr	0.95	0.95	0.96	1.00				
May	0.99	0.99	0.96	0.94	1.00			
Jun	0.95	0.96	0.97	0.97	0.95	1.00		
Jul	0.95	0.94	0.94	0.92	0.95	0.93	1.00	
Aug	0.99	0.98	0.97	0.96	0.98	0.95	0.97	1.00

There is a cyclic pattern in the call arrivals. This is very clearly evident from Fig. 6.6 that plots the calls of the 8 months against the working day number. Treating each months' volumes as a multidimensional vector and taking the first 20 working days' volumes (so that each vector is in the same dimension), we can examine the dimension of the vectors by looking at their correlations as well as their principal components. The first principal component accounts for 96.7 % of the total variation. The correlation matrix is shown in the exhibit. The high correlation and high contribution by the first eigen value indicate that the call volumes are have a stable pattern. Also the average volumes of the 6 months remain same. This can be examined by carrying out a two-way analysis of variance [4] on the number of call

arrivals by taking working day number as one of the factors and Month as the other factor. The summary of the analysis is shown in Fig. 6.7. This analysis reveals that month to month variation is insignificant and that Working Day Number is the only significant factor.

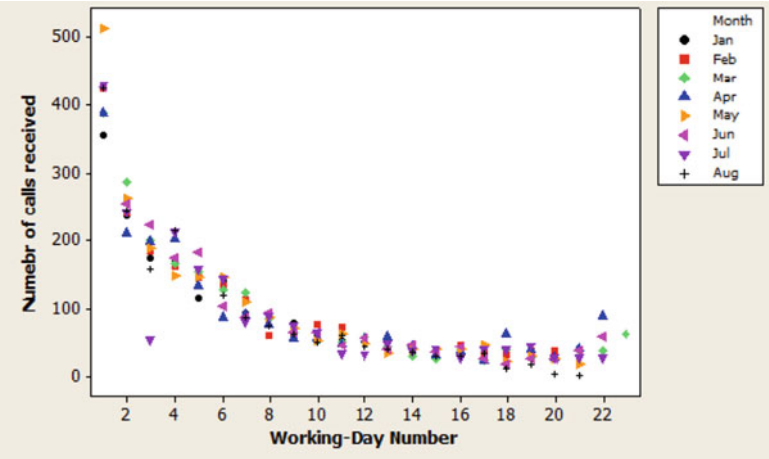


Fig. 6.6 Working day-wise call volumes

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Working Day	22	1647988	1646761	74853	148.17	0.000
Mon	7	2667	2667	381	0.75	0.627
Error	141	71230	71230	505		
Total	170	1721885				
S = 22.4761 R-Sq = 95.86% R-Sq(adj) = 95.01%						

Fig. 6.7 Analysis of variance

In the light of the above observations, it sounds reasonable to use the average number of calls of a working day, say i th working day, computed from the 8 months data as the expected number of call arrivals, r_i , for any month. Denote this expected number by \bar{r}_i . The values of \bar{r}_i and the sample standard deviations ($\hat{\sigma}_i$) of r_i s are summarized in the last two columns of Table 6.1. The utility of the software tool will now be demonstrated.

6.7.1 Plan for Sep’2010

First consider the case of planning for September 2010. In this case FPD is September 01, 2010 and LPD is September 30, 2010. The user must ensure that the sheet “Volumes” contains the data for August 2010 and September 2010. The call volumes for August should be the pending calls as on 1st September, and the call volumes for September are the arrivals forecasted (r_i s). In this case, the call volumes for August will all be zero as there are no pending calls as on 1st September. After updating the volumes sheet, the user should enter the FPD, LPD and the TAT (=10 for this category of calls) in the sheet “Direct” and click on “Load Data” button. This will load the data and the resulting sheet is shown in Fig. 6.8.

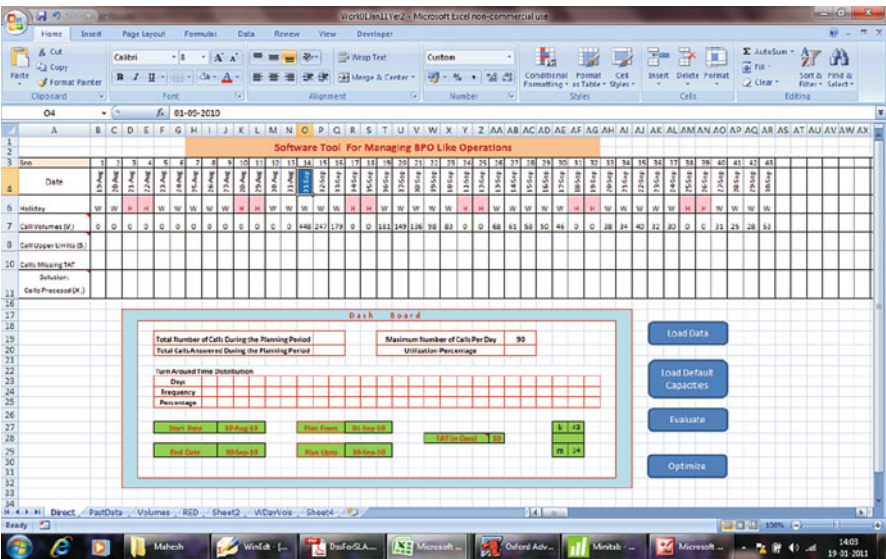


Fig. 6.8 Sheet “Direct” after loading data

6.7.2 The Company’s Solution

Next, to examine the decision of deploying three resource persons, that is 120 calls per day planned, the user should enter 120 against Maximum Number of Calls Per Day, click on “Load Default Capacities” button, and then click on “Evaluate” button. The resulting sheet is shown in Fig. 6.9. There will be no pending calls at the end of the month, the TAT distribution shows that all calls will be addressed within at most 7 working days, and the utilization is 59 %.

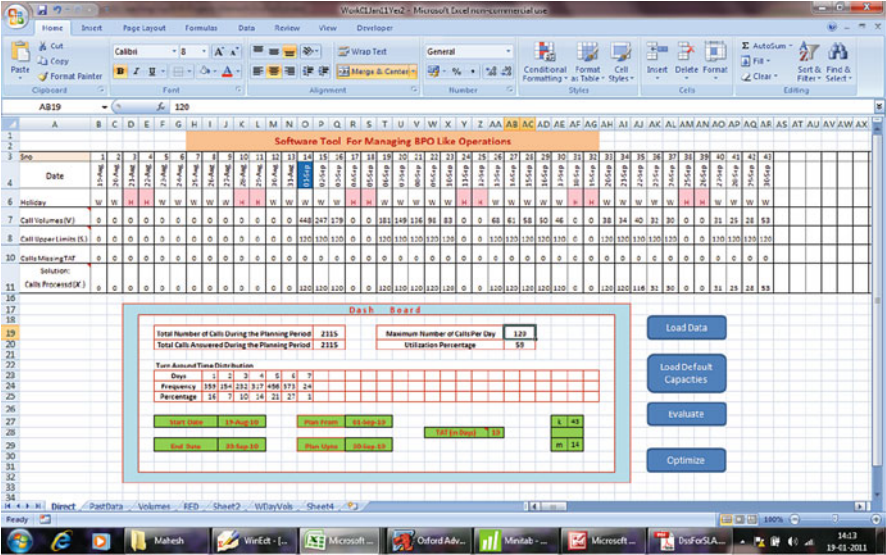


Fig. 6.9 Effect of deploying three resource persons

6.7.3 Stage-1 Optimal Solution

To find out the minimum number of resource persons required for September, the user should click on the “Optimize” button. Under the optimal solution, the maximum number of calls to be addressed on any day is 90, the TAT goes up to 10 days and the utilization goes up to 73 %. The optimal solution here will have all x_i corresponding to working days equal to 90. This means that instead of deploying three resource persons (equivalent to 120 calls), it is possible to manage with two resource persons with a little overtime provision.

6.7.4 Stage-2 Optimal Solution

From Stage-1 solution, it is known that the maximum of x_i is 90. With this information the Stage-2 problem is solved. In the resulting solution, some of the x_i s get reduced ($x_{14} = 88, x_{36} = 83, x_{37} = 69, x_{40} = 62, x_{41} = 61, x_{42} = 50, x_{43} = 41$, and other x_i s corresponding to working days are equal to zero). The final optimal solution is displayed in Fig. 6.10. After Stage-2 optimization, the utilization percentage reduces from 73 % to 67 %. However, one should be cautious in interpreting the results of Stage-1 to Stage-2 optimizations. The Stage-2 optimization determines how much more spare time, that is, the savings in terms of number of calls, can be extracted from the optimal solution of Stage-1. This can be worked out as $\frac{100(u_1 - u_2)}{u_1}$ where u_i is the utilization percentage of i th stage optimization, $i = 1, 2$. For this example it is equal to 8 % ($= \frac{100(73-67)}{73}$).

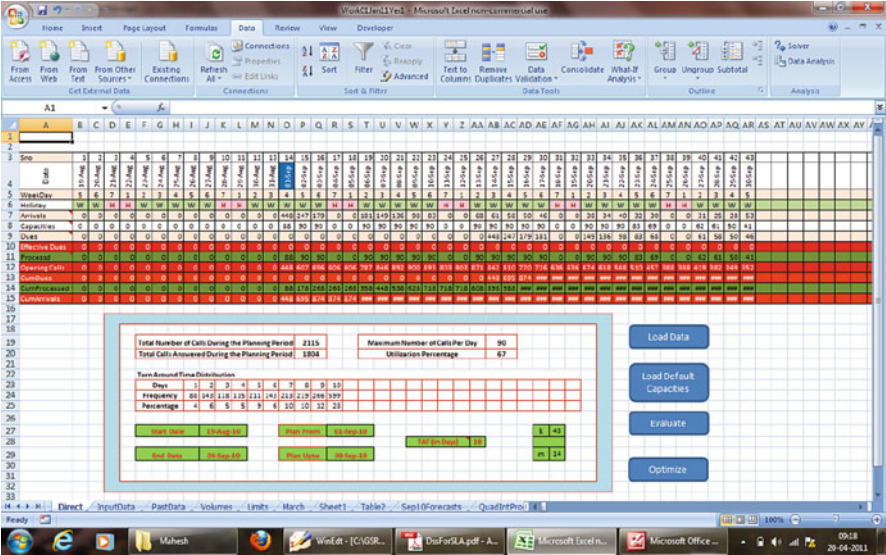


Fig. 6.10 Optimal resources required for September 2010

6.7.5 What if the TAT Is Reduced to 8 Days?

This question can be examined by running the tool by modifying TAT to 8. The user has to Load Data and then Optimize. Under the optimal solution, the maximum number of calls to be addressed on any day is 104, and the utilization is 68 %.

6.7.6 Dynamic Review

Imagine that the manager has decided to go for addressing a maximum of 100 calls on any day. Suppose the actual arrivals on 1st September is 523 as opposed to the forecasted 448. The manager wants to understand the impact of this deviation in the arrival. This can be done by replacing 448 (v_{12}) by 523 on 1st September, and evaluating the solution. The resulting solution will address all the calls by the end of the month but 21 % of the calls will be addressed on the 9th working day and 1 % of the calls on the 10th working day. The utilization is 73 %.

6.7.7 What if Resources Are Restricted on Some Days?

Imagine that the manager has planned for 100 calls per day and some of the resource persons will not be available on 6th and 9th of September. Based on this, the

manager can plan for at most 50 calls on 6th and at most 70 calls on 9th. How does the solution get affected? The user has to change the Call Upper Limits of 6th and 9th to 50 and 70 respectively and then evaluate. The solution implies that it is still possible to complete all the calls by month end with a TAT of 10 days and utilization will be 71 %.

6.7.8 Analysis of Past Data

It was observed that the call arrival pattern is not influenced by month. Therefore, the forecasts used for the month of September can as well be used for any of the past months. It may be interesting to examine, what would have been the situation had this tool been applied to all the past months from January to August. In this case, four types of solutions are examined. These are labeled as “Solution-OF,” “Solution-OA,” “Solution-QA” and “Solution-CA.” The description of each of these solutions is given below.

Solution-OF: This solution is obtained by using forecasted volumes and the min-max criteria explained earlier.

Solution-OA: This solution is obtained by using the actual arrivals and the Maximum Number of Calls Per Day obtained by the Solution-OF.

Solution-MA: This solution is obtained by using the actual arrivals and the average of the calls addressed during working days.

Solution-CA: This is the solution actually implemented by the company for the actual arrivals.

Before describing the results of the four types of solutions, it is necessary to elaborate on the calls answered on a normal day and the estimation of Q_3 , the third quartile.

6.7.9 Calls Answered on a Normal Day

From the past data it is observed that there is lot of variation in the calls addressed each day. In fact, it is noticed from the data that calls were answered on many holidays. It is found that due to logging problems, there could be a mix up of calls that fall on a working day which is either preceded or succeeded by a holiday. In order to assess the number of calls that can be addressed on a working day, it is felt reasonable to use only the calls pertaining to **normal working days**, where a day is considered normal if it is preceded and succeeded by working days. Over the 8 months data, there are 102 normal working days and the data on the calls addressed on these days had two outliers. A graphical summary of the the data (excluding the two outliers) is shown in Fig. 6.11. An estimate of the average number of calls per day is 92. The four types of solutions mentioned above are obtained for

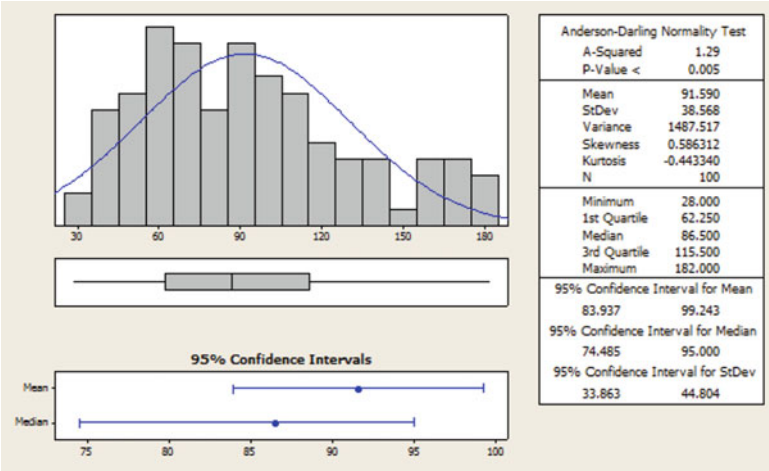


Fig. 6.11 Summary of number of calls addressed on normal working days

each of the months. With the help of the software tool this is just a matter of clicking the buttons and the solutions can be obtained in no time. To examine the four solutions and compare them, the results are summarized in Fig. 6.12. The summary comprises the following. The first column specifies the month and the figures in the parentheses specify the *Chi-square statistic* ($= \sum \frac{(r_i - v_i)^2}{v_i}$) for the month's forecasts (here r_i s are the actuals and v_i s are the forecasts). The second column presents the maximum number of calls on a single day addressed during the month. The third column presents the percentage of calls addressed against the expected/available calls during the month. The fourth column presents the utilization percentage and the rest of the columns present the TAT distribution.

As per the current planning of the company (which was applied to all the 8 months), 5.7 resource persons are assigned but only three resource persons (at most) are deployed for the calls of the category in question. Each resource person is expected to address 40 calls per day. From the results of Fig. 6.12, optimal solutions based on forecasted volumes recommended at most 120 calls for all the 8 months with just one exception (126 calls in May Solution-OA). Further, the same solutions work even under the actual calls received. This means that if the optimal solutions are implemented, then they are robust against the forecasting errors. The optimal solutions suggest that three resource persons would be good enough, as opposed to the 5.7 of the company's plan, for handling the calls without breaking TAT conditions. The percentage of available calls addressed is high (93 or above with just one exception). If one plans for the average number of calls per day (92), then the TAT condition is violated in all the 8 months. The number of actual calls addressed each day has more variation than desired. If this is better controlled and the average is taken closer to the number suggested by the optimal solution, then there is

Month	Company	Max. Calls	Answered Calls (%)	Utilize (%)	TAT Distribution Percentages														
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Jan (85%)	Solution-OF	94	89	68	1	8	8	4	7	6	10	13	19	18					
	Solution-OA	93	90	68	0	8	10	4	6	8	9	14	23	14					
	Solution-MA	92	89	68	0	8	10	4	5	8	8	12	22	18	0				
	Solution-CA	170	89	68	0	0	2	4	7	3	3	11	26	39					
Feb (40)	Solution-OF	105	93	71	0	0	5	6	5	13	12	17	20	18					
	Solution-OA	105	91	71	0	0	5	6	5	13	12	20	25	9					
	Solution-MA	92	80	71	0	0	3	8	5	8	12	7	10	11	19	13			
	Solution-CA	177	88	79	0	0	0	0	9	3	12	22	21	30					
Mar (50)	Solution-OF	116	100	68	5	1	4	9	8	12	13	11	15	17					
	Solution-OA	116	100	68	4	2	4	9	9	12	12	11	14	18					
	Solution-MA	92	83	71	0	0	0	6	7	7	11	7	7	6	4	8	19	12	
	Solution-CA	183	88	76	0	0	0	6	6	4	5	25	8	20	22				
Apr (141)	Solution-OF	109	100	73	1	1	3	10	8	9	13	14	20	17					
	Solution-OA	104	95	73	0	0	1	11	9	10	15	14	18	18					
	Solution-MA	92	84	73	0	0	0	5	8	6	10	9	10	13	17	18			
	Solution-CA	221	81	62	0	0	0	0	3	16	11	9	11	24	21	0			
May (65)	Solution-OF	117	95	65	0	0	0	5	9	9	14	15	24	21					
	Solution-OA	126	97	65	0	0	0	7	9	9	12	13	26	21					
	Solution-MA	92	71	65	0	0	0	0	7	9	9	7	7	10	7	8	10	13	7
	Solution-CA	182	93	85	0	0	0	3	9	10	9	21	42	3					
Jun (80)	Solution-OF	107	99	73	1	1	4	7	8	10	13	16	20	15					
	Solution-OA	111	100	73	2	2	4	6	8	11	14	14	20	14					
	Solution-MA	92	83	73	0	0	0	5	5	6	10	9	10	7	9	11	22		
	Solution-CA	156	92	81	0	0	0	0	5	6	11	29	18	29					
Jul (124)	Solution-OF	103	98	71	0	1	8	8	8	10	10	14	18	18					
	Solution-OA	98	97	71	0	0	7	9	9	9	12	16	15	18					
	Solution-MA	92	91	71	0	0	5	7	6	6	11	13	13	15	20				
	Solution-CA	205	92	72	0	0	0	0	0	10	13	12	22	37	3				
Aug (144)	Solution-OF	102	98	71	0	2	6	7	10	12	10	14	19	17					
	Solution-OA	101	100	67	0	1	6	6	10	11	10	15	17	18					
	Solution-MA	92	96	71	0	0	5	5	7	7	10	7	13	12	23	4			
	Solution-CA	179	100	74	0	0	0	6	13	13	21	27	16						

Fig. 6.12 Summary of results of past data analysis

scope for clearing most of the available calls of each month within the same month. The utilization percentages are around 65–75. The deviations between the forecasts and the actual arrivals are measured by the Chi-square statistic. Since the number of working days is approximately 21 days in each month, the Chi-square statistic may be compared with the tabulated chi-square which is equal to 33 for 21 degrees of freedom. Though the deviations are high, the solutions suggest that if the company plans for three resource persons, then the calls can be addressed efficiently.

6.8 Summary

The main activity of a ITES/BPO industry is to receive transactions/calls from the client base and address them within TAT, the stipulated time. One of the major activities of the industry is that of planning and management of resources, the agents/associates who address the calls from the clients. The problem is too complex to get solutions manually based on experience and human intelligence, and calls for a decision support system which will be extremely useful and inevitable for efficient management of the operations. In the absence of the use of DSSs, the industries are facing a common challenge of increasing gap between demand and supply due to inefficient management. This work has evolved from one such industry, an ITES company that is struggling in planning and management of resources.

This chapter presents the details of development and utility of an Excel based software tool that will be extremely useful in planning and management of human resources. It describes how the problem is viewed from a scientific angle and the formulation of the problem as a integer-linear programming problem. The tool is applied to a live example from the ITES company which has sought solution to the problem.

A number of situations that a manager encounters while planning and managing the resources and how they can be handled using the software tool are discussed in the chapter. A forecasting model is proposed for one specific category of calls along with a chisquare metric to indicate the extent of agreement between the forecasts and the actual calls received. Similarly, a metric to indicate the utilization of resources is also proposed. The effectiveness of deploying the software tool is demonstrated by applying the tool on past data and comparing solutions of different types including the one actually executed in the past.

Lastly and most importantly, the main objective of this work is to promote business excellence through scientific based and low cost solutions among industries where operations are similar to the ones for which the software tool is suitable.

Acknowledgements This study was carried out for a multinational company based in Hyderabad. Mr. Mahesh Giri, a senior executive from this company, initiated this study as a part of his course requirement at the Indian Statistical Institute, Hyderabad. He played a very active role in undertaking and carrying out this study, development of the software and getting it implemented in the company.

References

1. Aksin, Z., Armony, M., Mehrotra, V.: The modern call center. *Prod. Oper. Manag.* **16**(6), 665–688 (2007)
2. Baker, K.R.: Workforce allocation in cyclical scheduling problems: a survey. *J. Oper. Res. Soc.* **27**, 155–167 (1976)

3. Bhulai, S., Koole, G., Pot, A.: Simple methods for shift scheduling in multi-skill call centers. *Manuf. Serv. Oper. Manag.* **10**(3), 411–420 (2007). <http://www.aukepot.com/pdf/ShiftAppendix.pdf>
4. Bogartz, R.S.: *An Introduction to the Analysis of Variance*. Praeger, Westport (1994)
5. Brucker, P., Qu, R., Burke, E.: *Personnel scheduling: models and complexity* (2009). <http://www.cs.nott.ac.uk/~ekb/Publications/pers-sched.pdf>
6. Heyman, D.P., Sobel, M.J.: *Stochastic Models in Operations Research*, vol. II. Dover Publications, Mineola (2004)
7. Power, D.J.: What is a DSS? *On-Line Executive J. Data-Intensive Decis. Support* **1**(3) (1997). <http://www.taborcommunications.com/dsstar/971021/100015.html>
8. Power, D.J.: A brief history of decision support systems. <http://dssresources.com/history/dsshistory.html>, version 4.0 (2007)
9. Robbins, T.R., Medeiros, D.J.: Does the Erlang C model fit in real call Centers? In: Johansson, B., Jain, S., Montoya-Torres, J. Hugu, J., Yücesan, E. (eds.) *Proceedings of the 2010 Winter Simulation Conference*, Baltimore

Chapter 7

Intelligent Modeling and Practical Solutions

Abstract Industries often approach consultants with specific questions or problems. There is a great deal of demand for consultants in this regard. Some problems, though straightforward in terms of formulation, need special techniques for solving them. This chapter presents five live cases of industrial problems, some requiring special solutions and some requiring statistical modeling. One case deals with an important problem of determining the possible frequencies of amino acid molecules in a synthesized polypeptide – a typical integer linear programming problem application but requires an efficient solution procedure. Two of the cases are applications of multivariate analysis techniques – one dealing with comparing a company's product with a reference product, and the other is about a procedure for rating a product based on multiple characteristics. Another case deals with the assessment of reliability of CFL bulbs. Besides estimating the reliability of bulbs, this case uses Markov chains and simulation to determine the number of test benches required for testing the bulbs. The last case of this chapter deals with evaluating two insurance schemes, again a nice application of Markov chains.

7.1 Introduction

Industries often approach consultants with specific problems or questions. Some problems require routine modeling or analysis techniques but some require special solutions as the routine techniques are either inadequate or inefficient to solve them. This chapter presents some live cases where the industry has approached the author seeking solutions. It is hoped that the cases presented in this chapter will be useful to the readers in modeling and solving problems of similar nature.

This chapter has five sections. Each section presents a live case of decision making problem. The first two cases are from a pharmaceutical industry. The first of these two cases is an important problem of determining the possible frequencies

amino acid molecules in a synthesized polypeptide prepared by the company. This is a typical integer programming problem but requires an efficient solution procedure. The ideas of this case study can be applied to some other interesting problems. This study is presented in Sect. 7.2.

The second case deals with determining whether the behaviour of one of the products made by the company is different from that of a standard reference product. This is a typical application of multivariate analysis techniques such as principal component analysis and cluster analysis. This study is presented in Sect. 7.3.

The third case is from a bulbs manufacturing company and it is about the reliability of bulbs and planning testing facilities. The company was told by a consultant that the reliability of their bulbs was very low which became a major concern for them. Hence the company wanted a proper assessment of the reliability of their bulbs. This study is presented in Sect. 7.4.

The fourth case is from a paper industry and it is about deriving a procedure for rating a product, paperboard in this case, based on multiple characteristics of the product. This case also uses multivariate techniques for the solution and it is presented in Sect. 7.5.

The fifth case is about evaluating two insurance schemes. The client in this case is actually an individual who wanted a scientific solution to his decision making problem. The problem is that he has the option of choosing one of two possible pension schemes and he wanted to know which scheme should he chose. This study is presented because of its interesting solution which is a nice application of Markov chains. This study is presented in Sect. 7.6.

7.2 Determination of Polypeptide Composition

This study presents a case on how an important practical problem at a pharma industry was solved. Essentially the problem is a simple integer linear programming problem [1, 12] but the requirement is that of obtaining all solutions. In the process, the question of uniqueness of solutions to the problem is encountered. The approach to addressing this problem has lead to some interesting ideas. New formulations are devised to determine the uniqueness and to obtain multiple solutions when the solution is not unique. It is pointed that the ideas could be extended to general problems.

7.2.1 Background

The company is working on a polypeptide synthesised with four amino acids. A peptide is a molecule consisting of a chain of two or more amino acids. Depending on the number of amino acids in the chain, peptides are called dipeptides, tripeptides, tetrapeptides, and so on. The company is working on a polypeptide synthesized with four amino acids. By hydrolyzing the random polypeptide with

selective enzymes, polypeptide chain is cleaved into small fragments. The molecular weights of the cleaved fragments are determined by analyzing them using Liquid chromatography–mass spectrometry (LC-MS) (http://en.wikipedia.org/wiki/Liquid_chromatography-mass_spectrometry cited on Nov 06, 2014). The general problem of interest is to determine the sequence of amino acids in the chains of the fragments because knowing the primary structures of the fragments helps, sometimes, in deducing the original structure of the peptide by taking advantage of overlapping pieces. The actual problem posed by the company for this project is a subproblem of the above and is described in the following subsection. For more information on the polypeptides and the problems related to it see the sites http://en.wikipedia.org/wiki/Peptide_bond, http://en.wikipedia.org/wiki/Peptide_synthesis and <https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/protein2.htm>.

7.2.2 Statement of the Problem

The *random polypeptide* is prepared by synthesizing with four acids, namely, L-glutamic acid (*G*), L-alanine (*A*), L-tyrosine (*T*), and L-lysine (*S*) with the mole fraction of 0.141, 0.427, 0.095, and 0.337 respectively. In a layman's understanding this simply means that in the polypeptide that is finally prepared, the proportions of the four acids *G*, *A*, *T* and *S* are 0.141, 0.427, 0.095, and 0.337 respectively (note that the sum of these four numbers is 1). The polypeptide is then cooked/digested which results in the formation of sequences of the four acids. Each sequence may be regarded as a *DNA sequence* of the four acids. After the process of cooking/digestion, the molecular weight of the total substance prepared is in the range of 2000–20000 *daltons*. In the LC-MS analysis, each of the sequences is cut into smaller size subsequences of the four acids. The subsequences are called fragments. Now, each fragment has a certain number of molecules and the amino acids (Fig. 7.1) are in a particular sequence in the fragment. For example, a fragment has 4 *G*s, 2 *A*s, 3 *T*s and 2 *S*s, and the arrangement of the acids in this fragment may be *SGAGTAGGTTS*. It is also possible that another fragment has the same composition of the four acids but the sequence is *SAAGGTTTGGs*. The weight of each fragment is obtained in the LC-MS analysis and theoretically it should satisfy the following equation:

$$147.05x_g + 89.05x_a + 181.07x_t + 146.11x_s - 18.026(x_g + x_a + x_t + x_s - 1) = w, \quad (7.1)$$

where x_g , x_a , x_t , x_s are the number of *G*, *A*, *T* and *S* acid molecules respectively; w is the atomic weight of the total fragment. The coefficients of x_g , x_a , x_t and x_s are the molecular weights of the respective amino acids. The component $18.026(x_g + x_a + x_t + x_s - 1)$ in the equation is the weight of water molecules that are sandwiched between acid molecules in the sequence. Normally the weight w is in the range of 500–2000. The problem posed by the company is: Given w , find the composition

of the fragment and the possible sequences of the given fragment. In other words, find all possible solutions (x_g, x_a, x_t, x_s) satisfying Eq. (7.1) and list all possible sequences for the given weight w . The client wants a ready reckoner table that can be used for quick reference or a software tool that will throw up the solution by taking the weight as the input.

Another important question that the company had in mind, though not studied under this project, is whether the proportions of amino acids in the original mix influences the composition of the amino acids in the fragments. In other words, if (x_g, x_a, x_t, x_s) is treated as a random vector in a random fragment, does the distribution of (x_g, x_a, x_t, x_s) depend on the mixture proportions?

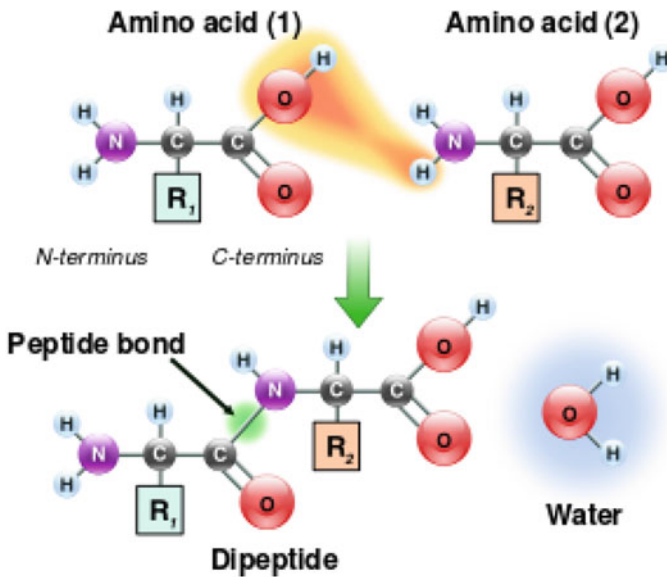


Fig. 7.1 Amino acids

7.2.3 Approach and Evolution of Ideas

The company's problem is that of finding x_g , x_a , x_t , and x_s for an observed w in the range 500–2000, and then list down all possible sequences for the possible composition. It must be noted that if Eq. (7.1) has no integer solution for a given w , then it means that w is not a possible observation. Since the problem is to prepare a ready reckoner, we must treat w also as a variable and solve the following integer programming problem: Find positive real w and nonnegative integers x_g , x_a , x_t and x_s such that

$$\begin{aligned}
w - 129.024x_g - 71.024x_a - 163.044x_t - 128.084x_s &= 18.026 \\
w &\geq 500 \\
w &\leq 2500
\end{aligned} \tag{7.2}$$

$w \geq 0$ is real and x_g, x_a, x_t, x_s are nonnegative integers.

Since the company wants all the solutions, we must have a procedure to determine all the solutions. One way of getting all solution is described below.

First solve (7.2) and get a feasible w . Having got the feasible w , get all possible solutions to (7.1) with this w fixed. Since the company is also interested knowing whether a solution is unique, we must first find out for a given feasible w whether the solution is unique. One way of addressing the uniqueness question is the following. Though simple, the idea deserves a mention here.

7.2.4 How to Address the Question of Uniqueness?

We can manipulate with the integer programming formulations to address the question of uniqueness and also ways to find out all solutions. In the process we use a formulation that will find out a situation where the solution is not unique. It is worth presenting these manipulations as they give interesting ideas. For a feasible w , the solution to (7.1) is unique if, and only if, the optimum objective value of the following problem is zero.

$$\begin{aligned}
&\text{Maximize } \sum_{i=1}^4 |u_i - v_i| \\
&\text{subject to} \\
&\quad g(u) = 18.026 \\
&\quad g(v) = 18.026 \\
&\quad u \text{ and } v \text{ are nonnegative integer vectors in } R^4, \\
&\quad \text{where } g(x) = w - 129.024x_g - 71.024x_a - 163.044x_t - 128.084x_s, \\
&\quad \text{and } x = (x_g, x_a, x_t, x_s).
\end{aligned} \tag{7.3}$$

Though the objective function involves absolute values, it can be linearized using the representation of a real number r as a difference of two nonnegative reals ($r = r^+ - r^-$, $r^+ \geq 0$, $r^- \geq 0$) and the formula $|r| = r^+ + r^-$.

Suppose (7.1) has multiple solutions. How to get all the solutions? Again, we can formulate this as a integer programming problem. Suppose u is a solution to (7.3) already obtained. To get another solution, solve the following problem:

$$\begin{aligned}
&\text{Find nonnegative integer vector } v \in R^4 \text{ satisfying} \\
&\text{subject to} \\
&\quad g(v) = 18.026 \\
&\quad \sum_{i=1}^4 |v_i - u_i| \geq 1
\end{aligned} \tag{7.4}$$

The above idea can be extended to obtain one more solution different from two already obtained solutions, say, u and v . We just have to add one more constraint to the above problem with respect to the second solution. This procedure can be continued until we get all the solutions. However, this will be an inefficient procedure. We will see an elegant solution to the problem of finding all solutions.

It turned out that for the client's problem, there are more than 11000 values of w in the range of 500 to 2500 so that (7.1) has a feasible solution. In fact, for each of these values of w , the solution is unique. Again, one might be curious to know how we arrive at these conclusions. The answer to this problem is the elegant solution to the client's problem presented in Sect. 7.2.6.

One more question that arises, out of curiosity, in connection with this problem is the following: Is there a w for which Eq. (7.1) has at least two solutions, and if so, what is smallest such w ? The formulation below was devised to find an answer to this question.

$$\begin{aligned}
 &\text{Minimize } w \\
 &\text{subject to} \\
 &\quad g(u) = 18.026 \\
 &\quad g(v) = 18.026 \\
 &\quad \sum_{i=1}^4 |u_i - v_i| \geq 1 \\
 &\quad u \text{ and } v \text{ are nonnegative integer vectors in } R^4,
 \end{aligned} \tag{7.5}$$

This problem was solved using a professional OR package, and it took more than 15 minutes of CPU time and yet the optimum solution could not be obtained. However, during the run time, the smallest w with at least two solutions was found to be 3883.08. This might provide a useful information to the client.

7.2.5 Extension of Ideas to General Problems

The above ideas can be extended to more general problems. For example, if we are interested in knowing whether a system of equations $f(x) = 0$, where f is a function from R^n to R^m , has a unique solution, examine the solution to the problem given by

$$\begin{aligned}
 &\text{Maximize } \sum_{i=1}^n |u_i - v_i| \\
 &\text{subject to} \\
 &\quad f(u) = 0 \\
 &\quad f(v) = 0
 \end{aligned} \tag{7.6}$$

Solution is unique (when it exists) if, and only if, the optimum objective value to (7.6) is zero.

Next consider the optimization problem:

$$\begin{aligned} &\text{Minimize } f(x) \\ &\text{subject to} \\ &g(x) = 0 \end{aligned} \tag{7.7}$$

If \bar{x} is a solution to this problem, then the problem has a unique solution if, and only if, the problem (7.8) below has zero as its optimum objective value.

$$\begin{aligned} &\text{Maximize } \sum_{i=1}^n |x_i - \bar{x}_i| \\ &\text{subject to} \\ &g(x) = 0, \\ &f(x) = f(\bar{x}) \end{aligned} \tag{7.8}$$

7.2.6 Solution to the Client's Problem

In the previous subsection, we have seen some theoretical ideas for solving the client's problem. But those ideas involve solving a large number of integer linear programming problems [2, 12, 14] which are difficult to solve and time consuming. We must explore different alternatives to solving the problem. Since the client problem has only four variables, it is worthwhile trying computer programming techniques to solve the problem. In fact, in this case, a simple program involving four nested *for loops* will provide a smart solution to the problem. The programming logic for the solution is explained in the next paragraph.

Since $w \leq 2500$, the maximum possible value of x_g is equal to integral part of $\frac{2518.026}{129.024}$ plus 1 which is 20 in this case. Similarly, the maximum possible values of x_a , x_t and x_s are 36, 16 and 20 respectively. Now the trick is to enumerate all four tuples (x_g, x_a, x_t, x_s) and compute the corresponding w . If a four tuple has its w within the range 500 to 2500, then record the weight and its tuple, otherwise discard it. Now, the question of uniqueness can be easily answered by finding out the frequencies of the recorded w s. If the frequency f_w of a recorded w is more than one, then the problem corresponding to that w will have multiple solutions, in fact, f_w solutions to be precise. It turned out that $f_w = 1$ for all w in the range 500 to 2500. The total number of 4-tuples scanned in the for loop is equal to 230400 of which 11945 have their w s in the range 500–2500. From the computer program, the ready reckoner is easily prepared. The program was written using Microsoft Excel macro and the reckoner is also stored in the same Excel file. The essential part of the programme is outlined below for better understanding of the reader.

```
k = 2
For xg = 0 To 20
  For xa = 0 To 36
    For xt = 0 To 16
      For xs = 0 To 20
```

```

w = 129.024 * xg + 71.024 * xa
    + 163.044 * xt + 128.084 * xs + 18.026;
If (w >= 500 And w <= 2500) Then
    Print (w, xg, xa, xt, xs);
    k = k + 1;
End If
Next
Next
Next
Next

```

7.2.7 Conclusions of the Peptide Study

This section presented an important case study for a pharma industry. The information that the solutions will be unique for all the fragments encountered (in the range 500–2500 of w) is important and useful to the company. The solution procedure demonstrates that sometimes complex problems have simple solutions. Solving the problem and providing solution has actually lead to some interesting ideas about some typical problems that are often encountered in theory and the ideas may be useful elsewhere. One such problem encountered here is that of uniqueness of solutions to a system of equations. This has provided an opportunity to the analyst to come up with new and interesting ideas. Finally, it turned out that the theoretical formulations of the practical problem may not always be effective in terms of providing quick solutions. Therefore, it is important that the analysts should explore the problem and solution from various angles. In this case, a simple computer program could provide a very effective solution to the problem. The most important contribution of this work is that it has provided an effective solution to a live problem from industry and in the process some interesting theoretical ideas evolved.

7.3 Comparing Company's Product with Reference Product

This study was carried out for a pharma industry. The problem was about establishing competence of the company in producing one of their products. The product is produced in batches. One of the requirements in establishing the competence is that of establishing batch consistency. In this regard, this study was taken up to compare the consistency of the reference product, also known as innovator's product, over different batches with that of the company's product. The comparison is to be made with respect to the markers of a Mass Spectrum (MS) analysis.

For this study, some background of multivariate analysis is required. The reader may refer to [3, 6, 10, 11].

Twenty-five markers, the variables in the analysis coded as x_1 to x_{25} , have been identified for the purpose of comparison. Data are provided for the two products – the reference product (RP) and the company’s product (CP) – on these variables. There are 39 data points (25-tuple vectors) for RP and 6 data points for CP. Each data point corresponds to a batch. The objective of the project is to see if there are any significant differences between RP and CP. The data are analyzed using multivariate methods [3, 6, 10, 11].

7.3.1 Analysis of the Reference Product Data of 39 Observations

The summary statistics of the original variables are given in Table 7.1. It may be noted that the means and variances of the 25 components (variables) are varying significantly. This is evident from the Dot plot shown in Fig. 7.2. Some visible outliers from the Dot plot are: batch numbers 31, 32, 33, 38 and 39 (accompanied with numbers in the Dot plot). The correlation matrix of the variables is given in Table 7.2. It may be noted that many of the variables are highly correlated. This means that the variables have significant interdependence.

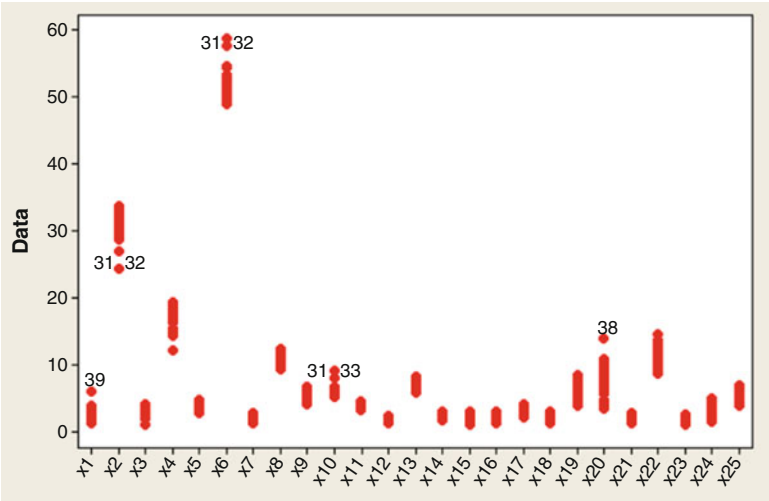


Fig. 7.2 Dot plot of the reference product data. The ordinate represents the marker values

Table 7.1 Basic statistics of the data on reference product

Variable	Mean	St dev	Variance	Min	Q_1	Median	Q_3	Max
x_1	2.05	0.86	0.74	1.35	1.67	1.82	1.96	6.01
x_2	31.11	1.79	3.21	24.26	30.48	31.39	32.17	33.63
x_3	2.75	0.67	0.45	1.09	2.27	2.70	3.12	4.17
x_4	17.10	1.52	2.31	12.10	16.51	17.27	18.23	19.37
x_5	3.33	0.48	0.23	2.77	3.02	3.23	3.45	4.81
x_6	51.73	2.12	4.49	48.91	50.16	51.16	52.81	58.80
x_7	1.73	0.32	0.10	1.21	1.56	1.70	1.82	2.75
x_8	10.91	0.74	0.54	9.38	10.44	10.89	11.37	12.40
x_9	5.20	0.55	0.31	4.10	4.79	5.23	5.50	6.68
x_{10}	5.98	0.92	0.85	5.09	5.42	5.81	6.12	9.18
x_{11}	3.95	0.29	0.08	3.27	3.73	3.94	4.14	4.56
x_{12}	1.73	0.26	0.07	1.33	1.52	1.75	1.86	2.41
x_{13}	6.80	0.54	0.30	5.92	6.40	6.70	7.00	8.13
x_{14}	1.85	0.32	0.10	1.59	1.65	1.76	1.83	2.99
x_{15}	1.53	0.49	0.24	1.13	1.28	1.35	1.41	3.00
x_{16}	1.58	0.38	0.14	1.24	1.39	1.48	1.61	3.04
x_{17}	3.39	0.37	0.13	2.23	3.22	3.43	3.64	4.03
x_{18}	1.63	0.47	0.22	1.18	1.35	1.50	1.60	3.01
x_{19}	6.02	1.21	1.45	3.95	5.02	5.92	6.59	8.48
x_{20}	7.82	2.45	5.98	3.40	6.54	8.09	9.53	13.98
x_{21}	1.68	0.34	0.12	1.34	1.50	1.54	1.74	2.70
x_{22}	11.35	1.26	1.60	8.73	10.43	11.20	12.16	14.59
x_{23}	1.30	0.40	0.16	1.01	1.10	1.15	1.23	2.64
x_{24}	2.93	0.86	0.74	1.46	2.13	2.95	3.52	5.04
x_{25}	4.85	0.64	0.41	3.92	4.49	4.72	4.96	6.81

Table 7.2 Correlation matrix of the variables for reference product

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	
1.00																									
x_2	-0.36	1.00																							
x_3	0.71	-0.43	1.00																						
x_4	0.16	-0.20	0.21	1.00																					
x_5	0.51	-0.41	0.65	-0.29	1.00																				
x_6	0.20	-0.71	0.24	-0.55	0.55	1.00																			
x_7	0.49	-0.49	0.69	-0.03	0.94	0.44	1.00																		
x_8	-0.28	0.32	-0.23	-0.34	-0.14	-0.02	-0.25	1.00																	
x_9	0.80	-0.36	0.75	0.44	0.55	-0.01	0.65	-0.39	1.00																
x_{10}	0.43	-0.56	0.58	-0.29	0.94	0.69	0.90	-0.14	0.46	1.00															
x_{11}	0.50	-0.07	0.26	0.01	-0.02	0.07	-0.10	0.32	0.23	-0.04	1.00														
x_{12}	0.49	-0.35	0.33	-0.36	0.72	0.55	0.61	-0.01	0.36	0.73	0.25	1.00													
x_{13}	0.59	-0.32	0.46	-0.33	0.59	0.52	0.46	0.18	0.43	0.64	0.53	0.71	1.00												
x_{14}	0.68	-0.57	0.60	-0.25	0.80	0.67	0.72	-0.02	0.52	0.84	0.37	0.80	0.83	1.00											
x_{15}	0.86	-0.40	0.79	-0.06	0.73	0.39	0.68	-0.14	0.75	0.66	0.44	0.70	0.75	0.83	1.00										
x_{16}	0.92	-0.38	0.73	0.10	0.55	0.26	0.53	-0.20	0.75	0.54	0.49	0.52	0.70	0.77	0.86	1.00									
x_{17}	0.23	0.14	0.06	0.11	-0.34	-0.19	-0.41	0.27	0.09	-0.31	0.56	-0.10	0.33	0.01	0.09	0.21	1.00								
x_{18}	0.73	-0.57	0.74	-0.17	0.92	0.61	0.89	-0.21	0.67	0.91	0.16	0.72	0.70	0.90	0.85	0.77	-0.17	1.00							
x_{19}	0.65	-0.10	0.60	0.04	0.49	0.07	0.49	-0.51	0.62	0.36	0.07	0.33	0.30	0.35	0.66	0.57	0.03	0.55	1.00						
x_{20}	0.48	-0.28	0.58	0.43	0.41	-0.07	0.57	-0.58	0.68	0.33	-0.10	0.10	0.01	0.18	0.39	0.44	-0.19	0.47	0.62	1.00					
x_{21}	0.87	-0.36	0.74	0.09	0.51	0.25	0.47	-0.06	0.76	0.47	0.62	0.53	0.76	0.71	0.90	0.88	0.35	0.69	0.54	0.33	1.00				
x_{22}	0.50	-0.15	0.38	0.06	0.09	0.10	0.06	0.17	0.46	0.11	0.67	0.29	0.65	0.43	0.55	0.57	0.60	0.28	0.21	0.04	0.71	1.00			
x_{23}	0.68	-0.58	0.70	-0.17	0.78	0.62	0.73	-0.04	0.58	0.79	0.35	0.78	0.81	0.91	0.92	0.74	0.02	0.87	0.47	0.26	0.79	0.48	1.00		
x_{24}	0.55	-0.31	0.48	0.29	0.30	0.05	0.38	-0.44	0.63	0.31	0.12	0.01	0.26	0.31	0.39	0.54	0.11	0.44	0.40	0.58	0.48	0.19	0.27	1.00	
x_{25}	0.67	0.08	0.43	-0.05	0.16	-0.03	0.05	0.05	0.50	0.06	0.58	0.32	0.58	0.41	0.67	0.65	0.48	0.33	0.42	0.08	0.77	0.74	0.48	0.26	1.00

7.3.2 Principal Component Analysis (PCA)

In order to examine if there are differences within 39 batches of the reference product itself, the principal component analysis and the cluster analysis are used [3, 6, 10]. Since the variances of the original variables are significantly varying, the PCA is carried out on the standardized values (called the z-scores) of the original variables. Since the variables seem to be highly interdependent, there is hope for understanding the behaviour of the data with smaller number of new variables (the principal components) so that the differences, if any, in the 39 batches of the reference product can be spotted. The cluster analysis is also used to spot the possible differences and the results will be crosschecked with the outcomes of the PCA.

Figure 7.3 presents the eigenvalues, their proportions and the cumulative proportions up to the first nine principal components. The Screeplot (used for assessing the true dimension of the data space) of the principal components is shown in Fig. 7.4. The first three principal components account for nearly 80 % of the total variation and hence the first three principal components may be used to understand the differences in the batches.

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	12.198	3.887	3.420	1.291	0.871
Proportion	0.488	0.155	0.137	0.052	0.035
Cumulative	0.488	0.643	0.780	0.832	0.867

	PC6	PC7	PC8	PC9
Eigenvalue	0.674	0.491	0.464	0.385
Proportion	0.027	0.020	0.019	0.015
Cumulative	0.894	0.913	0.932	0.947

Fig. 7.3 Summary of principal component analysis of reference product data

The first three principal components and their component correlations are presented in Table 7.3. The first three principal component scores are presented in Table 7.4. The scatter plot of first and second principal components is shown in Fig. 7.5 and that of the first three principal components is shown in Fig. 7.6.

From Fig. 7.5 it is clear that there are clusters within the batches of the reference product. The picture is clearer in Fig. 7.6 which takes into account the first three principal components accounting for 78 % of the total variation. From the figure, batches 31, 32, 33, 36, 37, 38 and 39 appear to be outliers. It may be recalled that batches 31, 32, 33, 38 and 39 also appear as outliers in the Dotplot (Fig. 7.2).

To sum up, the principal component analysis exhibits differences among the 39 batches of the reference product.

Table 7.3 Component loadings and correlation for reference product

	Principal component loadings			Correlations with original variables		
	PC1	PC2	PC3	PC1	PC2	PC3
x_1	0.244	−0.150	−0.116	0.853	−0.296	−0.215
x_2	−0.154	−0.168	0.013	−0.537	−0.331	0.024
x_3	0.231	−0.017	−0.153	0.808	−0.034	−0.283
x_4	−0.012	−0.128	−0.406	−0.043	−0.252	−0.752
x_5	0.232	0.242	0.053	0.811	0.477	0.099
x_6	0.142	0.230	0.282	0.496	0.453	0.522
x_7	0.223	0.263	−0.065	0.780	0.519	−0.120
x_8	−0.060	−0.142	0.372	−0.211	−0.280	0.687
x_9	0.224	−0.079	−0.267	0.782	−0.157	−0.494
x_{10}	0.226	0.268	0.102	0.790	0.529	0.188
x_{11}	0.109	−0.354	0.154	0.379	−0.697	0.286
x_{12}	0.205	0.103	0.224	0.716	0.202	0.414
x_{13}	0.225	−0.103	0.248	0.787	−0.203	0.459
x_{14}	0.257	0.063	0.175	0.899	0.124	0.324
x_{15}	0.274	−0.062	0.019	0.955	−0.122	0.035
x_{16}	0.254	−0.133	−0.057	0.887	−0.263	−0.106
x_{17}	0.017	−0.414	0.070	0.058	−0.817	0.129
x_{18}	0.269	0.150	0.024	0.940	0.295	0.044
x_{19}	0.179	−0.012	−0.223	0.624	−0.024	−0.412
x_{20}	0.136	0.086	−0.398	0.476	0.170	−0.736
x_{21}	0.251	−0.207	−0.004	0.878	−0.409	−0.007
x_{22}	0.149	−0.341	0.099	0.520	−0.673	0.183
x_{23}	0.264	0.037	0.132	0.922	0.073	0.244
x_{24}	0.144	−0.043	−0.281	0.502	−0.085	−0.520
x_{25}	0.159	−0.343	0.040	0.554	−0.676	0.075

Table 7.4 First three PC scores for reference product

Batch	PC1	PC2	PC3	Batch	PC1	PC2	PC3	Batch	PC1	PC2	PC3
1	−2.369	−0.624	3.163	14	−1.339	1.331	−0.548	27	−0.118	0.652	−1.205
2	−0.092	0.629	0.168	15	−1.009	1.068	−2.051	28	−2.163	1.097	−1.660
3	−2.667	−1.658	3.190	16	−0.608	0.283	−0.844	29	−0.692	−0.315	−1.468
4	−1.944	0.626	−1.097	17	−0.010	−0.101	−1.087	30	−2.031	−0.736	0.968
5	−2.576	−0.739	−0.747	18	0.021	−2.342	3.627	31	6.938	6.150	1.398
6	−0.438	−0.214	−0.489	19	−1.430	1.337	−0.837	32	9.299	2.852	2.549
7	−2.632	−0.710	1.898	20	−2.491	−0.438	−0.181	33	2.337	5.940	2.708
8	−2.874	−0.730	3.591	21	−2.557	0.366	−0.803	34	0.361	−0.968	−1.708
9	−2.902	−2.141	3.691	22	−2.351	0.477	−1.213	35	−1.566	0.180	−1.621
10	−2.505	−0.281	1.031	23	−2.976	−0.403	1.259	36	7.406	−3.934	0.270
11	−1.520	0.462	−0.561	24	−0.372	−0.111	−2.099	37	5.532	−2.528	0.965
12	−0.648	−0.292	−1.872	25	−2.195	1.753	−2.282	38	6.380	−2.253	−1.348
13	−0.620	−0.083	−1.559	26	−0.156	−0.184	−1.309	39	9.578	−3.415	−1.884

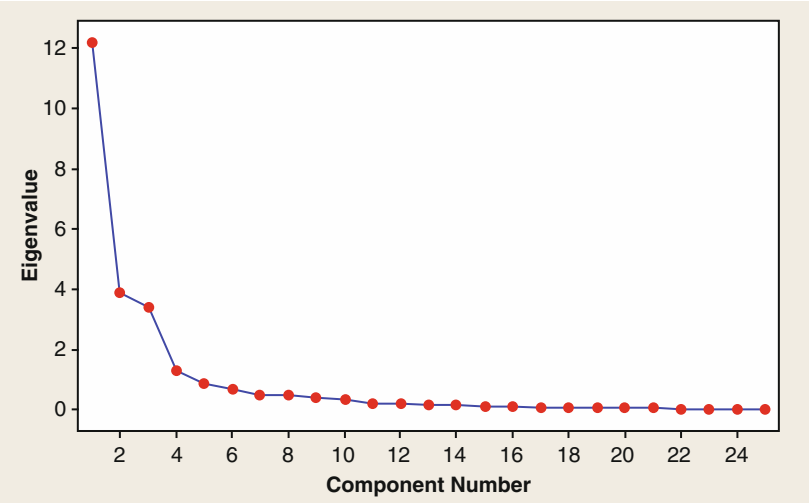


Fig. 7.4 Scree plot for the reference product data

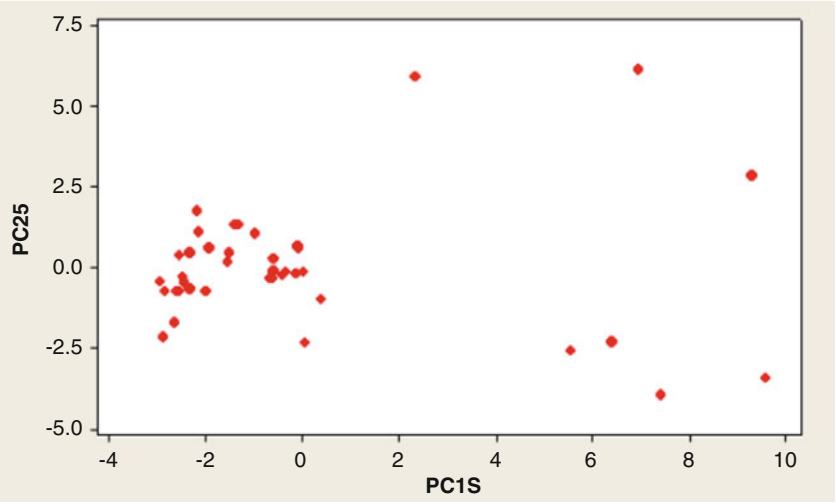


Fig. 7.5 Scatter plot of the first two principal components

7.3.3 Cluster Analysis (CA)

Cluster Analysis is used to examine and identify clusters within the 39 batches of the reference product. The Ward linkage method [6] is adopted for identifying the clusters. The number of clusters is judged from the screeplot and the dendrogram. In this case four clusters seemed to be appropriate. The four clusters and the distances between them are presented in Fig. 7.7. Cluster 1 has 11 batches, cluster 2

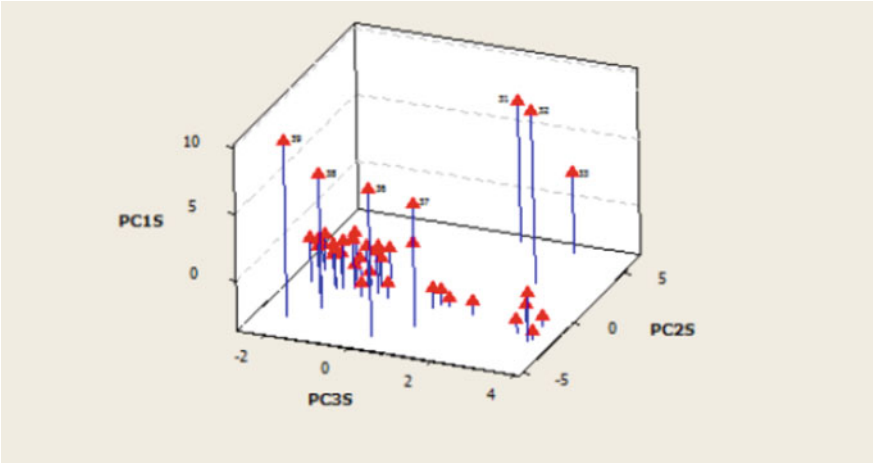


Fig. 7.6 Scatter plot of the first three principal components

Cluster 1 = $C_1 = \{1, 3, 7, 8, 9, 10, 18, 20, 21, 23, 30\}$;
Cluster 2 = $C_2 = \{2, 4, 5, 6, 11, 12, 13, 14, 15, 16, 17, 19, 22, 24, 25, 26, 27, 28, 29, 34, 35\}$;
Cluster 3 = $C_3 = \{31, 32, 33\}$;
Cluster 4 = $C_4 = \{36, 37, 38, 39\}$.

Cluster distances				
	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0.00	3.73	10.39	10.18
Cluster2	3.73	0.00	9.27	9.02
Cluster3	10.39	9.27	0.00	8.62
Cluster4	10.18	9.02	8.62	0.00

Fig. 7.7 Clusters and the distance between them for reference product

has 21 batches, cluster 3 has 3 batches and cluster 4 has 4 batches. The batch numbers are shown against each cluster. The 3-D scatter plot of the first three principal components is shown in Fig. 7.8, but this time with the clusters identified.

From Fig. 7.8, it can be observed that Cluster 3 and Cluster 4 are distinguished by the variables PC2 and PC3, but as far as PC1 is concerned, these two clusters have more or less the same PC1 scores. Therefore, to examine the differences between Cluster 3 and Cluster 4, one must examine the PC2 and PC3. The original variables that influence these two PCs are those which have large absolute numbers in the component correlations. For PC2, variables x_{17} , x_{11} , x_{22} and x_{25} are the ones that have large component correlations. Similarly, for PC3, variables x_4 , x_{20} and x_8 are the ones that have large component correlations.

Using PC1, we can make out three distinct groups: (i) Cluster 1, (ii) Cluster 2 and (iii) Clusters 3 and 4 (see Fig. 7.8). There are a large number of original variables

that have large component correlations in PC1 (see Table 7.3). It may be observed that many of these variables have high correlations (see Table 7.2).

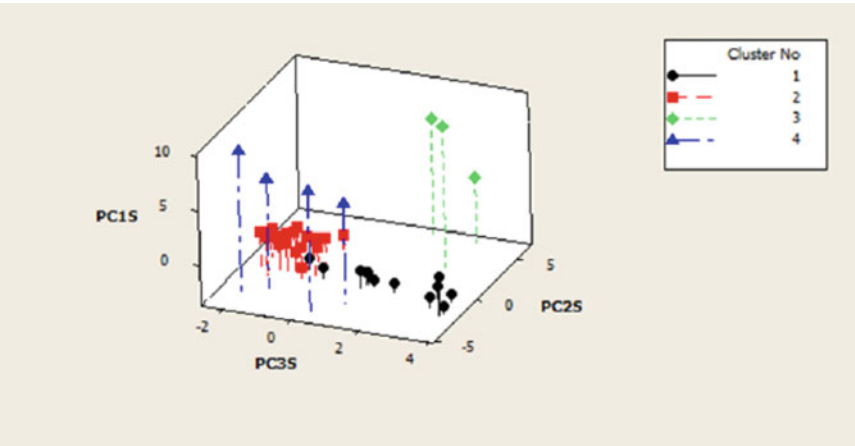


Fig. 7.8 Scatter plot first three PCs showing clusters

One more exploration was carried out with the data analysis by analysing only the 32 batches of Clusters 1 and 2. It was observed that batch 18 turned out to be an outlier. See the scatter plot of the first and second PC scores for this analysis in Fig. 7.9.

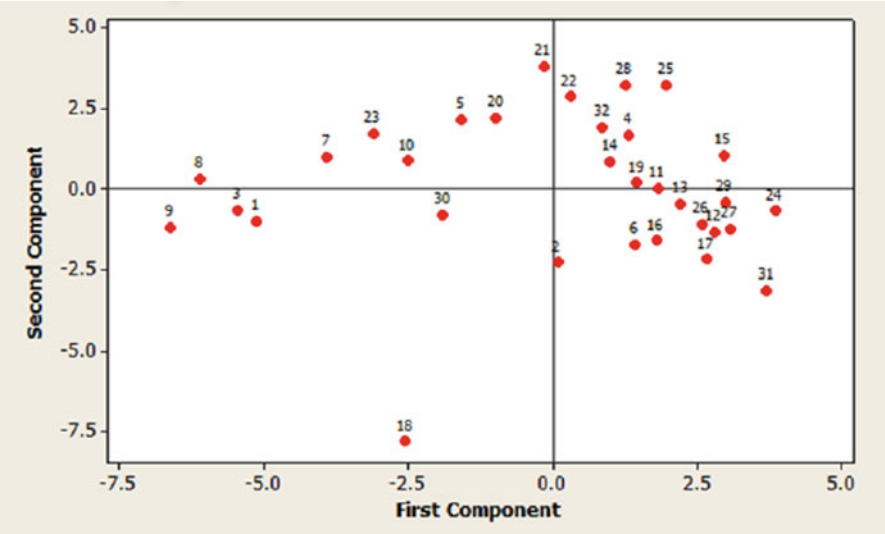


Fig. 7.9 Scatter plot first two PC scores of PCA of 32 batches of clusters 1 and 2

7.3.4 Combined Analysis of the Reference and Company's Product

To compare the Company's product with reference product, the principal component analysis and the cluster analysis were carried out on all the batches (39 batches of the reference product and the six batches of the company's product). The results are presented below. The summary statistics are presented in Table 7.5 and the sample correlation matrix is presented in Table 7.6.

The contributions of the first 8 PCs are as follows:

	PC1	PC2	PC3	PC4
Eigenvalue	9.079	4.594	4.152	1.922
Proportion	0.363	0.184	0.166	0.077
Cumulative	0.363	0.547	0.713	0.790

	PC5	PC6	PC7	PC8
Eigenvalue	1.658	1.003	0.704	0.387
Proportion	0.066	0.040	0.028	0.015
Cumulative	0.856	0.896	0.925	0.940

1. First three principal components contribute to 71 % of the total variation. The screeplot is shown in Fig. 7.10. The first three principal components and the component correlations are presented in Table 7.7. The intra and inter cluster distances are presented in Fig. 7.11.
2. Cluster analysis is carried out with four clusters and five clusters separately. The 3-D scatter plots for these two are shown in Figs. 7.12 and 7.13 respectively. The cluster means and standard deviations for the case of five clusters are presented in Table 7.8.
3. It should be noted that in both the cases, the company's product (six batches) is clearly shown as a standout. Further, the six company batches are divided into two clusters (see Figs. 7.12 and 7.13). These clusters are distinguished both by PC1 and PC3. The original variables that influence (having large component correlations) the first three principal components are highlighted in Table 7.7.
4. The seven batches of reference product (batch numbers 31, 32, 33, 36, 37, 38 and 39) are shown under one cluster in both the cases (the two cases being four clusters and five clusters). It may be recalled that when the reference product data were analysed separately, these seven batches got divided into two clusters – one containing three batches and the other containing four batches. The variable that differentiates these two clusters is PC2. And the original variables that have high (negative) correlation with PC2 are x_{21} , x_{22} and x_{25} . The remaining 32 batches of reference product also fall into two clusters (see Case 2) but they are closer compared to other clusters (examine the distances presented in Fig. 7.11). Cluster wise summary statistics are presented in Table 7.8.
5. When the number of clusters is increased from 4 to 5, the reference batches 1 to 31, 34 and 35 got divided into two clusters, and the seven batches comprising of batches 31–33 and batches 36–39 remained in one cluster only.

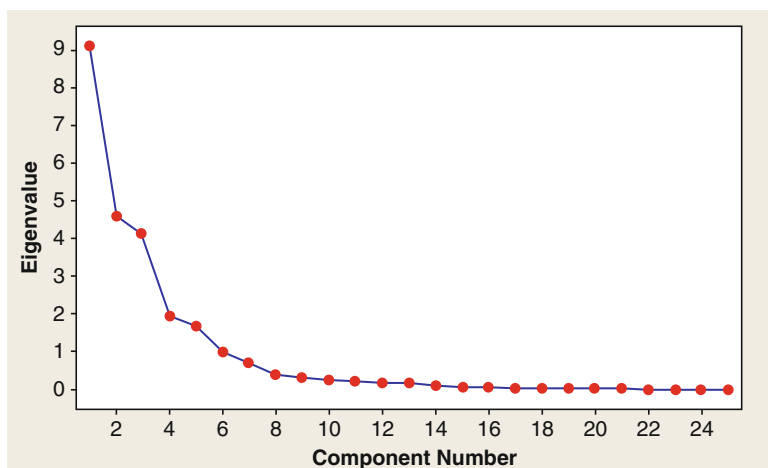


Fig. 7.10 Screeplot for the combined data

7.3.5 Summary

In this study, the product in question has two sources – the reference or the innovator and the company. The main objective of the study is to understand if there are differences between the reference and the company's products. Twenty five components of the product are considered for the analysis, and the requirement of the study is to find out if there are differences between the reference and the company's product with respect to these characteristics. Since there are multiple characteristics, the study is carried out using multivariate analysis techniques. The principal component analysis and the cluster analysis are employed in the analysis.

There are 39 observations (batches) on the reference product and only six observations on the company's product. The analysis is carried out in two stages. In the first stage, only the reference product data were analyzed and in the second stage all the batches (reference and company's) were analyzed together.

It is observed that, in the first stage analysis, that there are differences even in the case of reference product. Four clusters could be identified as distinct clusters. After deleting two of the clusters, it was found that one of the batches (batch number 18) was still an outlier.

In the second stage analysis where the two products were analyzed together, it was found that the company's product emerged as a clear standout. Further, within the six batches of the company's product, there were two clusters each containing three batches. Various statistical measures are presented in the report for looking into the details and examining the data.

Case 1: Number of clusters = 4

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	32	197.756	2.41641	4.13862
Cluster2	7	165.096	4.71095	6.85088
Cluster3	3	32.087	3.21723	4.04268
Cluster4	3	16.253	2.32233	2.51173

Exhibit 3 (Contd.)

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0.0000	6.49474	10.4336	8.2409
Cluster2	6.4947	0.00000	9.2359	8.5973
Cluster3	10.4336	9.23592	0.0000	11.7090
Cluster4	8.2409	8.59734	11.7090	0.0000

Case 2: Number of clusters = 5

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	8	27.018	1.75297	3.11434
Cluster2	24	107.174	2.07593	2.92837
Cluster3	7	165.096	4.71095	6.85088
Cluster4	3	32.087	3.21723	4.04268
Cluster5	3	16.253	2.32233	2.51173

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0.0000	3.2548	7.29591	11.3988	8.4379
Cluster2	3.2548	0.0000	6.41461	10.2218	8.3347
Cluster3	7.2959	6.4146	0.00000	9.2359	8.5973
Cluster4	11.3988	10.2218	9.23592	0.0000	11.7090
Cluster5	8.4379	8.3347	8.59734	11.7090	0.0000

Fig. 7.11 Output of cluster analysis

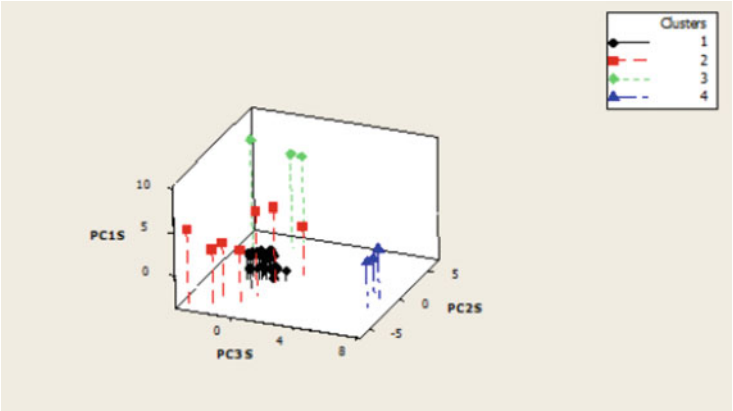


Fig. 7.12 Scatter plot showing four clusters of the combined data

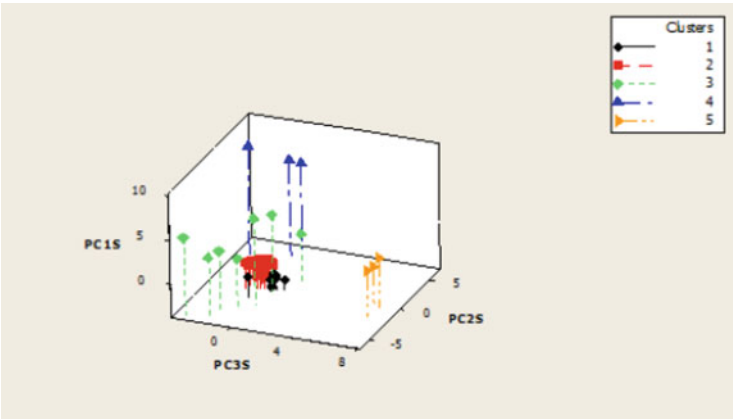


Fig. 7.13 Scatter plot showing five clusters of the combined data

Table 7.5 Basic statistics of the data on reference product

Variable	Mean	St dev	Variance	Min	Q_1	Median	Q_3	Max
x_1	2.446	1.427	2.036	1.354	1.708	1.860	2.673	7.687
x_2	30.919	1.801	3.244	24.257	30.025	31.081	31.996	33.625
x_3	2.964	0.968	0.937	1.091	2.332	2.808	3.235	6.342
x_4	16.897	1.644	2.703	12.098	15.807	17.245	18.219	19.374
x_5	3.516	0.823	0.677	2.769	3.027	3.261	3.564	6.607
x_6	52.134	2.355	5.546	48.913	50.445	51.511	53.238	58.796
x_7	1.843	0.516	0.266	1.207	1.583	1.711	1.864	3.746
x_8	11.101	1.833	3.360	7.351	10.257	10.890	11.475	17.104
x_9	5.403	0.990	0.980	4.096	4.841	5.242	5.601	8.719
x_{10}	6.393	1.729	2.989	5.087	5.467	5.885	6.292	12.556
x_{11}	3.799	0.499	0.249	2.343	3.635	3.880	4.110	4.562
x_{12}	1.738	0.276	0.076	1.326	1.503	1.752	1.885	2.457
x_{13}	6.992	0.853	0.728	5.918	6.442	6.749	7.251	9.814
x_{14}	1.808	0.370	0.137	0.946	1.616	1.736	1.880	2.992
x_{15}	2.134	1.787	3.193	1.127	1.293	1.381	2.341	8.231
x_{16}	1.591	0.368	0.135	1.240	1.395	1.494	1.620	3.044
x_{17}	3.384	0.376	0.141	2.228	3.217	3.429	3.641	4.030
x_{18}	1.675	0.451	0.203	1.177	1.378	1.523	1.908	3.007
x_{19}	5.914	1.223	1.496	3.950	4.897	5.913	6.570	8.484
x_{20}	8.325	2.778	7.717	3.400	6.647	8.275	9.963	15.598
x_{21}	1.664	0.356	0.127	0.938	1.474	1.540	1.755	2.701
x_{22}	11.121	1.743	3.038	6.059	10.254	11.195	12.169	14.588
x_{23}	1.425	0.574	0.329	1.011	1.102	1.161	1.480	3.154
x_{24}	2.733	1.023	1.047	0.393	2.023	2.761	3.427	5.036
x_{25}	4.811	0.624	0.389	3.840	4.470	4.718	4.955	6.808

Table 7.6 Correlation matrix of the variables for reference product

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}
x_1	1.00																								
x_2	-0.41	1.00																							
x_3	0.87	-0.43	1.00																						
x_4	-0.05	-0.20	0.21	1.00																					
x_5	0.82	-0.41	0.65	-0.29	1.00																				
x_6	0.36	-0.71	0.24	-0.55	0.55	1.00																			
x_7	0.80	-0.49	0.69	-0.03	0.94	0.44	1.00																		
x_8	-0.18	0.32	-0.23	-0.34	-0.14	-0.02	-0.25	1.00																	
x_9	0.88	-0.36	0.75	0.44	0.55	-0.01	0.65	-0.39	1.00																
x_{10}	0.80	-0.56	0.58	-0.29	0.94	0.69	0.90	-0.14	0.46	1.00															
x_{11}	-0.25	-0.07	0.26	0.01	-0.02	0.07	-0.10	0.32	0.23	-0.31	1.00														
x_{12}	0.45	-0.35	0.33	-0.36	0.72	0.55	0.61	-0.01	0.36	0.58	0.26	1.00													
x_{13}	0.47	-0.32	0.46	-0.33	0.59	0.52	0.46	0.18	0.43	0.27	-0.38	0.26	1.00												
x_{14}	0.28	-0.57	0.60	-0.25	0.80	0.67	0.72	-0.02	0.52	0.40	0.57	0.76	0.03	1.00											
x_{15}	0.91	-0.40	0.79	-0.06	0.73	0.39	0.68	-0.14	0.75	0.87	-0.49	0.34	0.42	0.07	1.00										
x_{16}	0.62	-0.38	0.73	0.10	0.55	0.26	0.53	-0.20	0.75	0.38	0.26	0.54	0.47	0.64	0.31	1.00									
x_{17}	0.24	0.14	0.06	0.11	-0.34	-0.19	-0.41	0.27	0.09	0.04	0.45	0.08	-0.01	0.21	0.12	0.22	1.00								
x_{18}	0.61	-0.57	0.74	-0.17	0.92	0.61	0.89	-0.21	0.67	0.62	-0.08	0.66	0.56	0.65	0.45	0.76	-0.14	1.00							
x_{19}	0.10	-0.10	0.60	0.04	0.49	0.07	0.49	-0.51	0.62	-0.08	0.14	0.15	0.14	0.26	-0.12	0.48	-0.03	0.42	1.00						
x_{20}	0.47	-0.28	0.58	0.43	0.41	-0.07	0.57	-0.58	0.68	0.32	-0.50	-0.05	0.35	-0.14	0.42	0.34	-0.24	0.47	0.44	1.00					
x_{21}	0.19	-0.36	0.74	0.09	0.51	0.25	0.47	-0.06	0.76	-0.13	0.30	0.27	0.52	0.38	-0.08	0.69	0.15	0.55	0.50	0.22	1.00				
x_{22}	-0.27	-0.15	0.38	0.06	0.09	0.10	0.06	0.17	0.46	-0.54	0.34	-0.09	0.40	0.09	-0.45	0.31	0.20	0.09	0.36	0.00	0.71	1.00			
x_{23}	0.44	-0.58	0.70	-0.17	0.78	0.62	0.73	-0.04	0.58	0.27	-0.43	0.27	0.89	0.07	0.41	0.46	-0.20	0.67	0.24	0.47	0.62	0.36	1.00		
x_{24}	0.02	-0.31	0.48	0.29	0.30	0.05	0.38	-0.44	0.63	0.03	0.55	0.13	-0.35	0.53	-0.22	0.40	0.25	0.21	0.34	0.06	0.25	0.05	-0.33	1.00	
x_{25}	0.21	0.08	0.43	-0.05	0.16	-0.03	0.05	0.05	0.50	-0.17	0.42	0.22	0.34	0.33	-0.05	0.58	0.41	0.27	0.43	0.00	0.74	0.66	0.30	0.23	1.00

Table 7.7 Component loadings and correlation for reference product

	Principal component loadings			Correlations with original variables		
	PC1	PC2	PC3	PC1	PC2	PC3
x_1	0.298	0.03	-0.005	0.898	0.064	-0.01
x_2	-0.196	0.016	-0.052	-0.591	0.034	-0.106
x_3	0.297	0.057	-0.077	0.895	0.122	-0.157
x_4	-0.049	-0.025	-0.266	-0.148	-0.054	-0.542
x_5	0.304	0.146	-0.047	0.916	0.313	-0.096
x_6	0.188	0.003	0.227	0.566	0.006	0.463
x_7	0.302	0.154	-0.055	0.910	0.33	-0.112
x_8	-0.073	-0.165	0.41	-0.220	-0.354	0.835
x_9	0.279	0.113	-0.142	0.841	0.242	-0.289
x_{10}	0.298	0.176	-0.034	0.898	0.377	-0.069
x_{11}	-0.078	-0.212	-0.365	-0.235	-0.454	-0.744
x_{12}	0.213	-0.084	-0.13	0.642	-0.18	-0.265
x_{13}	0.165	-0.202	0.314	0.497	-0.433	0.64
x_{14}	0.169	-0.17	-0.271	0.509	-0.364	-0.552
x_{15}	0.277	0.168	0.084	0.835	0.36	0.171
x_{16}	0.221	-0.281	-0.113	0.666	-0.602	-0.23
x_{17}	0.019	-0.093	-0.223	0.057	-0.199	-0.454
x_{18}	0.279	-0.171	0.032	0.841	-0.367	0.065
x_{19}	0.062	-0.262	-0.062	0.187	-0.562	-0.126
x_{20}	0.163	-0.026	0.138	0.491	-0.056	0.281
x_{21}	0.088	-0.422	0.031	0.265	-0.905	0.063
x_{22}	-0.078	-0.411	0.08	-0.235	-0.881	0.163
x_{23}	0.181	-0.207	0.332	0.545	-0.444	0.677
x_{24}	0.034	-0.122	-0.368	0.102	-0.261	-0.750
x_{25}	0.044	-0.372	-0.076	0.133	-0.797	-0.155

7.4 Estimation of Reliability of CFL Bulbs

This project was carried out for an electric bulbs manufacturing company. One of the types of bulbs, CFL bulbs (Fig. 7.14), are given a warranty of 6000 h of life. As per the company's quality practices, five bulbs are taken from each day's production and are put on test. The company has asked one of their consultants to evaluate the reliability of bulbs with respect to their life. The company wanted to know whether the estimate given by the consultant was correct. Also, the company wanted to understand the methodology of evaluating the reliability. The test adopted is a combination of time and number censored. A pertinent question in this regard is: what should be the number of test stations in order to ensure that the daily tests are carried out smoothly. This question is answered using a queuing model approach and the answer is obtained using simulation technique. A computer program is developed to compute the study state probabilities [5].

Table 7.8 Cluster Means and standard deviations

Cluster/size Statistics		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
Cluster 1	Mean	2.2	29.4	3.08	15.13	3.98	47.98	2.77	12.16	5.77	6.75	5.63	3.62
$n = 7$	Std	2.18	7.7	2.6	1.81	3.01	13.16	4.21	1.17	3.84	3.86	4.64	5.77
Cluster 2	Mean	2.72	31.06	3.57	18.19	4.19	49.92	2.83	11.45	6.3	6.81	5.05	2.96
$n = 7$	Std	4.43	1.64	4.7	2.01	4.96	4.53	5.66	4.1	5.36	5.47	6.03	6.68
Cluster 3	Mean	3.89	26.8	4.44	15.5	5.02	48.95	3.57	11.2	7.03	8.54	5.75	4.12
$n = 7$	Std	1.75	8.18	1.88	3.05	2.46	15.21	3.82	1.41	3.27	3.24	4.57	5.61
Cluster 4	Mean	5.62	23	5.48	13.73	6.28	42.77	4.79	8.5	9.11	11.93	5.67	5.04
$n = 7$	Std	1.98	12.67	0.63	5.63	0.66	23.31	2.82	1.23	1.3	0.69	4.91	5.98
Cluster 5	Mean	3.41	23.48	3.52	12.63	4.32	43.39	3.59	15.05	6.51	7.7	5.09	4.61
$n = 7$	Std	0.35	13.1	0.99	4.43	1.79	23.64	3.6	3.38	3.03	2.86	5.27	6.26

Table continued

Cluster/size Statistics		x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}
Cluster 1	Mean	8.4	3.93	3.55	3.81	5.88	3.93	7.06	6.7	4.49	13.65	4.36	5.28	7.96
$n = 7$	Std	4.37	6.4	6.92	7.2	6.8	7.9	7.12	7.65	8.82	5.78	9.62	9.66	9.02
Cluster 2	Mean	7.69	3.11	2.81	2.96	4.8	3.06	7.47	9.94	3.26	12.3	2.92	4.83	6.35
$n = 7$	Std	5.91	7.06	7.33	7.51	7.34	7.9	7.23	7.04	8.49	6.86	8.97	8.81	8.68
Cluster 3	Mean	9.05	4.64	4.81	4.67	5.8	5.21	9.71	11.51	5.34	14.27	5.45	6.78	8.74
$n = 7$	Std	4.04	6.22	6.55	7.02	6.97	7.6	6.3	6.24	8.76	5.94	9.52	9.42	9.06
Cluster 4	Mean	9.09	5.46	10.3	5.81	7.52	6.51	8.7	13.25	6.61	11.05	7.29	8.33	9.96
$n = 7$	Std	4.02	7.03	4.47	8.13	7.65	9	8.21	6.21	10.93	8.64	11.81	11.78	11.36
Cluster 5	Mean	10.78	4.8	7.3	5.69	6.97	6.44	9.6	15.21	7.21	15.41	8.52	6.81	10.35
$n = 7$	Std	2.83	7.47	6.47	8.21	8.03	9.04	7.7	4.96	10.53	5.91	10.98	12.8	11.11

7.4.1 Background of CFL Problem

The bulbs are given a warranty of 6000 h of life. The company wanted to estimate the reliability of bulbs with respect to their life and addressed the problem to a consultant. Having got an answer from the consultant, the company wanted to know whether the estimated reliability was correct or not. In this connection, company has approached the author to help them in this problem. Besides evaluation of the reliability of bulbs, this project examines a pertinent question in this regard: How many test stations are required to carry out the daily tests smoothly. This question is analyzed using a queuing model approach and the answer is obtained using simulation of the system. A computer program is developed to simulate the system and compute the study state probabilities.

7.4.2 Life Testing

The company produces CFL bulbs. The bulbs carry a warranty of 6000 h of life. From daily production, five bulbs are sampled on each working day (Monday to Friday) and are put on test for evaluation of the life. Generally life testing is carried



Fig. 7.14 Compact fluorescent lamp (CFL) gives the same amount of light compared to general-service incandescent lamps, use one-fifth to one-third the electric power, and last 8–15 times longer

out as follows. A certain number of bulbs, say k bulbs, are put on test. In most cases the test is censored in one of the two following ways:

- (i) **Time Censored:** the test is continued for a fixed period of time and the number of bulbs that have failed during the period is noted along with the times of their failures,
- (ii) **Number Censored:** the test is continued until a fixed number of bulbs from the k bulbs put on test have failed and the failure times of the failed bulbs are noted.

Company's Test Procedure

The method adopted by the company is as follows. From the daily production five bulbs are randomly selected and put on test. At 100 h from the beginning of the test, three of them are suspended and the remaining two are continued. The selection of the two is randomly made. The failure times of the bulbs that fail in the first 100 h are recorded. At the end of the 2000 h from the beginning, only one is continued for further observation and the other, if any, is suspended. The failure time(s) of the bulb(s) that fails (fail) during the period 100–2000 h is (are) recorded. The test for the bulb that lasts beyond 2000 h, if any, is continued up to 6000 h and at 6000 h the test is terminated. If the bulb fails between 2000 and 6000 h, its failure time is recorded.

Though the tests are performed from the daily production, it is observed that the data are not recorded in a systematic manner. One of the contributions of this project is that a nice format was suggested for recording the data. The company was impressed with the suggestion and readily agreed to implement. The company felt that the new format suggested for recording the test data not only brought clarity in understanding the status but greatly helped in monitoring quality [8, 9] as well.

7.4.3 Data Collection

It is understood that the company maintains a record in which the failure times of the bulbs that have failed is noted. It was difficult to collect data on the bulbs that were suspended. It will be erroneous to analyze only the data of the bulbs whose exact failure times are recorded. Following this, it was recommended that the past data be compiled for one month in the format suggested in Table 7.9. According to this format, a row is maintained for each bulb. The first column specifies the date on which the bulb is put on test; the second column specifies the bulb number; the third column specifies the actual number of hours that the bulb has lasted provided the bulb has failed during the testing period (0–6000 h), else a zero is entered in this column; the fourth column will have a 1 if the bulb is suspended at 100 h, else a zero is entered; the fifth column will have a 1 if the bulb is suspended at 2000 h, else a zero is entered; the sixth column will have a 1 if the bulb is suspended at 6000 h, else a zero is entered. This format will facilitate smooth processing of the data.

It may be noted that the sum of the entries of each row of the last three columns is either 0 or 1. If this sum is one, it means that the test of that bulb is censored. In fact, if x_4 , x_5 and x_6 are the entries of any row of the last three columns, then $100x_4 + 2000x_5 + 6000x_6$ is equal to the time at which the bulb testing is suspended provided $x_4 + x_5 + x_6 = 1$.

When this study was taken up, the company executive expressed the difficulty in extracting complete past data. Consequently, it was decided that the data for one month be collected for the purpose of the study. Data were compiled for the month of April 2011. These data had the test results for 130 bulbs. Of these, 75 were suspended at the end of 100th hour, 24 were suspended at the end of 2000th hour, and 23 were suspended at the end of 6000th hour. The failure times of the remaining 8 bulbs are: 69, 128, 229, 618, 4226, 4503, 5884 and 5991 h. Henceforth, these data will be referred to as *April data*.

7.4.4 Objectives

The primary objective of this study is to evaluate the reliability of bulbs with respect to their life and compare it with what the consultant had given. A secondary objective is to evaluate the number of test stations that will be required to conduct the test

Table 7.9 Format for data collection

Test date	Lamp no.	Actual life (h)	Suspended at 100 h	Suspended at 2000 h	Suspended at 6000 h
01-Apr	1	0	0	0	1
01-Apr	2	618	0	0	0
01-Apr	3	0	1	0	0
01-Apr	4	0	1	0	0
01-Apr	5	0	1	0	0
02-Apr	11	0	0	0	1
02-Apr	12	0	0	1	0
02-Apr	13	0	1	0	0
02-Apr	14	0	1	0	0
02-Apr	15	0	1	0	0
...
29-Apr	126	0	0	0	1
29-Apr	127	0	0	1	0
29-Apr	128	0	1	0	0
29-Apr	129	0	1	0	0
29-Apr	130	0	1	0	0
30-Apr	131	229	0	0	0
30-Apr	132	0	0	1	0
30-Apr	133	0	1	0	0
30-Apr	134	0	1	0	0
30-Apr	135	0	1	0	0

smoothly. This question arose while discussing with the concerned executive. This is an important question because the company needs to know how many test stations should be provided so as to ensure that the testing is carried out smoothly without interruption.

7.4.5 Evaluation of Reliability Function

As described above, the type of test conducted uses a combination of time and number censoring. If X is the random variable representing the life of a bulb and has distribution function F , then the reliability function of the bulb is given by

$$R(t) = P(X > t) = 1 - F(t), \quad (7.9)$$

where t is the time in hours. Since 6000 h has been specified as the warranty period, the reliability of the bulb is equal to $R(6000)$. The reliability of the bulbs is estimated by first identifying the distribution F using the life test data available. The Minitab package (<http://www.minitab.com/en-us/>) has a user-friendly module on Reliability. It allows identification of life time distribution using three types of data, namely, *right censored*, *left censored* and *arbitrary censored*. The right censored data needs two entries for each bulb tested, namely (x_i, C_i) , where x_i is the time and C_i is the indicator taking values either 0 or 1. If $C_i = 1$, then x_i is the failure time, and if

$C_i = 0$, it means that the bulb has not failed until time x_i . Left censored data are similarly interpreted with the difference that when $C_i = 0$ it means that the bulb has failed at some time during $(0, x_i)$ but the exact failure time is not known. The arbitrary censored data comprises a pair of times (x_i, y_i) which means that the bulb has failed some time between x_i and y_i . When $x_i = y_i$, it means that the bulb has failed exactly at time x_i .

Clearly, the type of life test listed in Table 7.9 is right censored data and hence need to be analyzed using *Distribution Analysis (Right Censor)* module. Under this module, there is a tool to examine if any of the known distributions such as Weibul, normal, lognormal, exponential, etc., fits closely to the observed data. The tool uses probability plots for the distribution identification. In practice, often one finds that none of these distributions fits exactly, meaning that the p-values are generally smaller than 0.05. However, one might choose the distribution that more or less shows satisfactory trend on the probability plot to be the closest approximation to the given data and use the same to carry out further analysis. For the April data presented above, the distribution identification plots are shown in Figs. 7.15 to 7.17. Two distributions, Weibul and lognormal, are the closest fits among the eleven distributions tried for the observed data. In case of life of electronic items,

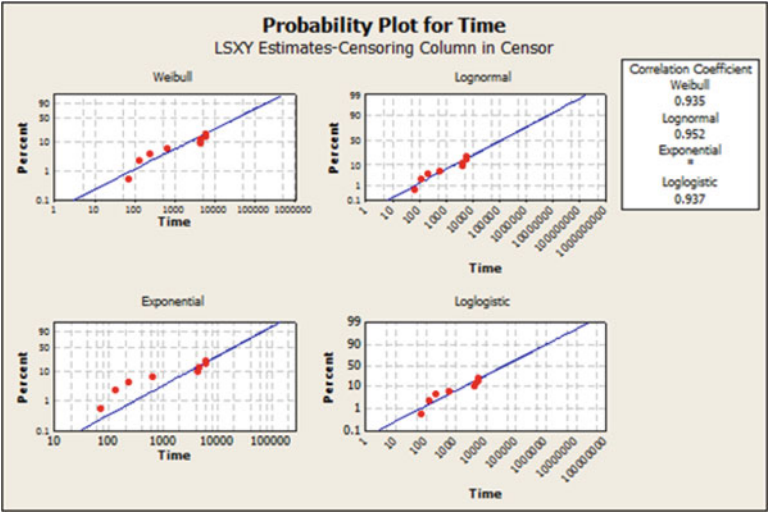


Fig. 7.15 Distribution identification plots

experience suggests that Weibul is a commonly encountered distribution. In view of this, the life time distribution of X is taken as Weibul. A closer look at the distribution using Weibul is drawn using the Distribution Overview plot of the Minitab reliability module. This gives the picture of the density function, the probability plot, the reliability function and the hazard function, all in one graph. The graph is

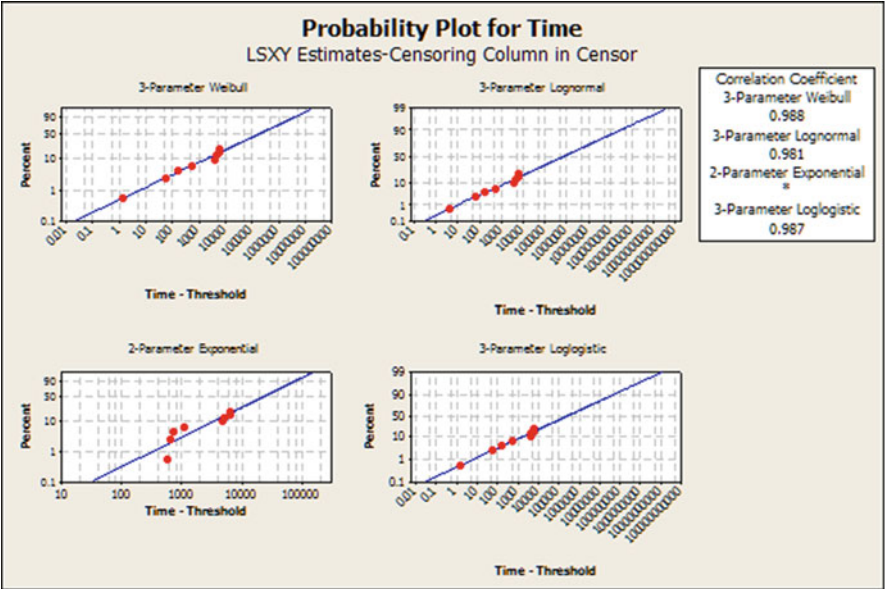


Fig. 7.16 Distribution identification plots

shown in Fig. 7.19. The parameters of the fitted distribution are: shape parameter $\beta = 0.07005$, and the scale parameter $\alpha = 54221$. The estimated mean time to failure or equivalently the average life of a bulb is equal to 68589 h and the median life is 32132 h. The reliability function is given by

$$R(t) = e^{-(\frac{t}{\alpha})^\beta} = e^{-(\frac{t}{54221})^{0.07005}} \tag{7.10}$$

The parametric analysis module of Minitab can be used to estimate the probabilities and percentiles. In particular, the reliability, $R(6000)$, is estimated using this module. The output of the analysis is presented in Fig. 7.20. As part of the analysis, the probability plot is also drawn which is shown in Fig. 7.18. As per this analysis, the estimated reliability is 80.74 % with a 95 % confidence interval of (65.17 %, 89.86 %).

7.4.6 Consultant’s Estimate

The company also provided the analysis of their consultant. It is understood that the consultant has used 20 randomly selected bulbs for the analysis. These 20 bulbs were

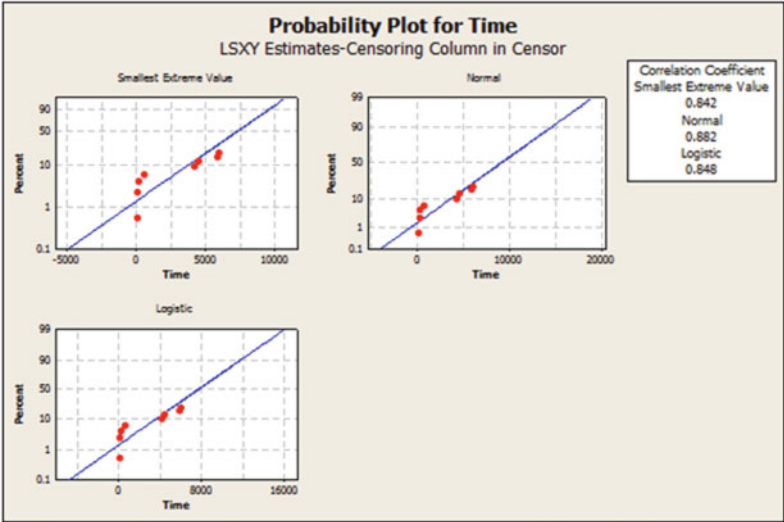


Fig. 7.17 Distribution identification plots

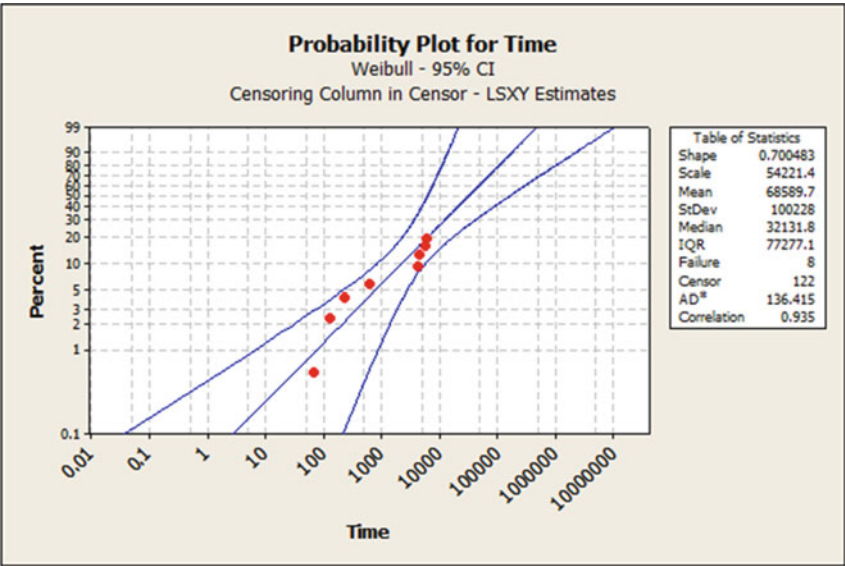


Fig. 7.18 Weibull probability plot for the life data

tested by the consultant and the analysis is based on the corresponding test results. The data and the analysis are presented in Fig. 7.21. Out of 20 bulbs shown in the analysis, 19 have failed before 1250 h (failure times provided) and one is reported as having discontinued at 1250 h. From this, one estimate of the probability that a bulb fails within 1250 h is approximately 0.95. But from the data (of April), at least

51 bulbs have lasted more than 2000 h. Therefore, an estimate of a lower bound for $R(2000)$ is approximately equal to $0.39 (= 51/130)$.

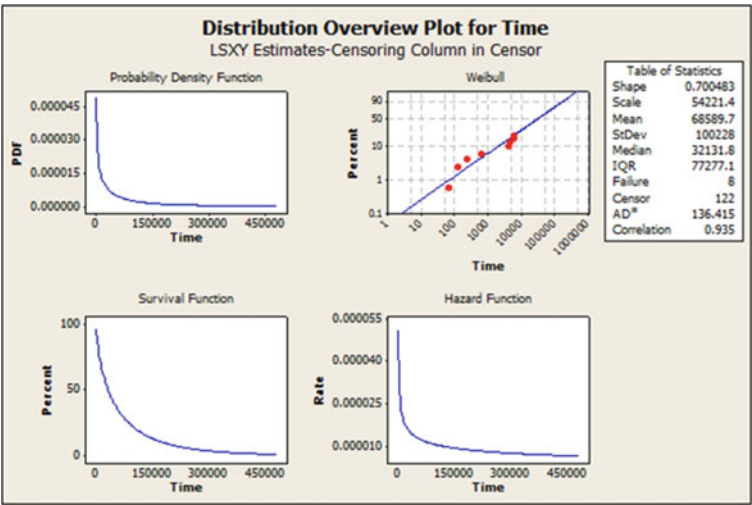


Fig. 7.19 Weibul fitting for the life data

In contrast, the consultant’s estimate for the reliability is 51.30 % as opposed to 80.74 % estimated above. The consultant seems to have used three-parameter Weibul distribution. Adopting the same distribution for the April data, the three parameter Weibul distribution is fitted and is given in Fig. 7.22. As per this, the estimated reliability is 84.67 %.

To compare, the 3-parameter Weibul is also fitted to the data provided by the consultant. The probability plot and the distribution fit are shown in Fig. 7.23. According to this, the estimated reliability is 0 % as opposed to 51.30 % projected by the consultant. Therefore, there is a need for understanding the details of consultant’s analysis.

7.4.7 Simulation for Determination of Number of Test Holders

In this subsection, the number test stations required for carrying out the test as per the present norms, that is, putting five bulbs on each of the working days (Monday to Friday) and adopting the right censor procedure at 100, 2000 and 6000 h as described earlier is determined. Let X_i denote the number bulbs on test on Day i . Note that $\{X_i : i = 0, 1, 2, \dots\}$ is a Markov chain. If N is the number of test stations to be provided, then for smooth testing it may be reasonable to assume that N should be such that $P(X_i \leq N) \geq 0.95$ for any Day i . Clearly, this is a queuing system, and X_i denotes the number of customers (test stations) in the system on Day i [13]. The

Exhibit 1: Distribution Analysis: Time

Variable: Time

Censoring Information

Count

Uncensored value

8

Right censored value

122

Censoring value: Censor = C

Estimation Method: Least Squares (failure time(X) on rank(Y))

Distribution: Weibull

Parameter Estimates

Parameter	Estimate	Standard Error	95.0% Normal CI	
			Lower	Upper
Shape	0.700483	0.213293	0.385666	1.27228
Scale	54221.4	52535.4	8117.80	362163

Log-Likelihood = -89.501

Goodness-of-Fit

Anderson-Darling (adjusted) = 136.415

Correlation Coefficient = 0.935

Characteristics of Distribution

	Estimate	Standard Error	95.0% Normal CI	
			Lower	Upper
Mean (MTTF)	68589.7	84363.8	6156.00	764221
Standard Deviation	100228	156159	4729.16	2124188
Median	32131.8	26785.5	6271.39	164630
First Quartile (Q1)	9156.35	5153.81	3038.15	27595.4
Third Quartile (Q3)	86433.5	94639.3	10108.1	739085
Interquartile Range (IQR)	77277.1	90296.2	7824.30	763232

Fig. 7.20 Parametric analysis of the bulb lives data

arrival and departure distributions are governed by the Weibul distribution and the censoring procedure. In order to determine N , one must obtain the steady state probabilities of the system. These are obtained by simulating the system using Monte Carlo simulation. The arrival times are simulated using the Weibul distribution with the estimated parameters using April data. A computer program for this is written using Excel macros. The system is simulated for 1000 days. An extract of the simulated data is presented in Table 7.10. The simulated system is shown as a time series plot in Fig. 7.24. From the simulated data, it can be safely assumed that the system is stable after Day 700 onward. In fact, the system stabilizes after 250 days (recall that 250 days corresponds to 6000 h of final censor of any day’s bulbs). Therefore, to understand the steady state probabilities, the descriptive statistics of data on $X_i, i \geq 700$, is presented in Fig. 7.25. The steady state probabilities summarized from $X_i, i \geq 700$, are presented in Table 7.11.

7.4.8 Summary

This project is taken up to estimate the reliability of CFL bulbs. The bulbs are given a warranty of 6000 h. The bulbs are subjected to right censored life on a sample basis on every working day. The company wanted to understand the reliability of

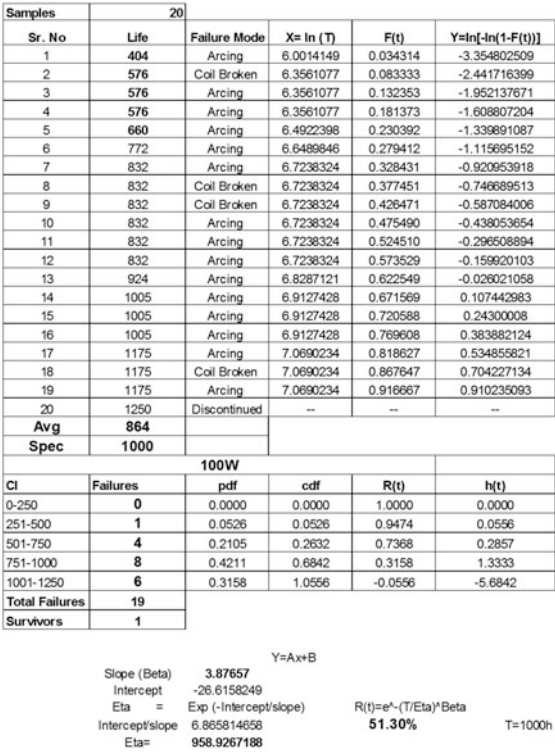


Fig. 7.21 Consultant’s analysis of the bulb lives data

the bulbs and also wanted to know whether the estimate given by one of their consultants was correct. One month’s data were collected from the past records and the parametric reliability analysis was performed. The reliability model was fitted with the two- parameter Weibul distribution which was the closest among the standard distributions tried. With the fitted Weibul, the estimated reliability of the bulbs is 80.74 % with a 95 % confidence interval of (65.17 %, 89.86 %). There appears to be some inconsistencies in the consultant’s analysis.

In order to determine the number of test stations for smooth testing, the system of number of stations is simulated using Monte Carlo simulation. The system stabilizes after 250 days. Approximately 350 test stations will be required to conduct the test smoothly.

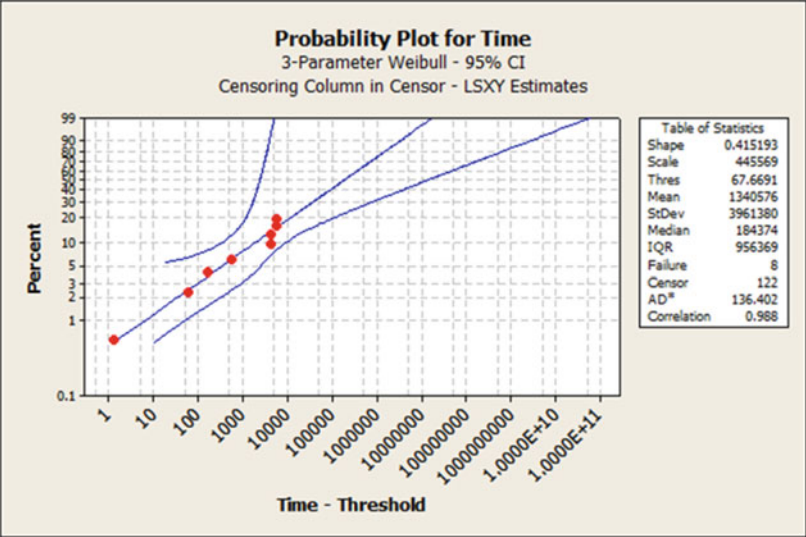


Fig. 7.22 Three-parameter Weibull fitting for April Data

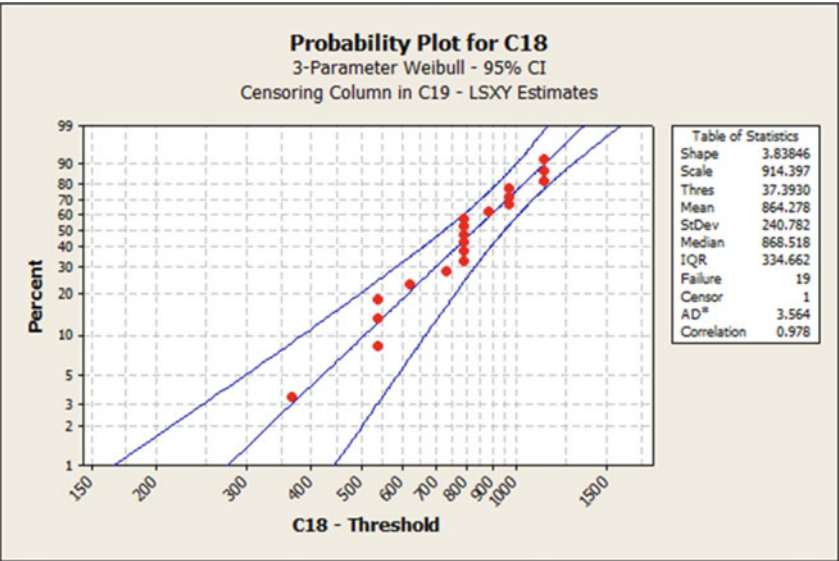


Fig. 7.23 Three-parameter Weibull fitting for consultant’s data

Table 7.10 Simulation of bulbs testing

Day	Bulb 1	Bulb 2	Bulb 3	Bulb 4	Bulb 5	Bulbs On Test
1	511394.2	80516.9	38876.7	40288.3	30104	5
2	121672.4	48104.3	31567.3	411102.8	42117	10
3	33889.8	574842.4	21782.5	56959.7	13137.8	15
4	167162.1	14225.8	8252.6	4922.5	23356.6	20
5	19275.5	45787.9	3312.9	46901.2	8646.3	22
6	3967.6	18777.3	6582.9	451124.4	22597	19
7	5055.3	159939.7	20781.5	341415	65172.7	26
8	46354.3	356.3	221172.5	172972.9	929.5	28
9	43500.9	11928.4	24245.4	69281.5	4915	30
238	80711.1	212664.6	4137.5	152626.9	1430.5	324
239	1492.7	34788.5	219401.3	39521.5	238620	325
240	345499.1	25653.7	30013.2	14104.2	211.8	326
997	37792.8	13468.8	18480.1	48615.4	536250.7	340
998	79755.3	88139.4	72539.7	45655.2	12147.2	340
999	44645.9	39323.4	72160.9	3098.2	159.2	339
1000	2052.5	184349.4	3521.3	96326.4	202765.7	334

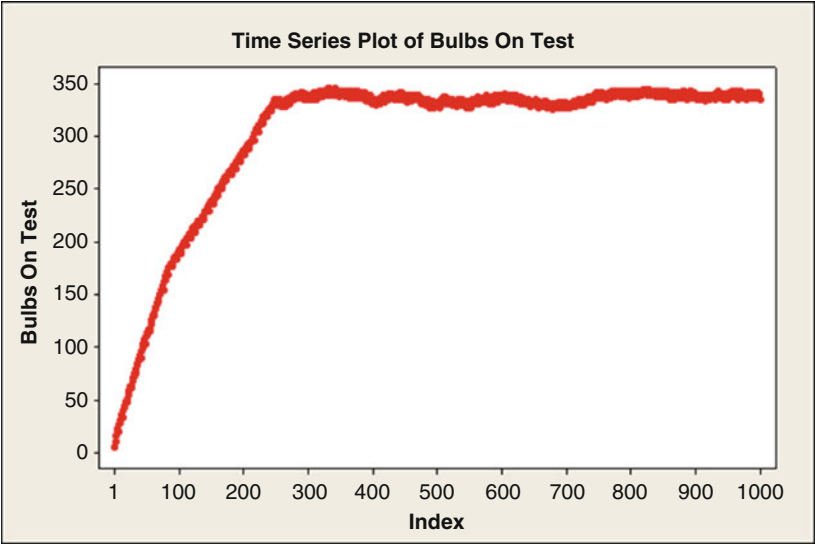


Fig. 7.24 Time series plot for the simulated system data

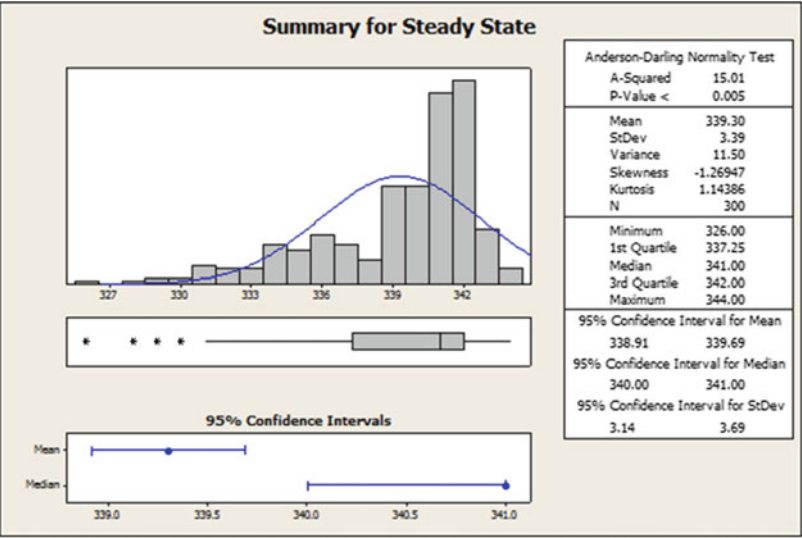


Fig. 7.25 Steady state statistics

Table 7.11 Steady state probability distribution

State	326	328	329	330	331	332	333	334	335
Probability	0.003	0.003	0.007	0.007	0	0.017	0.017	0.043	0.037
State	336	337	338	339	340	341	342	343	344
Probability	0.053	0.043	0.027	0.107	0.1	0.21	0.223	0.06	0.017

7.5 Product Quality Rating System for Paper Boards

This project was carried out for a paperboard manufacturing company. Paper boards are thick papers, usually with weight of 200 g/m² (GSM) or above, used for a variety of purposes, the most common application being packaging material (Fig. 7.26). The quality of a paperboard is determined based on a number of characteristics such as grammage (GSM), caliper, bulk, bursting factor, tensile strength, moisture content, ash percentage, brightness, opacity, etc. Using a multivariate approach, an attempt is made to develop a quality rating system for the paperboards for the company. From the analysis of the data, Euclidean distance from the target specifications seems to work well to derive a measure of quality index for the product. Further exploration has been recommended.

7.5.1 The Quality Rating Problem

This project aims at developing a Product Quality Rating System (PQRS) for the paperboards for a leading paperboard manufacturing company. There are a number of quality characteristics of a paperboard that determine the quality of the board.

Some of these are grammage (GSM), caliper, bulk, bursting factor, tensile strength, moisture content, ash percentage, brightness, opacity, etc. Paperboards are produced in rolls and these parameters are measured on each roll. Many of the quality characteristics are interrelated and hence the quality of the board must be judged on the basis of the multivariate response. The main problem that the company is facing is to find an answer to the question: How to rate the quality of each board produced. This project is undertaken with this aim of evolving a system to rate the boards produced using a multivariate response approach. For the purpose of the study, it has been decided to evolve a system for the APM65 paperboard to start with. Once an effective system is evolved, the same can be extended to all the boards.



Fig. 7.26 Paperboards are used to make bags, cartons, composite containers and tubes, cups, disposables, envelopes, office files and so on. The Paper and Paperboard market size (2007) had a value of 630.9 billion USD and a volume of 320.3 million metric tons. About 50% of all produced paper is used for packaging, followed by printing and writing. Cartons make up one third of paper and board packaging and 15 % of all packaging. <http://en.wikipedia.org/wiki/Paperboard#citerefind5-0>, cited Nov 12, 2014

7.5.2 Approach

To start with, a meeting was held with all concerned technical personnel and a brainstorming session was carried out. Following this, the important quality characteristics were identified and short listed. Though most of these quality characteristics had specifications, from the past data analysis it was found that no boards were observed with all the characteristic values within specification limits. From this it was not clear whether it is possible to produce ideal boards (an ideal board is contemplated as a board with all its quality characteristics within the specifications). One possibility is that the interrelationships among the quality characteristics may perhaps restrict feasibility of meeting the individual specifications. Therefore, it was decided to treat rolls close to specifications as good quality boards. In fact, when the technical personnel were asked to identify some good rolls from the past, they chose rolls with quality characteristics close to specifications and did not have complaints from customers. Using multivariate data on the past rolls, we first classify the rolls as **Good** (good quality rolls) and **Poor** (poor quality rolls) using exploratory data analysis techniques. Here, we use Euclidean distance from the target specifications as a measure to identify the good and poor rolls. Then Discriminant Analysis is used to see if there is a clear distinction between Good and Poor rolls using only the multivariate data. If this works well, we can use the discriminant function to classify future rolls as good or poor and use Mahalanobis distance [7, 12] as a measure of quality index.

The discriminant function [6] is a function of the quality characteristics and is computed from the observed data. Suppose x_1, x_2, \dots, x_k are the quality characteristics. We observe the values of x_1, x_2, \dots, x_k for each of the good and poor rolls. Using these observations, two discriminant functions, $f(x_1, x_2, \dots, x_k)$ and $g(x_1, x_2, \dots, x_k)$ are computed for good and poor rolls respectively. Given the values (x_1, x_2, \dots, x_k) for a roll, we can then compute two pairs of measures (d, p) – one using f and the other using g . Here d stands for squared distance and p for posterior probability. If p is more for f compared to g , then we treat the roll as good, other wise we consider the roll to have some quality problem. In order to validate the efficacy of the discrimination model, we set aside some good and some poor quality rolls from the selected rolls, and cross check if they will be correctly classified by the model. If the model is efficient, then we pool all the good rolls and build the Mahalanobis distance function to characterize the good rolls.

7.5.3 Data Collection and Analysis

The APM65 paperboard is taken up for the purpose of this study. A list of the variables along with their specifications identified in the meeting with the technical personnel is presented in Fig. 7.27. Data on 62 rolls were collected from the past production. The measurements on the variables selected for the study were compiled from the inspection records. When asked, the technical personnel identified

S.No.	Variable	Symbol	Specifications	
			Min	Max
1	GSM Avg.	x_1	57	63
2	Std.Dev. Of GSM	x_2	68.4	75.6
3	Caliper Avg.	x_3		
4	Std.Dev. Of Caliper	x_4		
5	Bulk Avg.	x_5	1.2 Avg.	
6	Std.Dev. Of Bulk	x_6		
7	Bursting Strength Avg. (BS Avg.)	x_7		
8	Bursting Factor Avg. (BF Avg.)	x_8	18	
9	Tearing Strength	x_9		
10	Tearing Factor	x_{10}	48	
11	TS MD	x_{11}		
12	TS CD	x_{12}		
13	BL CD	x_{13}	2500	
14	Cobb TS	x_{14}	23	27
15	Cobb WS	x_{15}	23	27
16	Brightness	x_{16}	88	92
17	Opacity	x_{17}	80	83
18	Ash %	x_{18}		
19	Moisture %	x_{19}	5%	6%
20	Smooth TS Avg.	x_{20}	120	160
21	Smooth TS Std.Dev.	x_{21}		
22	Smooth WS Avg.	x_{22}	200	250
23	Smooth WS Std.Dev.	x_{23}		
24	Porosity Avg.	x_{24}		600
25	Formation Index	x_{25}	120	
26	L Value	x_{26}	90	
27	A Value	x_{27}	4.5	
28	B Value	x_{28}	-9.7	
29	Whiteness	x_{29}	123	
30	Delta E	x_{30}		1

Note: The specification are for the individuals

Fig. 7.27 Quality characteristics with specifications

34 of these rolls as good quality rolls based on the closeness to specifications and customer feedback. The balance 28 rolls will be referred to as *unspecified* rolls. Since we are looking for good quality rolls, it was expected that rolls with characteristic values close to specifications would be of good quality. With this in mind, taking the target specifications as the reference points, the distance (the Euclidean distance) is computed for each roll. While computing this distance, only those characteristics which have specifications were considered (see Fig. 7.27). A histogram of the distances is shown in Fig. 7.28. The histogram suggests a bimodal distribution. Taking a clue from this, the rolls are classified into two categories. Those rolls with distance less than 1200 are labeled as G (for good), and the others

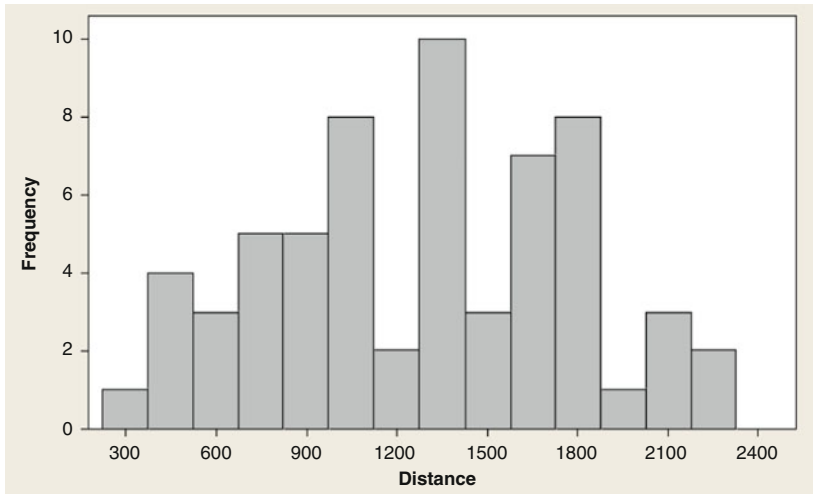


Fig. 7.28 Histogram of distances from target specifications

are labeled as P (for poor). This classification of G and P must not be confused with the Good and unspecified as judged by the technical personnel. Taking the classification of rolls by G and P as the category response variable and all the variables listed in Fig. 7.27 as predictor variables, the discriminant analysis was performed. Before performing the discriminant analysis, the calibration technique was adopted to validate the power of discrimination. For this purpose, 12 rolls out of the 62 were selected randomly and set aside for cross checking. The discriminant analysis was performed using the remaining 50 rolls' data. The summary of the discriminant analysis is shown in Fig. 7.29. Out of the 12 test rolls set aside for cross checking, six are from those labeled as G and the remaining six are from those labeled as P. In the cross validation only one roll was wrongly classified into P where as it has actually come from G. The summary of the classification of these 12 rolls is shown in Fig. 7.30. Canonical variables are the linear functions of the original variables. The biplot plots the original data using the first two canonical functions that contain maximum information about the data on the original variables. This plot can be used to see the distance between the two groups (designated by G and P). Figure 7.31 presents the biplot for the data on the 50 rolls used for building the discrimination model. From the figure, there is a clear demarcation between the two groups (Figs. 7.32–7.34).

7.5.4 Recommendations

From the analysis of the data, it seems that using the Euclidean distance from the target specifications seems to work well to derive a measure of quality index for the

Discriminant Analysis

Linear Method for Response: Index
Predictors: GSM Avg,GSM Stdev, Cal Avg, Cal Stdev,
Bulk Avg, Bulk Stdev, B.S Avg, BF Avg, Tearing Strenght,
Tear Factor, TS MD, TS CD, BL MD, BL CD, Cobb TS,
Cobb WS, Brightness, Opacity, Ash%, Moistures,
Smooth TS Avg, Smooth TS SD, Smooth WS Avg,
Smooth WS SD, Porosity Avg, Formation Index,
L value,A value,B value,Whiteness,Delta E

Group	G	P
Count	21	29

Summary of classification

Put into Group	True Group	
	G	P
G	21	0
P	0	29
Total N	21	29
N correct	21	29
Proportion	1.000	1.000

N = 50 N Correct = 50

Proportion Correct = 1.000

Squared Distance Between Groups

	G	P
G	0.0000	50.3152
P	50.3152	0.0000

Fig. 7.29 Summary of discriminant analysis

rolls. However, it is suggested that the management should further examine the rolls classified here as G and P more closely and try to find out the differences between the two groups. For the present, it may be recommended that the rule derived in the study may be implemented for the APM65 board. It is also proposed that a meeting can be held with all concerned personnel to discuss further on the findings of this study to examine the suitability of the approach suggested in this article.

Prediction for Test Observations

Observation	Pred Group	From Group	Squared Distance	Probability
1	G			
		G	105.605	1.000
		P	164.657	0.000
2	G			
		G	67.476	1.000
		P	102.558	0.000
3	G			
		G	119.112	1.000
		P	194.113	0.000
4	P			
		G	139.076	0.000
		P	112.777	1.000
5	G			
		G	107.703	1.000
		P	229.554	0.000
6	G			
		G	123.572	1.000
		P	151.811	0.000
7	P			
		G	89.664	0.002
		P	77.664	0.998
8	P			
		G	74.633	0.000
		P	57.581	1.000
9	P			
		G	176.343	0.000
		P	122.770	1.000
10	P			
		G	178.109	0.000
		P	67.259	1.000
11	P			
		G	287.934	0.000
		P	233.505	1.000
12	P			
		G	167.386	0.000
		P	71.096	1.000

Fig. 7.30 Summary of classification of 12 test rolls

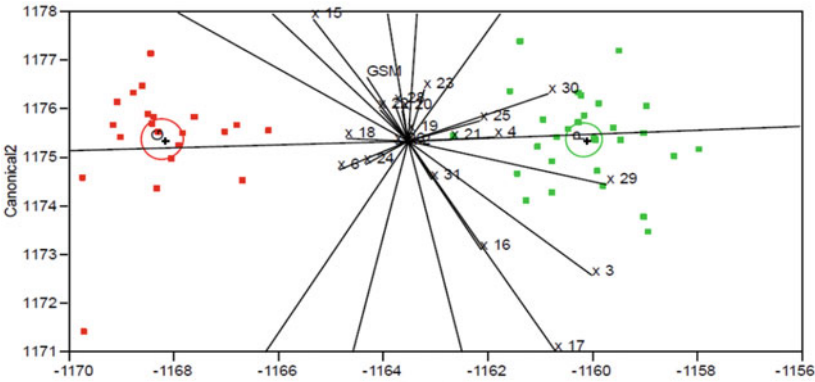


Fig. 7.31 Biplot of the first two canonical functions for the model data of 50 rolls. Classification is based the distance of 1200

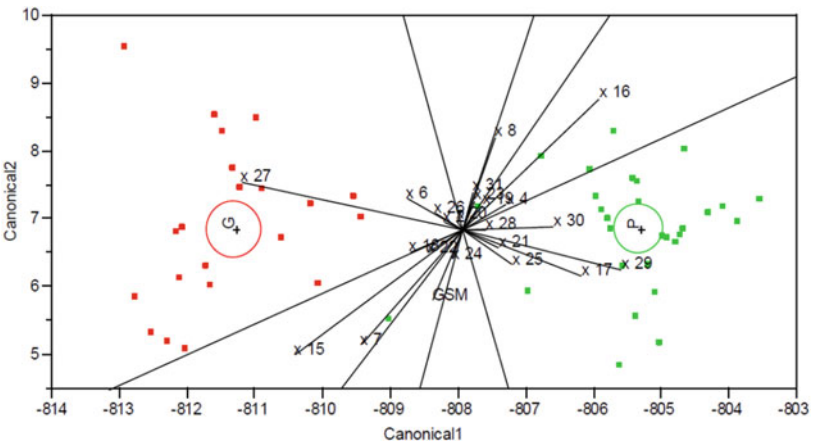


Fig. 7.32 Biplot of the first two canonical functions for the model data of 50 rolls. Classification is based the distance of 1300

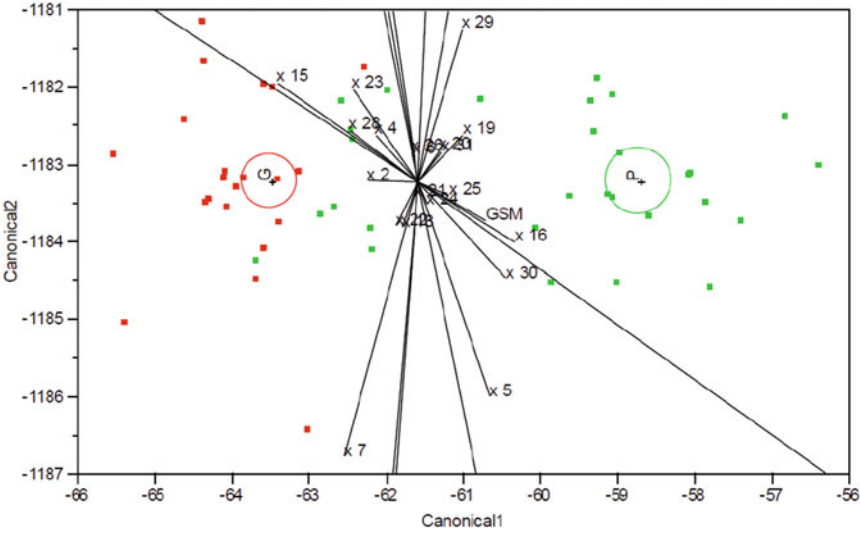


Fig. 7.33 Biplot of the first two canonical functions for the model data of 50 rolls. Classification is based the distance of 1500

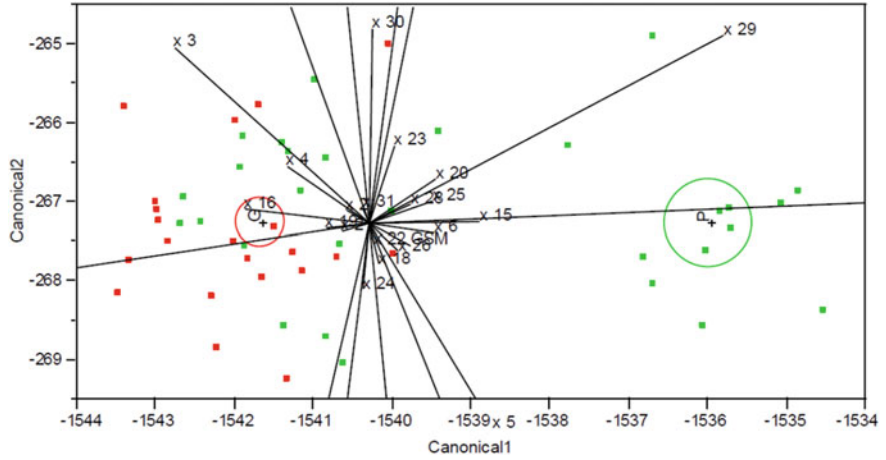


Fig. 7.34 Biplot of the first two canonical functions for the model data of 50 rolls. Classification is based the distance of 1700

7.6 Which Pension Scheme Is Better?

This study was done for an individual – an employee who is on the verge of his retirement. The employer offers two pension schemes to his employees and the employees can choose any one of the schemes. The individual we are talking about wanted to know which of the two schemes will be beneficial for him. Perhaps there may be many ways to find an answer to this decision making problem but the solu-

tion presented to this problem in this section will be interesting because the solution is a nice application of Markov Chains and the approach can be extended to many industrial problems.

7.6.1 *The Pensioner's Problem*

In the schemes below, the retiring employee will be called the pensioner, and his wife will be referred to as his survivor.

Scheme 1. The pensioner will receive a monthly pension of \$ 4783 as long as he is alive and upon his death his survivor will receive a monthly pension of \$ 3500 and one of his sons declared as his beneficiary will receive a monthly pension of \$ 1195. The survivor and beneficiary will get the pension as long as they live.

Scheme 2. The pensioner will get \$ 4486 per month as long as he is alive; and upon his death, his survivor will get \$ 2285 per month and each of his two sons will get \$ 1090 per month. The survivor and sons get their dues as long as they live.

7.6.2 *A Solution*

As the returns under two schemes depend upon the survival chances, the usual way to solve this problem looks at the expected returns. The survival chances depend on the life expectancies of the pensioner, survivor and the two sons. Unfortunately, no information is available on this. However, the problem can be modeled with unknown parameters, the survival chances, and solved for different possible values of the parameters. We can visualize the problem as a Markovian model [4, 5] which is briefly described in the next subsection.

7.6.3 *Markovian Model for the Problem*

To start with, let us first consider Scheme 1. Consider the system of persons involved in the problem under Scheme 1. The persons are: pensioner, survivor and the beneficiary. We can think the system as the survival positions of the three persons in question. With this, there are eight possible states for the system, and at any given time, the system is in one of these eight states. One of the possible states is that all the three are alive. We can denote this state with a 3-tuple lll , where the first letter of this 3-tuple stands for the survival position of the pensioner, the second letter for that of survivor and the third letter for that of beneficiary. Here 'l' stands for 'live.' Another possible state of the system is dll meaning that the pensioner is no more (d for 'dead') but the survivor and the beneficiary are alive. For mathematical treatment, the states are numbered as 1–8. Note that the amount received by the family in

any month depends only on the state of the system (see problem 7.5). If the system is in state *lll* or equivalently in state 1, the amount is \$ 4783, and if the system is in state *ddl* (state 7), the amount will be \$ 1195. The eight states, their numbers and the rewards (amounts received) are listed in Table 7.12.

Table 7.12 States of system under Scheme 1 and rewards

State	<i>lll</i> (1)	<i>lld</i> (2)	<i>ldl</i> (3)	<i>ldd</i> (4)	<i>dll</i> (5)	<i>dld</i> (6)	<i>ddl</i> (7)	<i>ddd</i> (8)
Reward (USD)	4783	4783	4783	4783	4695	3500	1195	0

Since the reward is received every month, the time horizon is taken as the first month (1), second month (2), third month (3), and so on. Let $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$ denote the state space of the system. Therefore, on any given month, the system is in j , where $j \in S$. Let X_k be the state of the system at the end of k th month from the time of retirement, $k = 0, 1, 2, \dots$. Here $k = 0$ stands for the beginning of first month immediately after retirement, $k = 1$ stands for end of first month or equivalently the beginning of second month, and so on. It is assumed that the initial state of the system is *lll* or $X_0 = 1$. Clearly, $\{X_k, k = 0, 1, \dots\}$ is a finite Markov chain over the state space S .

Let $P^{(k)}$ $k = 0, 1, \dots$ be the k -step transition matrix of the chain. That is, the (i, j) th element of $P^{(k)}$, denoted by $p_{ij}^{(k)}$, is the conditional probability that the system will be in state j at the end of month k given that the system is in state i at the beginning of month 0. In other words, $p_{ij}^{(k)} = \text{Prob}\{X_k = j | X_0 = i\}$. In general, the one-step transition probabilities depend on the time as well. That is, $\text{Prob}\{X_{k+1} = j | X_k = i\}$ depends on k as well (besides its dependence on i and j). But, if these one-step transition probabilities are independent of k , then the Markov chain is called a stationary Markov chain in which case the common one-step transition matrix is given by the probabilities $\text{Prob}\{X_1 = j | X_0 = i\}$.

To find an answer to the pensioner's problem, we shall assume that the Markov chain in question is a stationary one. This may not badly affect the conclusions. The elementary results on Markov chain assert that for a stationary Markov chain with one-step transition matrix P , the k -step transition matrix $P^{(k)}$ is equal to P^k (P raised to the power k). Using this result, it is easy to obtain the expected rewards over any give period of time.

Note that the state 8 (*ddd*) is an *absorbing state* of the system and the rest of the states are *transient*. The one-step transition matrix P is shown in Fig. 7.35. The first row of the figure shows the pension amounts for each state. The elements p, q, r, p', q' and r' in the matrix P are the following: p is the conditional probability of 'given that the pensioner is alive in a month, he will be alive in the next month'; q and r are similarly defined for the survivor and beneficiary respectively; $p' = 1 - p$, $q' = 1 - q$ and $r' = 1 - r$.

Let us now compute the expected pension to the family over different months. Let a_j be the pension to the family when the system is in state j . The a_j s are shown in the first row of Fig. 7.35. The starting state of the system is given by $X_0 = 1$. The expected pension on the first month is given by

Pension	4783	4783	4783	4783	4695	3500	1195	0
States	<i>lll</i> 1	<i>lld</i> 2	<i>ldl</i> 3	<i>ldd</i> 4	<i>dll</i> 5	<i>dld</i> 6	<i>ddl</i> 7	<i>ddd</i> 8
<i>lll</i> 1	pqr	pqr'	pq'r	pq'r'	p'qr	p'qr'	p'q'r	p'q'r'
<i>lld</i> 2	0	pq	0	pq'	0	p'q	0	p'q'
<i>ldl</i> 3	0	0	pr	pr'	0	0	p'r	p'r'
<i>ldd</i> 4	0	0	0	p	0	0	0	p'
<i>dll</i> 5	0	0	0	0	qr	qr'	q'r	q'r'
<i>dld</i> 6	0	0	0	0	0	q	0	q'
<i>ddl</i> 7	0	0	0	0	0	0	r	r'
<i>ddd</i> 8	0	0	0	0	0	0	0	1

Fig. 7.35 One-step matrix with rewards for Scheme 1

$$R_1 = \sum_{j=1}^8 a_j p_{1j} = (Pa)_1, \quad (7.11)$$

where $a = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)^t$ and $(Pa)_1$ is the first coordinate of the vector Pa . The expected pension for the second month is given by

$$R_2 = \sum_{j=1}^8 a_k p_{1j}^{(2)} = (P^2 a)_1. \quad (7.12)$$

The expected pension of k th month is given by

$$R_k = \sum_{j=1}^8 a_j p_{1j}^{(k)} = (P^k a)_1, \quad j = 1, 2, \dots \quad (7.13)$$

The total expected pension (undiscounted) over first n months is given by

$$\sum_{k=1}^n R_k = ([P + P^2 + P^3 + \dots + P^n]a)_1. \quad (7.14)$$

Let g_n denote the average expected pension per month based on the first n months expected pension under Scheme 1. That is,

$$g_n = \frac{\sum_{k=1}^n R_k}{n}. \quad (7.15)$$

Evaluation of Scheme 2

Under Scheme 2, the system has 16 states, represented by 4-tuples of the letters l and d . The states, pensions and the one-step transition matrix are shown in Fig. 7.36. Here p is the conditional probability that the pensioner will be alive in the following month given that he is alive in a given month; q is the conditional probability that the survivor will be alive in the following month given that she is alive in a given month; $r(s)$ is the conditional probability that the first (second) son will be alive in the following month given that he is alive in a given month. Let $p' = 1 - p$, $q' = 1 - q$, $r' = 1 - r$ and $s' = 1 - s$. The reward vector for Scheme 2 is shown in Fig. 7.36. The expected pension for month k is computed using formula given in Eq. (7.13) with the new one-step matrix and reward vector given in Fig. 7.36.

Let h_n denote the average expected pension per month based on the first n months expected pension under Scheme 2. The h_n is computed using the right hand side expression of the formula given by Eq. (7.15).

7.6.4 Comparison of Two Schemes

The two schemes can be compared by comparing the average expected monthly rewards g_n and h_n for $n = 1, 2, \dots$, for the same set of life expectancy parameters. We shall consider two sets of parameters for this comparison and present the results for each of these sets. The computations are carried out in Matlab.

Case 1: In this case we take $p = 0.99$, $q = 0.995$ and $r = s = 0.99999$. This may be considered as the case where the health of the parents is not so sound, wife's being marginally better than the husband.

Case 2 In this case we take $p = 0.995$, $q = 0.995$ and $r = s = 0.999999$. This is a slightly better case compared to Case 1.

For Case 1, the expected average monthly incomes are shown in Table 7.13; and a graphical comparison of the two schemes is shown in Fig. 7.37. We see that in this case, Scheme 1 is better up to 150 months, and if you are looking for average monthly rewards beyond 150 months, then Scheme 2 is better.

For Case 2, the expected average monthly incomes are shown in Table 7.14; and a graphical comparison of the two schemes is shown in Fig. 7.38. In this case, Scheme 1 is better up to 219 months and beyond that Scheme 2 is better.

Pension	4486	4486	4486	4486	4486	4486	4486	4486	4465	3375	3375	2285	2180	1090	1090	0
States	llll	llld	lldl	lldd	ldll	ldld	lddl	lddd	dlld	dlld	dlld	dlld	dlld	dlld	dlld	dlld
llll	pqr's	pqr's	pqr's	pqr's	pqr's	pqr's	pqr's	pqr's	p'qrs	p'qrs	p'qrs	p'qrs	p'qrs	p'qrs	p'qrs	p'qrs
llld	0	pqr	0	pqr	0	pqr	0	pqr	0	p'qr	0	p'qr	0	p'qr	0	p'qr
lldl	0	0	pqs	pqs	0	0	pq's	pq's	0	0	p'qs	p'qs	0	0	p'qs	p'qs
lldd	0	0	0	pq	pq	0	0	0	0	0	0	p'q	0	0	0	p'q
ldll	0	0	0	0	prs	prs	prs	prs	0	0	0	0	p'rs	p'rs	p'rs	p'rs
ldld	0	0	0	0	0	pr	0	pr	0	0	0	0	0	p'r	0	p'r
lddl	0	0	0	0	0	0	ps	ps	0	0	0	0	0	0	p's	p's
lddd	0	0	0	0	0	0	0	p	0	0	0	0	0	0	0	p'
dlld	0	0	0	0	0	0	0	0	qrs	qrs	qrs	qrs	q'rs	q'rs	q'rs	q'rs
dlld	0	0	0	0	0	0	0	0	qr	0	qr	0	q'r	0	q'r	q'r
dlld	0	0	0	0	0	0	0	0	0	qs	qs	0	0	0	q's	q's
dlld	0	0	0	0	0	0	0	0	0	0	0	q	0	0	0	q'
dlld	0	0	0	0	0	0	0	0	0	0	0	0	rs	rs	r's	r's
dlld	0	0	0	0	0	0	0	0	0	0	0	0	0	r	0	r'
dlld	0	0	0	0	0	0	0	0	0	0	0	0	0	0	s	s'
dlld	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Fig. 7.36 One-step matrix with rewards for Scheme 2

Table 7.13 Average monthly rewards under Schemes 1 and 2 under Case 2 parameters

Month	Scheme 1	Scheme 2	Month	Scheme 1	Scheme 2	Month	Scheme 1	Scheme 2
1	4781	4486	22	4737	4466	43	4659	4424
2	4780	4485	23	4734	4465	44	4655	4422
3	4779	4485	24	4730	4463	45	4650	4419
4	4777	4485	25	4727	4461	46	4646	4417
5	4776	4484	26	4724	4460	47	4642	4414
6	4774	4484	27	4721	4458	48	4637	4411
7	4772	4483	28	4717	4456	49	4633	4409
8	4771	4482	29	4714	4454	50	4628	4406
9	4769	4482	30	4710	4452	51	4623	4404
10	4767	4481	31	4707	4450	52	4619	4401
11	4765	4480	32	4703	4448	53	4614	4398
12	4763	4479	33	4699	4446	54	4609	4395
13	4760	4478	34	4696	4444	55	4605	4393
14	4758	4477	35	4692	4442	56	4600	4390
15	4756	4476	36	4688	4440	57	4595	4387
16	4753	4475	37	4684	4438	60	4590	4384
17	4751	4473	38	4680	4436	120	4585	4381
18	4748	4472	39	4676	4433	150	4069	4067
19	4745	4471	40	4672	4431	151	4063	4064
20	4742	4469	41	4668	4429	180	3897	3960
21	4740	4468	42	4663	4426	240	3578	3759

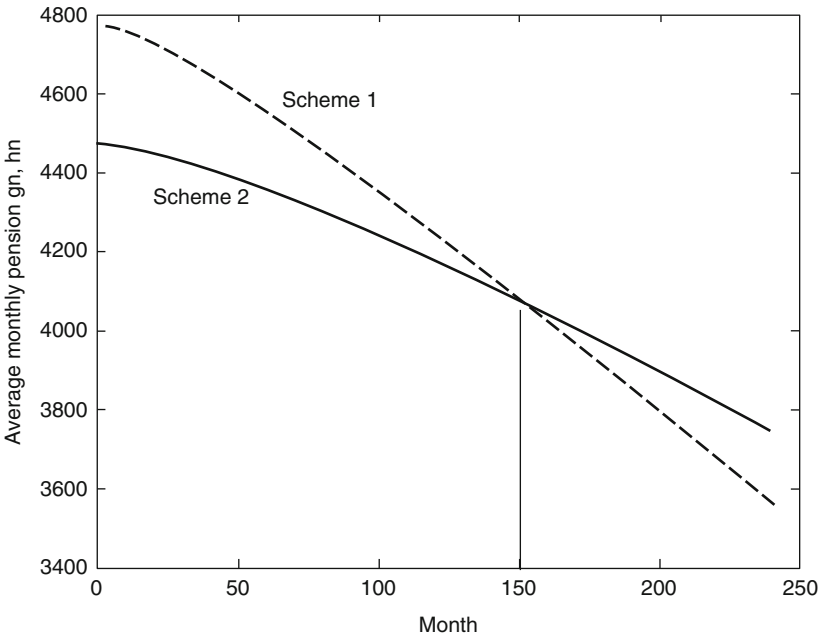


Fig. 7.37 Comparison of schemes under Case 1 ($p = 0 : 99$; $q = 0 : 995$ and $r = s = 0 : 99999$)

Table 7.14 Average monthly rewards under Schemes 1 and 2 under Case 2 parameters

Month	Scheme 1	Scheme 2	Month	Scheme 1	Scheme 2	Month	Scheme 1	Scheme 2
1	4782	4486	22	4759	4476	43	4716	4453
2	4781	4486	23	4757	4475	44	4714	4451
3	4781	4486	24	4756	4474	45	4712	4450
4	4780	4485	25	4754	4473	46	4709	4449
5	4779	4485	26	4752	4472	47	4707	4447
6	4779	4485	27	4750	4471	48	4704	4446
7	4778	4485	28	4749	4470	49	4701	4444
8	4777	4484	29	4747	4469	50	4699	4443
9	4776	4484	30	4745	4468	51	4696	4441
10	4775	4483	31	4743	4467	52	4693	4440
11	4774	4483	32	4741	4466	53	4691	4438
12	4773	4482	33	4739	4465	54	4688	4436
13	4772	4482	34	4737	4464	55	4685	4435
14	4770	4481	35	4735	4463	56	4682	4433
15	4769	4481	36	4732	4462	57	4680	4432
16	4768	4480	37	4730	4460	60	4677	4430
17	4766	4479	38	4728	4459	120	4459	4299
18	4765	4479	39	4726	4458	150	4335	4223
19	4764	4478	40	4724	4457	180	4208	4144
20	4762	4477	41	4721	4455	219	4040	4040
21	4761	4477	42	4719	4454	240	3951	3984

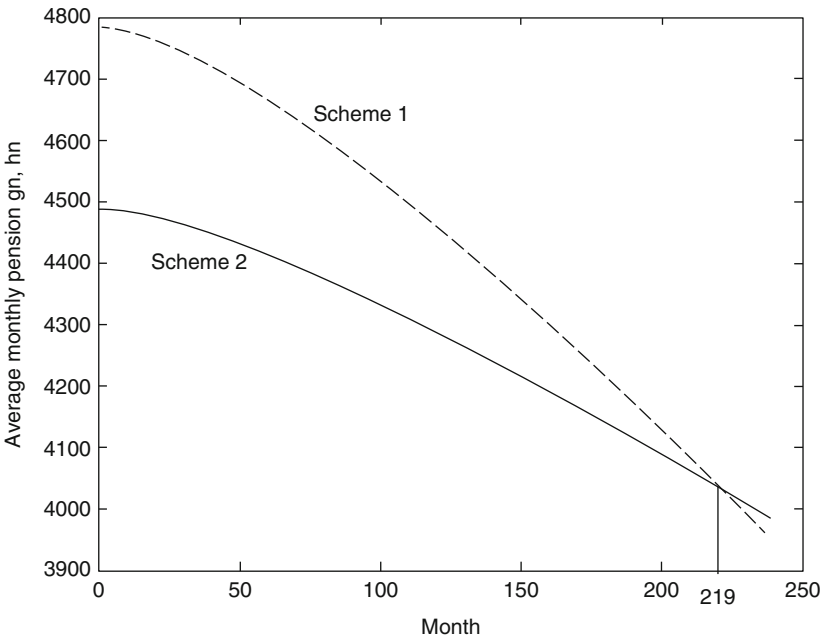


Fig. 7.38 Comparison of schemes under Case 2 ($p = q = 0 : 995$ and $r = s = 0 : 999999$)

7.6.5 Summary of Pernsioner’s Problem

In this section we have seen a nice application of Markov models to a simple problem raised by a pensioner. Though essential input data (on life expectancy parameters) are not provided, the problem could be given a solution. Using the solution, the user can evaluate the answers under various scenarios (different possibilities of life expectancy parameters) and weigh his options. The solution methodology will be useful to many live industrial problems.

7.7 Chapter Summary

This chapter presented five live case studies. These talk about the variety of problems that a consultant comes across. Some may be simple to answer but the solutions provided for these case studies help in logical thinking and formulation and providing intelligent solutions where necessary. The first one (Sect. 7.2) on determining frequencies of amino acid molecules present in a polypeptide is given an elegant solution using a simple computer program. The second (Sect. 7.3) and fourth (Sect. 7.5) are applications of multivariate methods to quality evaluation problems of products based on several product characteristics. The third one (Sect. 7.4) is

about evaluation of reliability of CFL bulbs and it provides an answer to important question regarding the reliability of bulbs. The last case study (Sect. 7.6) is actually an answer to a problem raised by an employee who is on the verge of his retirement and provides an answer to his question of which pension scheme he should choose.

Problems

7.1. Verify the one-step transition matrices given in Figs. 7.35 and 7.36. Write a code in Matlab or any other language you know to compute P^k for the two schemes and compute P^{60} (end of 5 years), P^{120} (end of 10 years), P^{180} (end of 15 years), and P^{240} (end of 20 years). What are the study state probabilities?

7.2. Consider Scheme 1 (in the Pensioner's problem). Let u_1 , u_2 and u_3 be three binary variables defined as follows: $u_1 = 1$ if pensioner is alive, $u_1 = 0$ other wise; $u_2 = 1$ if survivor is alive, $u_2 = 0$ other wise; $u_3 = 1$ if beneficiary is alive, $u_3 = 0$ other wise. Observe that $u_1 u_2 u_3$ s will correspond to the eight states of the system. Show that the pension corresponding to $u_1 u_2 u_3$ is given by $4783u_1 + (1 - u_1)(3500u_2 + 1195u_3)$. Develop a similar formula for Scheme 2.

7.3. Consider the CFL bulbs case study. Assume that the manufacturing cost of each bulb is Rs.45/-. If the company has to make a profit of Rs.25/- on each bulb, what should be the selling price of each bulb. Analyse this problem carefully and find the answer (Hint: The selling price is not Rs.70/-).

7.4. Write a computer programme to determine the number of test benches using Weibul distribution with derived parameters as in Sect. 7.4.

7.5. This problem is about the stationarity assumption of the Markov chains used. Let Q_k be the one-step transition matrix of transition probabilities from month $k - 1$ to month k , $k \geq 1$ in the pensioner's problem of Sect. 7.6. That is, the (i, j) th element of Q_k is equal to $\text{Prob}\{X_k = j | X_{k-1} = i\}$. Show that $P^{(k)} = Q_1 Q_2 \dots Q_k$. Assess the difference between the stationarity assumption and without it.

References

1. Bruni, R.: A Combinatorial Optimization Approach for Determining Composition and Sequence of Polymers. Department of Electronic and Information Engineering, University of Perugia, Perugia
2. Chen, D.-S., Batson, R.G., Dang, Y.: Applied Integer Programming: Modeling and Solution. Wiley, Hoboken (2010)
3. Flury, B.: Common Principle Components Analysis and Related Multivariate Models. Wiley, New York (1998)
4. Hoel, P.G., Port, S.C., Stone, C.J.: Introduction to Probability Theory. Universal Book Stall, New Delhi (2000)

5. Hoel, P.G., Port, S.C., Stone, C.J.: *Introduction Stochastic Processes*. Haughton Mifflin, Boston (1972)
6. Johnson, R.A., Wichern, D.W.: *Applied Multivariate Analysis*, 4th edn. Prentice Hall, Englewood Cliffs (1998)
7. Mahalanobis, P.C.: On the generalised distance in statistics. *Proc. Natl. Inst. Sci. India* **2**(1), 49–55 (1936)
8. Montgomery, D.C.: *Statistical Quality Control – A Modern Introduction*, 6th edn. Wiley, New York (2009)
9. Montgomery D.C., Friedman, J.J.: Statistical process control in a computer-integrated manufacturing environment. In: Keats, J.B., Hubele, N.F. (eds.) *Statistical Process Control and Automated Manufacturing*. Series in Quality and Reliability. Dekker, New York (1989)
10. Morrison, D.F.: *Multivariate Statistical Methods*. McGraw Hill, New-York (1990)
11. Muirhead, R.J.: *Aspects of Multivariate Statistics*. Wiley, New York (1982)
12. Nemhauser, G.L., Wolsey, L.A.: *Integer and Combinatorial Optimization*. Wiley, New York (1998)
13. Ramanathan, G., Kettenring, J.R.: Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28**, 81–124 (1972)
14. Taha, H.A.: *An Introduction to Operations Research*. Prentice Hall, New Jersey (2007)

Chapter 8

Modelling Optimum Utilization of Irrigation Resources

Abstract This chapter deals with a case study on the development of a model for the management of irrigation system. The motivation for this case study was a severe crisis faced by farmers of the Pennar Delta in Andhra Pradesh, India. This work was taken up at the request of a Superintending Engineer of the concerned Irrigation Department. Upon educating the engineer about the utility and effectiveness of optimization techniques for his problem, the engineer came forward for exploring a scientific approach and a mathematical model for finding a solution to the problem of planning water resources and scheduling the distribution of reservoir waters through the irrigation channels of the Pennar Delta. Given the requirements, availabilities of water at various nodes and the particulars of the irrigation network of the system, the problem can be formulated as a multicommodity network flow problem. This chapter presents an approach for solving the problem. The problem is formulated as a dynamic minimum cost network flow problem and provides an approach to solve the problem using static network flow models. Due to non-availability of required data, the scope of the work was confined to modeling the problem. This work also helped in training a post graduate student on live applications of modeling practical problems.

8.1 Introduction

This chapter deals with a case study on the development of a model for the management of irrigation system. A preliminary version of this chapter can be seen in [10]. The motivation for this case study was a severe crisis faced by farmers of the Pennar Delta in Andhra Pradesh, India. This work was taken up at the request of a Superintending Engineer of the concerned Irrigation Department. Upon educating the engineer about the utility and effectiveness of optimization techniques for his problem, the engineer came forward for exploring a scientific approach and a mathematical model for finding a solution to the problem of planning water resources

and scheduling the distribution of reservoir waters through the irrigation channels of the Pennar Delta. Given the requirements, availabilities of water at various nodes and the particulars of the irrigation network of the system, the problem can be formulated as a multicommodity network flow problem. This chapter presents an approach for solving the problem. The problem is formulated as a dynamic minimum cost network flow problem and provides an approach to solve the problem using static network flow models. Due to non-availability of the required data, the scope of the work was confined to modeling the problem. This project was also used for providing practical training to a post graduate student. Besides training the student on modelling the problem, he was assigned the task of developing a programme for solving network flow problems. This work is outlined in Sect. 8.5.

Efficient management of water for irrigation plays crucial role in countries which are facing severe water crisis. Several authors have studied the need and importance of water management for irrigation purposes [3, 13, 15]. The irrigation water management requires good statistical data maintenance systems and information technology for effective decision making and implementation. Prior to early 1990s the information technology was in the developing stages and this has partly been responsible for the lack of efficient statistical systems. One of the other main causes for this lacuna is the lack of awareness (among the administrators) of scientific methods and their necessity in decision making processes and management. But, with the level of advancement and growth of information technology that we have today, we should quickly design, develop and establish efficient statistical systems and implement them using electronic media so that our resources are best utilized. This can be achieved only if the people in the scientific and governing communities come together and put conscious and untiring efforts to bridge the gaps.

The prosperity of a country like India where agricultural produce is about 14 % of its GDP depends largely on efficient management of water resources and proper irrigation methods. Therefore, optimum utilization of resources pertinent to agricultural sector is very important for the growth. A large part of India's agricultural fields are fed by irrigation (http://planningcommission.nic.in/plans/planrel/fiveyr/11th/11_v3/11v3_ch2.pdf). Water supplied from rivers and reservoirs to irrigated lands is undoubtedly the most important of agricultural resources. Due ecological disturbances, the availability of water is becoming increasingly more difficult year after year. Therefore, it is utmost important for us to manage irrigation water in the best possible way [17]. This becomes particularly more relevant in the present day's global scenario of depleting water resources and acute water scarcity problem [13].

Pennar Delta ayacut (irrigation lands) of Nellore district in Andhra Pradesh serves about 2,50,000 acres of agricultural lands. Due to severe summer and draught conditions, this ayacut had faced acute water shortage problem during the past and had led to agitations by the farmers and the district administration was under tremendous pressure to resolve the imbroglio. Added to the acute water shortage problem, lack of adequate information on actual amount of water available and the water requirements (determined by various crop patterns and their extents) had complicated the problem. Motivated by this imbroglio, a proposal was made to the

concerned authorities in the district administration. Consequently this project has been initiated. Typically a project of this nature has to address a number of issues such as development of systems and procedures for acquiring and monitoring necessary data and information on electronic media, making the same available to all concerned personnel for planning and monitoring, developing efficient optimization tools for managing the resources, training the personnel in the administration, educating the farmers, advising and guiding them about crop patterns and scheduling etc. A complete solution to this problem requires concentrated efforts by a team of specialists from different areas and other resources such as financial support from the Government, etc. Optimization techniques have been used to model the irrigation water utilization [6, 7, 11, 14]. We have initiated the project with a limited scope so as to create awareness among the administrators and induce them into taking up the project in its entirety. This chapter presents the problem and the optimization model for the same. The details are elaborated in the next section.

8.2 The Irrigation System

The model irrigation system considered for this work has been derived from the Penar Delta System (PDS). A major or medium irrigation system such as this consists of a main reservoir, main channels, branch channels, distributary channels and several storage and/or regulation points. Water is carried to the agricultural fields (AFs) through these channels. A pictorial description of the system is shown in Fig. 8.1.

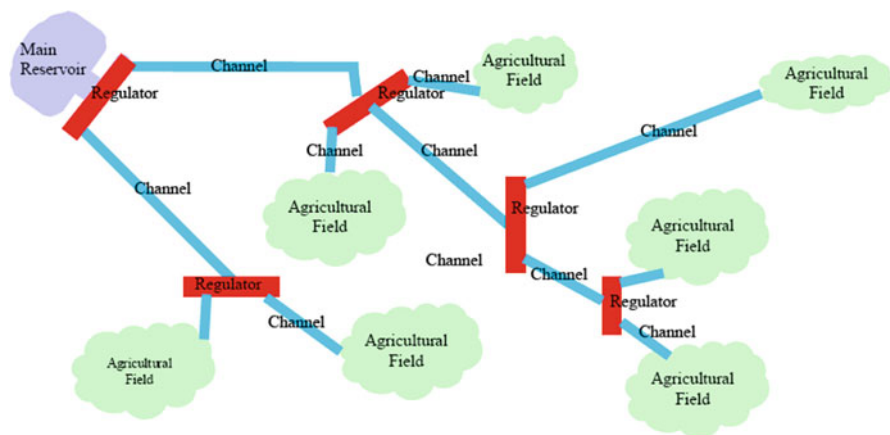


Fig. 8.1 Pictorial description of reservoir irrigation system

The collection of all AFs is called the project ayacut. Each channel is designed for a specific capacity to meet its intended requirements. The channel parameters such as cross sectional area, slope etc., will determine the delivery rates (usually

measured in cubic feet per second (cusecs)). The release of water in any channel is controlled by the regulators by opening their gates partially or fully. Generally the regulators have storage capacities. A number of factors such as crop pattern variation, timing of crops, soil conditions (water absorption properties), etc., determine the water requirements of the agricultural fields. The normal practice of supplying water to AFs is to declare the opening and closing dates of various regulators. This results in uniform supply of water to AFs which may be detrimental to the crops. This is because the water requirements of crops vary with different growth stages of the crops.

An important feature of the irrigation system under consideration is that every AF is connected to one and only one channel in the network. Technically speaking, the irrigation network is a tree in Graph Theory terminology [12, 16].

One of the constraints in releasing water at a regulator into a channel is that there should be certain minimum quantity of water at the regulator so that the required water head is there for the water to flow into the channel. For example, at Nellore regulator there are two branch channels. One of these two channels, the Kanupur canal, is at a higher altitude than the other. In order to supply water in this channel, one has to build up the water head above the channel bed level at the regulator and then release the water. The calculated quantity of water supplied in a channel and the actual quantity of water supplied are usually at variance. There are several reasons for this. Though the channels are designed for predetermined capacities, the actual capacities may be at variance with designed capacities. Silt formation in the channels, lack of proper maintenance, etc., also affect the water quantities supplied. In addition, water losses take place both in transmission as well as at storage points.

8.3 Objectives and Scope of the Project

As briefed in the introduction, a complete solution to the irrigation system involves a number of issues such as development of systems and procedures for acquiring and monitoring necessary data and information on electronic media, making the same available to all concerned personnel for planning and monitoring, developing efficient optimization tools for managing the resources, training the personnel in the administration, educating farmers, advising and guiding them about crop patterns and scheduling etc. Since such a task is well beyond the scope of present study, initially this project was taken up with the objective of developing an optimization model for the following problem: Given the water requirements of all AFs over a specified period, the quantities of water available at the main reservoir and other storage points, and the basic network of the irrigation system, build an optimization model to determine optimal utilization of water and the complete time schedule for the optimum distribution of water.

Keeping the above objectives in mind, steps were initiated to collect the basic information from PDS on the following items:

- Reservoirs: locations, capacities (designed and actual), storages.

- Canals (major, minor and sub minor): start and end points, flow rates (designed and actual), officials (posts) in-charge.
- Crop (final) units: survey number, area, types of crops, water requirements, wetting times, crop yields (expected and past data), water source (which regulator/canal).
- Control structures: regulators, sluices.

It has been noticed that the above information was not available in one place and compilation of the same is a very time consuming exercise. Pending this, the work was taken up for modeling the problem and developing a method to solve the problem.

8.4 Formulation and Solution to the Problem

In this section we shall give a detailed description of the irrigation water management problem and formulate the same as an optimization problem. The problem will be described through examples. First let us understand the basic network of the irrigation system. This is shown in Fig. 8.1. For ease of description, the same is reproduced in Fig. 8.2 with regulators and AFs labeled from 2 to 13. The main reservoir is labeled as 1. We shall refer the regulators and the AFs as nodes of the network. Thus, every node is identified with its label. With this labeling, every channel in the network can be identified uniquely by its tail and head nodes. For example, the channel connecting the regulator 2 and regulator 4 is uniquely identified by the pair (2, 4). Here, 2 is the tail node and 4 is the head node of the channel. Similarly, channel connecting regulator 5 and AF 11 is uniquely identified by the pair (5, 11) with 5 and 11 as its tail and head nodes respectively.

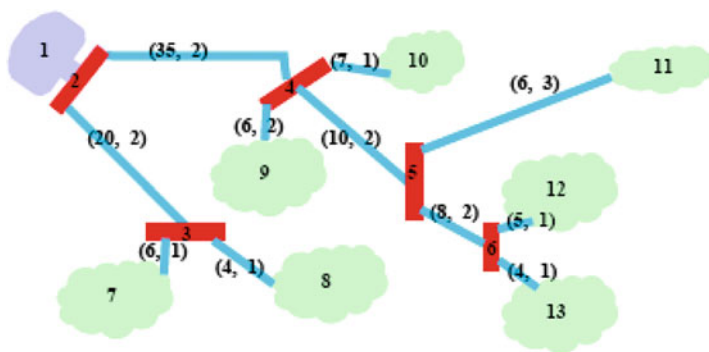


Fig. 8.2 Basic network of irrigation system

In Fig. 8.2, note that every channel is assigned a pair of numbers. For example, channel (2, 4) is assigned the pair (35, 2). Here, the first coordinate, 35, stands for

the maximum delivery capacity of the channel per unit time. The second coordinate, 2, stands for the time taken for delivering the water from the tail node to the head node. If we take day as the unit of time, then (35, 2) means that the channel (2, 4) can carry at most 35 units of water per day and it will take 2 days for the water released at regulator 2 to reach regulator 4.

Generally, regulators have the capacity to store certain amount of water. Therefore, we define $S_c(k)$ as the maximum amount of water that can be stored at regulator k . Also, when water is stored in reservoirs or at regulators, losses take place due to evaporation. In practice, the evaporation losses are affected by atmospheric conditions and surface area of the water exposed. For the purpose of this project, we shall make the simplifying assumption that the evaporation losses are directly proportional to the quantity of water stored and the duration of the storage. We shall use the notation $L_c(k)$ to denote the rate of evaporation loss at node k . Water losses take place during transmission as well. We shall use the notation $L_c(i, j)$ for rate of transmission loss in the channel (i, j) .

One of the water management problems here at a very macro level is to determine how much water to supply to each of the AFs over a given period of time (period being a season or a year). The inputs for such a problem will be the total amount of water available at various storage points during the period and the total requirement of water of the AFs. These two issues together with total channel capacities act as the constraints of the problem. The objectives may be defined by the water losses or meeting certain priorities among the AFs. It should be noted that the time schedule does not come into the picture of this problem. This problem can be formulated and solved as a minimum cost network flow problem. Since time factor is not involved in this problem, these are known as the static network flow problems. However, if the problem involves scheduling activities over a time horizon, one has to deal with dynamic network flow problems. In this project, we are concerned with dynamic network flow problems. A brief description of the static and dynamic network flow problems is given in the next section. See [9] for more details on these problems.

We shall now examine the more important problem of drawing a time-dependent schedule for a given set of requirements and availabilities of the problem. The problem shown in Figs. 8.1 and 8.2 becomes complex for the presentation and discussion. For this reason, we shall use a reduced size problem. The basic network for reduced problem is shown in Fig. 8.3. The number of AFs is reduced to 4 and the nodes are relabeled.

Suppose it is required to plan and develop the water distribution schedule for the month of March in a year. Table 8.1 summarizes the water requirements at the AFs and the water availabilities at regulation points. We see that AF5 requires 8 units of water on March 27th and 9 units on March 28th. Similarly, it can be seen from the table that no water is available prior to March 20th but on 20th, 20 units of water is available at regulator 2. On March 22nd there is an additional 55 units of water available at the main reservoir (node 1). Thus, the total water available during March is only 75 units. The total requirement in the month is equal to 72 units.

Besides the availabilities and requirements, the water distribution schedule has to satisfy the capacity constraints (of storage as well as channel transmission

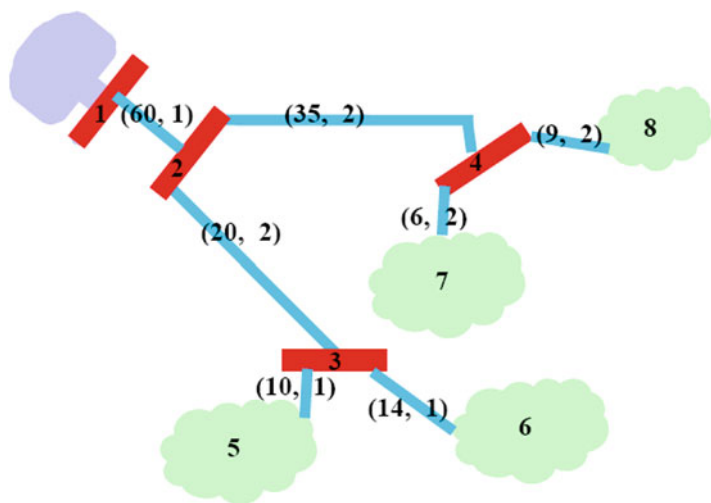


Fig. 8.3 Basic network of the reduced problem

Table 8.1 Water availability and requirement data

Date	Node	Requirement/Availability
March 20	2 (Regulator)	20
March 22	1 (Main reservoir)	55
March 26	AF7	6
March 27	AF5	8
March 27	AF6	14
March 27	AF8	6
March 28	AF5	9
March 28	AF6	14
March 28	AF7	6
March 28	AF8	7

capacities). Table 8.2 presents the data on storage capacities at the regulator points and also the evaporation losses. As explained, we shall assume that the storage losses are proportional to the quantities being stored. Therefore, we shall represent the storage losses in terms of percentages.

Table 8.2 Storage losses in percentages

Node	1	2	3	4
Storage capacity (S_c)	200	20	10	9
Storage loss (L_c)	3	2	1	1

Finally, we need to specify the transmission losses $L_c(i, j)$ of the channels in order to formulate the problem. The transmission losses are given in Table 8.3.

Table 8.3 Transmission losses in percentages

Channel	(1, 2)	(2, 3)	(2, 4)	(3, 5)	(3, 6)	(4, 7)	(4, 8)
Transmission loss	2	1	1	3	1	1	2

The water distribution and supply scheduling problem has to specify the amount of water to be released in each channel on each of the days during the planning horizon. This problem is solved by creating the extended network from the basic network and the input data given in Table 8.1. The extended network is shown in Fig. 8.4. It can be seen that the nodes in the extended network are represented by the node numbers in the basic network coupled with dates. Since there is water available at node 2 on 20th, we create the node represented by 2/20. Since water can reach nodes 3 and 4 on 22nd from 2, the nodes 3/22 and 4/22 are created. Similarly, the other nodes are created as shown in Fig. 8.4. In the water scheduling problem, we should have the option to bring the water to a particular node ahead of time and store it there for later use. As an example, note that nodes 5 and 6 put together require 22 ($= 8 + 14$) units of water on 27th. Since the transit times to 5 and 6 from 3 are one day each, we must have 22 units of water on 26th at node 3. As the channel capacity of (2, 3) is only 20 units, we cannot serve 5 and 6 on 27th unless water is brought to 3 (at least 2 units) before 26th and stored there. Suppose we bring some water, say 5 units, to 3 on 25th itself and store it there for a day. In the formulation of the problem we handle this situation by introducing a fictitious channel between the nodes 3/25 and 3/26 (in the extended network) and fix the flow in this channel as 5 units. Since the storage capacity of node 3 is 10 units, we impose the condition that the transmission capacity of the fictitious channel (3/25, 3/26) as 10. All storage decisions are thus represented by the blue colored arcs (arrows) in Fig. 8.4. The red colored arcs represent the usual flow channels between two regulator points. The flow channels from regulators to AFs are shown by the green color arcs. This way we can understand the extended network shown in Fig. 8.4. Note that two numbers a/b are specified for each of the arcs. Here a stands for water loss and b stands for the storage capacity or the maximum flow capacity of the corresponding arc depending upon whether the arc is a storage arc or a transmission arc. Since storage capacity at regulator 1 is 200 and the storage loss is 3, 3/200 is by the side of the blue arc connecting 1/22 and 1/23, and the arc connecting 1/23 and 1/24. Similarly, since the maximum flow capacity of the channel (2, 3) is 20 and the corresponding transmission loss is 1, we find 1/20 by the side of the arcs which corresponds to (2, 3). Finally, the water availabilities, 55 units and 20 units, are shown at the appropriate nodes in the extended network with a + sign ('+' standing for availability); and the AFs' requirements are shown with minus signs ('-' to indicate that these are requirements).

The dynamic network flow problems can be solved by converting them into static network flow problems. This is exactly what we have done for our problem. The extended network in Fig. 8.4 is actually a dynamic network flow problem presented in the form of a static network flow problem. All the inputs needed for solving the water scheduling problem are shown in the extended network. Once this is ready,

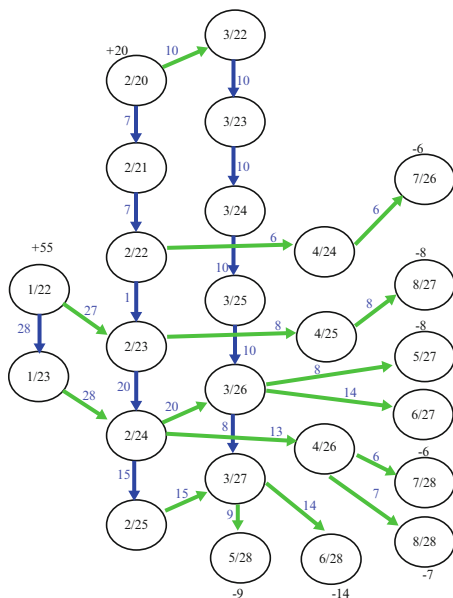


Fig. 8.5 Solution to the example

8.4.2 The Level Constraint

The distribution channels starting from a regulator may not be at the same altitude. Recall the Kanupur canal situation discussed in Sect. 8.2. Thus, different channels starting from a regulator may need different water heads for getting water. This situation can be handled by introducing artificial arcs in the network with certain minimum flow requirements. We shall illustrate this with an example. Let us suppose that channel (4, 6) in Fig. 8.3 is at a higher altitude and requires additional water head at regulator 4. Since additional head means additional quantity of water, whenever there is water requirement at 8, some minimum level should be maintained at 4. Let us assume that this minimum quantity is 5 units. Since AF 8 requires 8 and 7 units of water on 27th and 28th respectively, we should have 13 ($= 8 + 5$) and 12 ($= 7 + 5$) units of water at 4 on 25th and 26th respectively (note that the transit time is 2 days between 4 and 8). We employ the following trick to handle this situation. Introduce a dummy arc in the extended network (in Fig. 8.4) between the nodes 8/27 and 4/26 with a minimum flow requirement of 5 units. Introduce a dummy node 4/27 and declare 5 units of water as a requirement of this node. Connect the nodes 8/28 and 4/27 with another dummy arc and fix its minimum flow as 5 units. This augmented network will take care of the level constraint.

8.4.3 The Group Constraint

In the example, we have specified the AFs' requirements day-wise. Instead we may wish to specify requirements of one or more AFs grouped over days. Suppose we want to specify the requirement of AF 5 as 17 units over the 2 days, 27th and 28th, instead of specifying 8 units on 27th and 9 units on 28th separately. We can handle such constraints by adding a new node with a requirement of 17 units and connect it to the nodes 5/27 and 5/28 using two different arcs. No requirements are specified at 5/27 and 5/28.

8.4.4 Handling Priorities

In practice, often there is not sufficient water for serving all AFs. In such situations, it is not possible to serve water to all AFs and one has to resort to some sort of priorities based on some evaluation. Accordingly, one can attach some cost to each of the arcs ending with AFs and solve the problem.

In this subsection we have seen how the water distribution and the scheduling problem can be formulated as a network flow problem and how various practical constraints can be handled. Also, we have seen how a small size problem with only 8 nodes and 7 arcs got magnified into a problem of 28 nodes and 34 arcs (the extended network). Two factors determine the size of the problem: (i) number of nodes in the basic network and (ii) number of time periods in the planning horizon. It is learnt that the PDS has about 2,000 nodes in the basic network and the planning horizon has about 150 days. Therefore, if we take days as the basic units of time, then we will have about 3,00,000 nodes and about 6,00,000 arcs (decision variables) in the extended network. These problems can be solved quite efficiently with the computing power available today. However, if there is need to reduce the size of the problem, we may think of time units as weeks, and this will drastically reduce the size of the problem. Creating the input data for the extended network based on the basic network, the requirements and the water availabilities is very cumbersome and is very difficult to construct manually.

8.5 Software Development

Network models are one of the most widely used optimization tools. These have a wide range of applications [1, 2, 8, 12], and airline systems are analyzed and optimized using network models. Transportation, transshipment and assignment problems are special cases of network models. Many software and computer systems are analyzed and optimized using network systems. Communication networks is another important application area of network models. Oil and natural gas production and distribution systems use network models. The elegance and efficiency with

which network flow problems are solved have attracted many scientists to formulate a variety of problems as network flow problems. A number of commercial and public domain software packages are available to solve network flow problems. Since commercial software packages are expensive and the public domain software packages are not easily adoptable to our requirements, we thought it would be worthwhile developing our own package for solving network flow problems. As of now our software has two main components: (i) a code for solving a general minimum cost network flow problems and (ii) a code that will act as interface between our network solver and the users in the irrigation department. The latter takes the minimal inputs from the users, formulate the problem, prepare the input file for the solver, read the solution of the solver and present the results in customer specified formats. We wish to add a number of new modules to our software meant for other application areas.

We have adopted one of the most popular and efficient algorithms, namely the primal network simplex algorithm, for solving the minimum cost network flow problems. See [9] for a detailed description of the algorithm. We shall briefly outline some of the important aspects that we have considered in the development of our software. But before we do this let us take a quick look at the minimum cost network flow problem.

8.5.1 Minimum Cost Network Flow Problem

A directed network consists of a set of nodes $N = \{1, 2, \dots, n\}$ and a set $A = \{(i_1, j_1), \dots, (i_m, j_m)\}$ of m directed arcs. Each arc (i, j) is obtained by joining two nodes i and j from N . Here, i and j are called the tail and head nodes of the arc (i, j) . Two subsets S and T of N are specified, S is called the set of source nodes and T is called the set of sink nodes. A certain commodity is to be supplied from nodes of S to nodes of T through the arcs of the network. Each node i in S is assigned an integer a_i which means that i can supply at most a_i units of the commodity. Similarly, each node j in T is assigned an integer d_j which means that j requires d_j units of the commodity. Since the network is a directed network, arc (j, i) , if exists in A , is treated as different from (i, j) . Three numbers, l_{ij} , k_{ij} and c_{ij} , are associated with each arc (i, j) in A . Here l_{ij} and k_{ij} are called the lower and upper bounds of the arc (i, j) and the quantity of the commodity transferred through the arc, called the flow in the arc, should be within the limits l_{ij} and k_{ij} . Next, c_{ij} is the cost of flow per one unit of commodity. Given these inputs, the minimum cost network flow problem is to determine the flows in the network arcs so that the total cost of transportation is a minimum and that the requirements of T are met, availability constraints of S are not violated, and the flow restriction on the arcs are not violated. A mathematical statement of this problem can be stated as follows: Find the flows f_{ij} , $(i, j) \in A$ so as to

$$\begin{aligned}
&\text{Minimize} && \sum_{(i,j) \in A} c_{ij} \cdot f_{ij} \\
&\text{subject to} && f(N, j) - f(j, N) = d_j \text{ for each } j \in T, \\
& && f(N, j) - f(j, N) \leq a_j \text{ for each } j \in S, \\
& && f(N, j) = f(j, N) \text{ for each } j \in N \setminus (S \cup T), \\
& && l_{ij} \leq f_{ij} \leq k_{ij} \text{ for each } (i, j) \in A, \\
&\text{and} && f_{ij} \geq 0 \text{ for all } (i, j) \in A,
\end{aligned}$$

where $f(N, j) = \sum_{i \in N: (i, j) \in A} f_{ij}$ and $f(i, N) = \sum_{j \in N: (i, j) \in A} f_{ij}$. The total availability of the material is equal to $\sum a_i$ and the total demand is equal to $\sum d_i$. If $\sum a_i = \sum d_i$, then we call the network problem as a balanced network problem. We shall now illustrate the network model with an example.

Example. Consider the network with five nodes and seven arcs shown in Fig. 8.6. The arc lower bound, upper bound and the unit costs are shown as triplets (l_{ij}, k_{ij}, c_{ij}) .

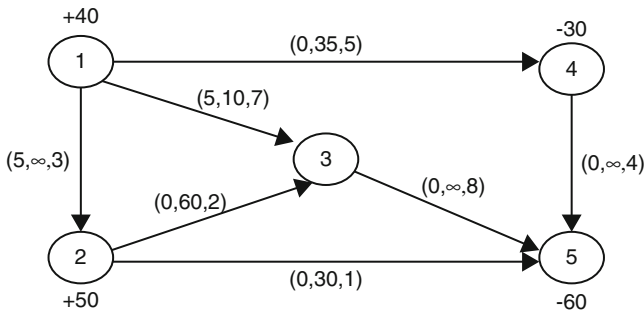


Fig. 8.6 A network problem with 5 nodes and 7 arcs

The first step in solving this problem is to augment the network with two additional nodes 0 and $(s + t + 1)$ arcs, where s and t are the number of nodes in S and T respectively, so as to make it a single source and single sink network problem. The arc between the two newly added nodes will be referred to as the main arc. We then convert the problem into a balanced network problem by defining suitable availability/requirement for the two new nodes if necessary. The augmented problem for the network problem in Fig. 8.6 is shown in Fig. 8.7. Like in the two-phase simplex method, a feasible solution to the network flow problem may be obtained by first considering the Phase-I objective function. We can obtain a feasible solution to a network flow problem, if one exists, using the augmented network in which all the unit costs (including the newly added arcs) are zero except for the main arc. The unit cost of the main arc is taken as 1. Since the augmented problem is feasible, we

stop as soon as we get an optimal solution. If in the optimal solution, the flow in the main arc is zero, then this solution is a feasible solution for the original problem. On the other hand, if the flow in the main arc is positive, then it means that the original problem has no feasible solution. When the original problem has no feasible solution, the requirements of the nodes of T are partially fulfilled by the optimal solution of the Phase-I. In such situations, we switch over to the original costs, takes zero as the cost for all the augmented arcs except the main arc. The cost of the main arc is treated as infinity. In our software, it is ensured that once the main arc becomes a nonbasic arc (that is, the flow in it becomes zero), it never enters the basis again.

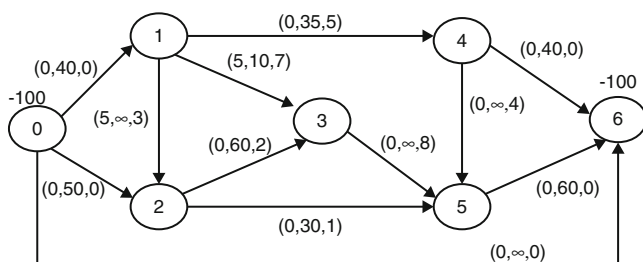


Fig. 8.7 A network problem with 5 nodes and 7 arcs

8.5.2 Resolution of Cycling and Stalling

We use efficient data structures in our program to store and update the data in each iteration of the algorithm. Though degeneracy and stalling problems occur very rarely in real world applications ([9] for details), a good software must take care of these issues. To avoid cycling due to degeneracy, we have implemented dropping arc choice rule developed by Cunningham [4, 5]. If the starting feasible solution corresponds to a strongly feasible partition (see pp. 327 [9]), then Cunningham's method does not allow cycling. Cunningham [4] has also suggested an efficient method of obtaining a solution corresponding to a strongly feasible partition from an arbitrary feasible solution. However, this part is yet to be implemented in our program. This is because we can easily obtain a solution corresponding to strongly feasible partition for the network problem arising out of water scheduling problem discussed in this report. Another rule suggested by Cunningham [4] to avoid stalling is that every arc be examined periodically and select it as entering arc if it is eligible at that time. This has been implemented in our program.

8.6 Summary

Motivated by the acute water shortage problem faced by the farmers in the recent times, we have initiated this work on water distribution planning and scheduling in irrigation projects. Agricultural irrigation uses considerable volumes of water and efficient management of water resources is utmost important. We have used the Penar Delta System of Nellore District in Andhra Pradesh, India, to build our model. The project concerns providing water distribution and scheduling for given requirements of the ayacut and given availabilities of water at various nodes of the irrigation network. The problem is formulated as a dynamic minimum cost network flow problem. We have developed a computer software to aid the users in irrigation department as a decision support system. This software has two main components the first of which is a network solver meant for solving general minimum cost network flow problems. The second component is an interface software between the network solver and the users in the irrigation department. We wish to broaden the scope of application of our software to many other areas of optimization.

Some parts of the western countries have been using the World Wide Web for efficient irrigation management practices. For a number of years, the North Dakota State University Extension Service has been using the Internet to deliver information, beginning with a Gopher server in the early 1990s and more recently with a World Wide Web server. To take full advantage of today's information technology and the computing power, we must quickly establish good database systems for effective planning and implementation.

References

1. Ahuja, R., Magnati, T., Orlin, J.: *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Upper Saddle River (1993)
2. Bazaraa, M., Jarvis, J., Sherali, H.: *Linear Programming and Network Flow*. Wiley, New York (1990)
3. Bouwer, H.: Integrated water management for the 21st century: problems and solutions. *J. Irrig. Drain Eng.* **128**(4), 193–202 (2002)
4. Cunningham, W.H.: A network simplex method. *Math. Prog.* **11**, 105–116 (1976)
5. Cunningham, W.H.: Theoretical properties of the network simplex method. *Math. OR.* **4**, 196–208 (1979)
6. Dudley, N.J., Burt, O.R.: Stochastic reservoir management and system design for irrigation. *Water Resour. Res.* **9**(3), 507–522 (1973)
7. Dudley, N.J., Howell, D.T., Musgrave, W.F.: Optimal intraseasonal irrigation water allocation. *Water Resour. Res.* **7**(4), 770–788 (1971)
8. Evans, J.R., Minieka, E.: *Optimization Algorithms for Networks and Graphs*. Mrcel Dekker, New York (1992)
9. Murty, K.G.: *Network Programming*. Prentice-Hall, Englewood Cliffs (1992)
10. Murthy, G.S.R., Murthy, A.L.N.: A network flow model for irrigation water management. *Algorithmic Oper. Res.* **7**, 83–93 (2013)
11. Paul, S., Panda, S.N., Nagesh Kumar, D.: Optimal irrigation allocation: a multilevel approach. *J. Irrig. Drain. Eng. ASCE* **126**(3), 149–156 (2000)

12. Ravi Ravindran, A.: Operations Research and Management Science Handbook. Taylor & Francis Group, LLC, CRC Press, New York (2008)
13. Richards A.: Coping with Water Scarcity: The Governance Challenge. Center for Global, International and Regional Studies Working Paper. University of California, Santa Cruz. <http://www2.ucsc.edu/globalinterns/wp/wp2001-4.pdf>
14. Vedula, S., Nagesh Kumar, D.: An integrated model for optimal reservoir operation for irrigation of multiple crops. *Water Resour. Res.* **32**(4), 1101–1108 (1996)
15. Wichelns, D.: New policies are needed to encourage improvements in irrigation management. *J. Irrig. Drain. Eng.* **130**(5), 366–372 (2004)
16. Wilson, J., Watkins, J.J.: *Graphs: An Introductory Approach*. Wiley, New York, Chichester (1990)
17. Zilberman, D., Chakravorty, U., Shah, F.: Efficient management of water in agriculture. In: Parker, D.D., Tsur, Y. (eds.) *Decentralization and Coordination of Water Resource Management*, Chapter 22. Kluwer Academic, Boston (1997)

Chapter 9

Statistics for Measuring Tyre Contact Patch

Abstract This project was carried out for a leading tyres manufacturing company in India and deals with the development of an alternative method to measure the land sea ratio, an important performance measure of tyres. The company's method for measuring the ratio depends on the skill of a technical person and uses a costly measuring equipment. It takes almost 3 days to get the measurements for each day's samples. An alternative method is being explored. The new method uses a statistical technique, a multivariate clustering method, and a simple computer program. The result is instantaneous, removes the subjectivity, disposes the need of a technical expert and the need of the costly measuring equipment. The present status of the project is that some fine tuning is required for the proposed method which is under exploration.

9.1 Introduction

It is difficult to imagine our world without tyres today. In our modern world, mobilization rests heavily on tyres. We need tyres everywhere – from bicycles to aeroplanes (Fig. 9.1). Also, the safety of human life depends on quality of tyres. This project was carried for a large tyres manufacturing company in India. Interestingly, the study has evolved from a student, a Statistics post graduate student, who was undergoing his practical training at the tyre industry in question.

The company where this project was carried out has manufacturing set-ups with advanced technologies in multiple locations and produces tyres for variety of vehicles starting from two and three wheelers to small and light commercial vehicles to farm vehicles to trucks and buses. They are into domestic and international markets. They produce both bias and radial tyres and have well over 100 types of tyres. The company has well established testing facilities and a big R&D establishment. The testing department is engaged in extensive testing and analysis activities of their daily produce and the newly designed models.



Fig. 9.1 Tyres for vehicles

Tyres have a number of important quality characteristics such as dimensions, weight, run out, alignment, visual defects and so on. Besides, there are a number of performance characteristics such as balance, camber thrust, circle of forces, dry traction, slip angle, stopping distance, contact patch, rolling resistance and so on.

This project is concerned with the contact patch, commonly referred to as footprint, which in turn is related to rolling resistance and stopping distance. Contact patch is the portion of a vehicle's tyre that is in actual contact with the road surface. Contact patch is measured through a characteristic known as land sea ratio. The problem is about determining the land to sea ratio. According to company's test procedure, it takes about 3 days to get a measurement of the land sea ratio for a tyre sample. It involves taking a carbon copy of the footprint, sending it to the company's corporate office where it is tested by an expert technician using a sophisticated testing instrument. In this project, an alternative method of determining the land sea ratio is developed using a multivariate statistical method of clustering. Using this method it is possible to determine the land sea ratio by just taking a scan of the carbon print and running a simple computer program. This new method has potential advantages. It is instantaneous, eliminates the need of technical expertise and costly testing equipment. The test result can be obtained in the manufacturing site itself, no need to send it to the corporate office. However, the procedure needs some fine tuning which is under exploration. It will be interesting to know the iterations in the evolution of the final test procedure developed. For this reason the entire effort in the development of the proposed method is presented in this chapter.

The organization of this chapter is as follows. The problem is described in Sect. 9.2. Section 9.3 presents the first solution attempted for the problem. A second thought on the solution method led to the use of computer technology clubbed with statistical techniques. Using digital images of footprints, data are read and the second solution uses the principal components [5, 7] of the color spaces data to determine the land sea ratio. For this reason, a brief discussion on the color spaces and the relevant data structures is presented in Sect. 9.4. The solution using the principal components is presented in Sect. 9.5. The final solution using clustering technique is presented in Sect. 9.6. Section 9.7 examines some issues with the proposed method and makes a comparison of some possible solutions. Section 9.8 summarises the chapter.

9.2 Problem Description

The friction between the tyres of a vehicle and the road surface is an important factor that governs the functioning of the vehicle for changing the direction or controlling the speed. The top portion of the tyre that comes in contact with the road or ground is called the tread of the tyre (<http://en.wikipedia.org/wiki/Tread>, cited on Nov 16, 2014). The tread is a thick rubber, or rubber/composite compound formulated to provide an appropriate level of traction that does not wear away too quickly. The tread pattern is characterized by the geometrical shape of the grooves, lugs, voids and sipes. Grooves run circumferentially around the tire, and are needed to channel away water.

Treads are often designed to meet specific product marketing positions. High performance tires have small void ratios to provide more rubber in contact with the road for higher traction, but may be compounded with softer rubber that provides better traction, but wears out quickly. Mud and snow tires are designed with higher void ratios to channel away rain and mud, while providing better gripping performance (for more detail see <http://en.wikipedia.org/wiki/Tire>, cited on Nov 16, 2014).

9.2.1 Contact Patch and the Land Sea Ratio

Lugs are that portion of the tread design that contacts the road surface. Voids are spaces between lugs that allow the lugs to flex and evacuate water. The portion that is in contact with the road at a given instant in time is the **contact patch**. If the total surface of the tread that is in contact with road is treated as *sea*, the lugs may be viewed as *land*. A footprint of the tread of a tyre is obtained by taking a carbon copy of the contact patch by placing the tyre on a flat table and applying a specified load. Figure 9.2 shows one such footprint. The area in the footprint that corresponds to lugs or the land is called the *land area*; and the rest of the area in the footprint, corresponding to the sea, is called the *sea area*. The percentage of land area in the total area (land plus sea area) is termed as land sea ratio.

A tyre with wider grooves and heavy siping (siping is a process of cutting thin slits across a rubber surface to improve traction in wet or icy conditions) etc. will have a low land sea ratio. Higher land sea ratio means higher lug in contact with the road resulting in better dry handling. A high land sea ratio will result in better grip in dry conditions, longer wear life and better noise emissions; while a low land sea ratio has better grip in wet conditions, better rolling resistance and better traction on loose terrains. Typical land sea ratios ranges are: 70 % for general touring tyres, 65 % for high performance tyres, 62 % for ultra high performance tyres and 40 % for winter tyres.

In a view to improve quality standards, the company recently introduced the measurement of land sea ratio of their tyres, and their procedure involved taking footprints of their sample tyres and sending them to their corporate office where



Fig. 9.2 Footprint of the tread of a tyre

the carbon copies are processed by an expert technician using an instrument called computer aided 3-dimensional interactive application (CATIA). The company was facing problems in getting timely results due to postal delays, availability of the technician and the actual time required for processing the carbon copies which was substantial in itself. Consequently, the problem was referred to us to explore the possibility of hastening the process or finding an alternative method of determining the land sea ratio.

9.3 The First Solution Using Simulation

When the problem was addressed, the first approach attempted was to use the standard method of Monte-Carlo simulation technique [4] for estimating areas of odd shapes. For the sake of completeness, it is briefly described below.

Suppose it is required to estimate the area of the irregular shape shown in Fig. 9.3. Fit a rectangle around the box so that the shape is completely sitting in the rectangle. Assume that the dimensions of the rectangle are $a \times b$ as shown in the figure. The coordinates of the rectangular box are as follows: the bottom left point of the box is $(0, 0)$, the bottom right corner of the box is $(a, 0)$, the top left point of the box is $(0, b)$, the top right corner of the box is (a, b) . Let n be a large positive integer. Choose n points randomly in the rectangular box, (x_j, y_j) , $j = 1, 2, \dots, n$. A random point in the box may be simulated as follows. Choose two random numbers r_1 and r_2 between 0 and 1 and then set $x_j = ar_1$ and $y_j = br_2$. Then (x_j, y_j) is a random point in the box. Some of the n points simulated fall on the irregular shape. Count the number of points (x_j, y_j) that are inside or on the irregular shape. Let this number be $k(n)$ (note that this number depends on k , hence the notation $k(n)$). Then, the area of the shape is given by $abk(n)/n$ and the land sea ratio is computed

as $s(n) = k(n)/n$. If n is sufficiently large, then $abk(n)/n$ will be close to the area of the irregular shape, and converges to the area of the irregular shape as n tends to infinity.

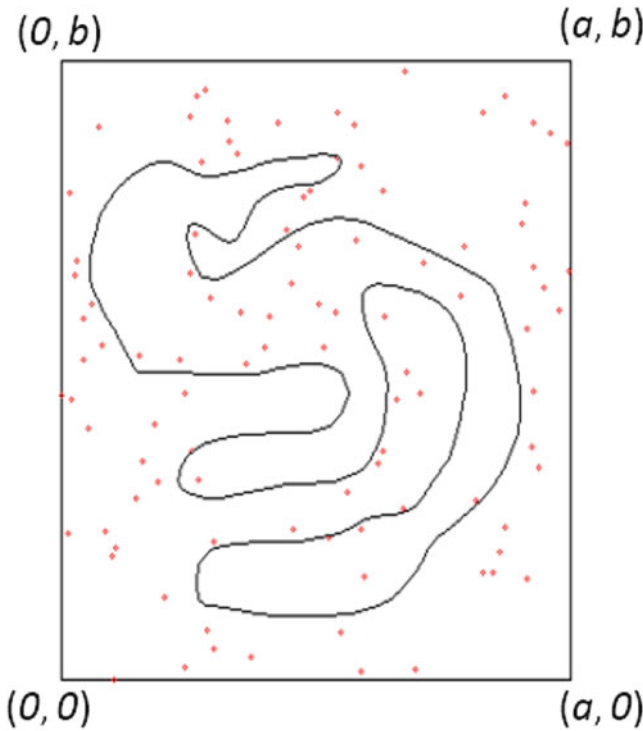


Fig. 9.3 Monte-Carlo simulation for estimating area of irregular shapes

The above method was applied to one sample footprint of a tyre whose CATIA measurement was 69.22 %. A computer program was written that fits a rectangle around the scanned copy of the footprint and shoots random points into the rectangle (see Fig. 9.4). Then, the points that fall on the land area are counted manually, and the land sea ratio is computed as 100 times the number of points landed on the land area divided by the total number of points shot. For this exercise, the total number of points shot was taken as 100. This exercise was repeated five times for the same footprint and the results are: 68 %, 67 %, 64 %, 68 % and 63 %.

Taking the CATIA measurement as the reference, it is seen that the simulation method is not satisfactory. Mean of five repetitions is on the lower side as well as the variation in the five repetitions is high. Moreover, the method is laborious. Even for the small number of 100 points, it is a very time consuming exercise to count the number of points on the land (actually, it is easier to count the number of points on the sea which was done).

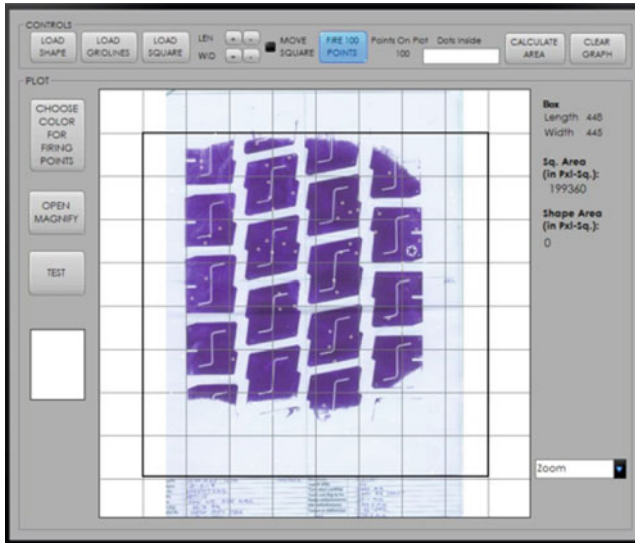


Fig. 9.4 Monte-Carlo simulation for estimating area of irregular shapes

Since this method is not a viable solution practically, we started exploring an alternative method for the purpose. At this juncture it struck to us that when the footprint is scanned on the computer, it actually converts the picture into a digital picture. Our second and the third (final) methods for *estimating* the land sea ratio are based on the digital pictures. Before presenting these methods, a brief overview of the digital pictures and the related data structures is presented in the next section.

9.4 Color Spaces

A color model is an abstract mathematical model describing the way colors can be represented as vectors of nonnegative integers. The most commonly used color model is known as RGB color model. This model portrays every color as a combination of the three primary colors Red, Green and Blue. In this model every color is captured as a 3-dimensional (3-D) integer vector in which each coordinate varies from 0 to 255. Thus, a total of $16777216 (= 256^3)$ different colors can be stored as 3-D vectors. Most of the electronic gadgets such as color televisions, video cameras, image scanners, video games, digital cameras, etc. use the RGB color model. Display devices such as television screens, computer monitor, etc. are made up of pixels in a 2-dimensional grid. Every picture is digitised into pixel grid where each point on the grid is an RGB vector. When this grid is projected on a screen, we see the image of the picture. For the purpose of this project, we can imagine every picture as 5-dimensional vector where the first two coordinates represent the position of the pixel on the grid, and the last three coordinates are the RGB vector in that position.

Therefore, when we scan a footprint, we are actually generating a matrix with five columns in which each row specifies the position of a pixel and the RGB vector in that position. Conceptually, we can think of a color picture as this matrix.

Like the RGB color model, there are other color models such as HSB, YUV, XYZ, and LAB etc. These color models or color spaces are inter-transformable i.e. these color spaces can be derived from one-another. Figure 9.5 presents different color spaces.

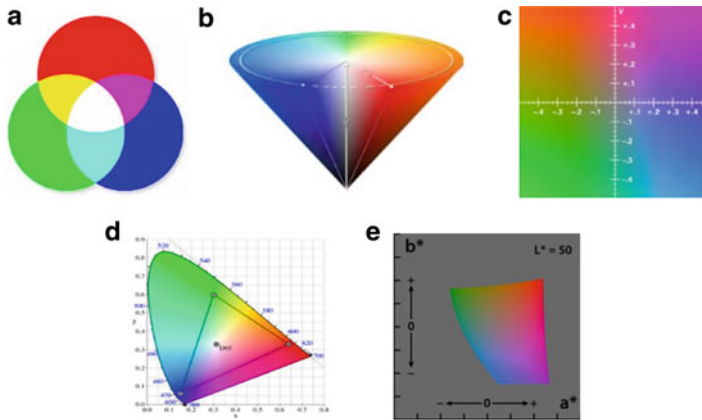


Fig. 9.5 Different colour spaces. (a) RGB. (b) HSB. (c) YUV. (d) CIE XYZ. (e) CIE LAB

The five color spaces RGB, HSB, YUV, XYZ and LAB and the relationships among them are briefly summarized below.

9.4.1 RGB Color Space

RGB [11] is the most basic color space and it represents the colors with 3-D nonnegative integer vectors whose values vary from 0 to 255. This color model produces the color by some combination of RGB coordinates. For example, green will be represented as $[R=0, G=255, B=0]$. It is possible to convert RGB to any other color space [3, 10]. The conversion of RGB into other color spaces is briefly presented in the following subsections.

9.4.2 HSB Color Space

This color model defines the color space in terms of three components [1]:

Hue (H) Hue defines the color type, that is, red, blue etc., it ranges from 0° to 360° , that is, it select colors from a color wheel as shown Fig. 9.5.

Saturation (S) Saturation defines the intensity of the color. It ranges from 0 to 100 %, where 0 represents no color and 100 represents intense color.

Brightness (B) This defines the brightness of the color. It also ranges from 0 to 100 % where 0 is black and 100 is white or other color depending on saturation.

To covert the RGB coordinates to HSB, let $Min = \min\{R, G, B\}$ and let $Max = \max\{R, G, B\}$ and apply the following formulae.

$$H = \begin{cases} \text{undefined} & \text{if } Max = Min \\ \frac{60(G-Min)}{Max-Min} & \text{if } R = Max, \text{ and } G \geq B \\ 360 + \frac{60(G-B)}{Max-Min} & \text{if } R = Max, \text{ and } G < B \\ 360 + \frac{60(B-R)}{Max-Min} & \text{if } G = Max, \\ 240 + \frac{60(R-G)}{Max-Min} & \text{if } B = Max, \end{cases}$$

$$S = \begin{cases} 0 & \text{if } Max = Min \text{ or } B = 0 \\ \frac{Max-Min}{2B} & \text{if } 0 < B \leq \frac{1}{2} \\ \frac{Max-Min}{2B} & \text{if } L > \frac{1}{2} \end{cases}$$

$$B = \frac{1}{2(Max - Min)}$$

9.4.3 YUV Color Space

YUV color model [2] defines the color space in terms of Luma and two chrominance components. It is used in NTSC and PAL video standards and models human perception of colors more closely than standard RGB. This space is defined by the following components: Y, the luma component or the brightness. It ranges from 0 % to 100 %; U and V are the difference component of blue-luminance and red-luminance. These are expressed as factors based on the YUV version. U ranges from -0.436 to 0.436 and V ranges from -0.615 to 0.615 . The conversion of RGB into YUV is given by the linear transformation

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.29900 & 0.58700 & 0.11400 \\ -0.14713 & -0.28886 & 0.43600 \\ 0.61500 & -0.51499 & -0.10001 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

9.4.4 XYZ Color Space

Also known as CIE 1931 XYZ [3], this color space is an absolute color space. It is based on the direct measurements from human eye and serves as the basis from which many other color spaces are defined. It is defined in terms of three components: X – analogues to red and it ranges from 0 to 0.9505; Y – analogues to green and it ranges from 0 to 1.0; and Z – analogues to blue, and it ranges from 0 to 1.089. The conversion of RGB into XYZ is given by the linear transformation

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

9.4.5 LAB Color Space

LAB color space is a color opponent space which is based on non-linearly compressed XYZ color space coordinates [6]. This space is created with an attempt to linearize the perceptibility of color differences. It is defined in terms of three components: L, the luminance; A, the red/green color opponent dimension; B, the yellow/blue color opponent dimension. LAB space cannot be directly obtained from RGB. Standardized RGB coordinates are first converted to XYZ and then from XYZ they are transformed to LAB.

$$\begin{aligned} L &= 116 \left[f \left(\frac{Y}{Y_n} \right) - 16 \right], \\ A &= 500 \left[f \left(\frac{X}{X_n} \right) - f(Y/Y_n) \right], \\ B &= 116 \left[f \left(\frac{Y}{Y_n} \right) - f(Z/Z_n) \right], \end{aligned}$$

where $Y_n = 1$, $X_n = 0.9505$, $Z_n = 1.089$ and

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{if } t > 0.008856, \\ 7.787t + 0.1379 & \text{otherwise.} \end{cases}$$

9.5 The Principal Component Approach

The method of simulation for determining the land sea ratio does not make use of the complete information that is possessed in the footprint. It rather depends on the random points shot in the rectangle and the points on the border case are susceptible

for misclassification. Since the scanned image of the entire footprint is available, it was thought of using the matrix data of the image discussed earlier for determining the land sea ratio.

Logically each pixel of the image should either belong to sea or land. Therefore, if the pixel points of the footprint can be classified into either a sea point or a land point, then the land sea ratio can be taken as 100 times the number of land points divided by the total number of points. If N is the total number pixels in the footprint, then we have N data points which are the RGB values at each of the pixel points. These data were captured for the footprint and analysed using principal component analysis.

The principal component analysis revealed that the first principal component itself accounted for 97.3 % of the total variation. This gave the clue that the first principal component could be used to classify the points. Consequently, the first principal component scores were computed for each of the points in the footprint. When the histogram of the first principal component score was plotted, it projected a clear bimodal distribution which is shown in Fig. 9.6.

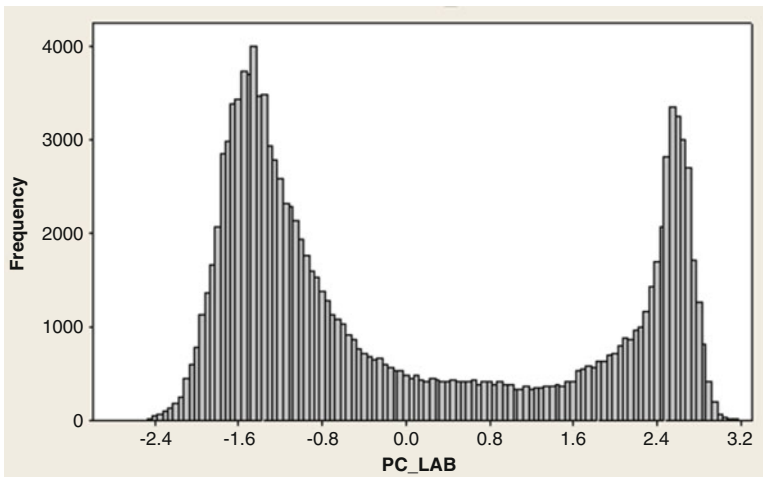


Fig. 9.6 Histogram of the first principal component

Let x denote the first principal component score. To classify the points into land points and sea points using x , it is necessary to find out a cutoff point for x , say x_0 , so that all points with $x < x_0$ are put in one cluster and the rest of the points are put in the other cluster. It is easy to identify which of these two clusters is the cluster of land points (in this case, the points with $x < x_0$ correspond to land points as there are more land points in the footprint).

An objective way of determining x_0 can be as follows. Fit a smooth frequency curve that is closest to the histogram, say $f(x)$, so that the function is a convex function between x_1 and x_2 where x_1 and x_2 are the values of x corresponding to the

two peaks of the histogram with $x_1 < x_2$. Once this is done, choose x_0 as the mid point of all points between x_1 and x_2 which minimize $f(x)$ in the interval $[x_1, x_2]$.

However, to avoid the above mathematical complexity, the cutoff point x_0 is decided by observation and taken as the point corresponding to the the class interval with lowest frequency in the change over region of the frequency curve.

Out of curiosity the above method was tried using all the five color spaces discussed earlier. That is, a sample footprint with known CATIA land sea ratio of 69.78 % was scanned and the data matrices for the footprint were obtained using each of the five color spaces. Using the first principal component of each of the five data sets, the land sea ratios were estimated. Table 9.1 presents the land sea ratios obtained by different methods. It is observed that LAB color space data gave estimate closer to CATIA reading.

Table 9.1 Land sea ratios of a footprint using five different color spaces and the first principal component

Method	CATIA	RGB	HSB	YUV	XYZ	LAB
Land sea ratio (%)	69.78	66.03	64.28	68.07	67.90	70.10

Encouraged by the results, the land sea ratios of four more footprints of different tyres were assessed using the LAB color space data. The results are presented in Table 9.2. Table 9.2 also presents the results of Monte-Carlo simulation alongside in the last column.

Table 9.2 Results of applying first principal component to more samples

Sample	Method		
	CATIA	LAB	Simulation
Footprint 1	69.22	68.91	66.00
Footprint 2	69.78	70.10	70.40
Footprint 3	73.04	72.96	71.60
Footprint 4	71.22	70.21	70.20

Another Interesting Observation

If x_1 , x_2 and x_3 are the original variables, the color space coordinates, then the principal components are linear functions of x_1 , x_2 and x_3 . Let y denote the first principal component. Then, $y = a_1x_1 + a_2x_2 + a_3x_3$ for some constants. The a_j s are obtained from the data. If a_j s are independent of the footprints of different samples, it indicates a functional relationship between the original variables x_j s and the principal components. It is observed that when the principal component analysis is performed on the five footprints using LAB values, the a_j s of the first principal component remained more or less same. The values of a_j s for the five footprints are summarized in Table 9.3. The consequence of this observation is that the linear function $y = 0.5706L - 0.5810A + 0.5800B$ can be used as an identifier to classify a point on the footprint as a land point or a sea point.

Table 9.3 Coefficients of first principal components for different footprints

Sample	L (x_1)	A (x_2)	B (x_3)
Footprint 1	0.570	−0.582	0.581
Footprint 2	0.575	−0.579	0.578
Footprint 3	0.568	−0.580	0.580
Footprint 4	0.571	−0.581	0.581
Footprint 5	0.569	−0.583	0.580

9.6 Approach Using Clustering Technique

The method presented in this section is based on multivariate clustering methods [5, 7–9]. It outdoes the two methods discussed in the earlier sections. Unlike the previous methods, this method uses complete data (in the method using principal components, part of the information possessed by the second and third principal components is discarded).

Given data on a multivariate vector such as color space coordinates (x_1, x_2, x_3), the multivariate clustering methods divide the data points into clusters such that points within a cluster are close to one another and that the inter cluster distances are maximized. There are a number of algorithms to cluster the points in a data set. A popularly known method for clustering is k-means clustering. Without getting into the theoretical detail, we shall simply apply a k-means clustering procedure to our problem of estimating the land sea ratio.

A usual problem that one encounters while clustering points in a data set is to figure out the correct number of clusters. Fortunately, we have a cake walk on this issue because the number of clusters in our case is 2, the cluster of land points and the cluster of sea points. We shall now apply a k-means clustering algorithm to a footprint to estimate the land sea ratio. We shall illustrate the application with the help of one of the sample footprints.

Consider the footprint picture of one of the sample footprints shown in Fig. 9.7. The footprint is scanned and stored in an image file named fp1511.png. When the file is flashed on a monitor, it uses a pixel grid with 367 rows and 311 columns. Therefore, the total number of pixels in this footprint is 114173($= 367 \times 311$). Corresponding to each of these pixels, we have a 3-dimension RGB vector. These vectors are our data points. Some randomly selected data points are shown in Fig. 9.8.

The first step in the new estimation procedure is to read the color space data of the footprint into a matrix of data points. Matlab has a function called “imread” using which the data can be extracted into a matrix, say, M . The size of M for the footprint is 114173×3 . Each row of M corresponds to a pixel position. Let (r_i, c_i) be the position of pixel of i th row of M . When the k-means clustering method is applied to M with $k = 2$ (k is the number of clusters), each row of M is assigned to one of the two clusters. One of these clusters corresponds to land points and the other to the sea points. Let n_L be the number of points in the cluster corresponding to the land points. Then, the estimate of the land sea ratio is taken as $100n_L/N$, where N is the total number of points or pixels in the footprint ($N = 114137$ for

footprint fp1511.png). The Matlab code for this procedure is given in Fig. 9.9. Upon running this, it is found that $n_L = 72777$. Therefore, the estimated land sea ratio using this method is equal to 63.76 %. It should be noted that this estimate is much smaller than the CATIA figure for the footprint which is 71.22 %. Before examining the possible reason for this discrepancy (see Sect. 9.7), we need to understand the impact of clustering method. This is discussed in the next paragraph.



Fig. 9.7 Image of a sample footprint

Pixel position		RGB data vector
Row	Column	
89	127	(94, 54, 151)
311	107	(92, 52, 148)
348	152	(87, 55, 131)
348	152	(103, 68, 148)
163	248	(223, 234, 254)

Fig. 9.8 Some randomly selected data vectors for footprint fp1511.png

Impact of Clustering Method on Land Sea Ratio Estimation

It has been observed that the clustering method assigns each data point to a cluster. There is a nice way to check the efficacy of clustering. This is done as follows. After assigning the data points to clusters using the clustering method find out the mean vector of each cluster. The cluster mean vectors are not necessarily RGB vectors.

For example, the cluster mean vectors for fp1511.png are (88.54, 60.58, 132.08) and (216.95, 221.31, 244.21). Convert these vectors to nearest RGB vectors. In this case, the nearest RGB vectors are (89, 61, 132) and (217, 221, 244). Obtain a new matrix \hat{M} by replacing the rows of M by the respective cluster mean vectors corrected to nearest RGB vector. That is, if i th row of M is assigned to the cluster with mean vector is (88.54, 60.58, 132.08), then replace the i th row of M by (89, 61, 132); other wise replace it by (217, 221, 244). Recall that each row of M , and hence each row of \hat{M} , corresponds to a pixel position in the pixel grid. Now project the picture of \hat{M} by placing the row vectors of \hat{M} in their respective pixel position in the pixel grid. Now compare the two pictures of M (the original picture) and \hat{M} (the *estimated picture*). The Matlab code does this. The two pictures for fp1511.png are shown in Fig. 9.10. From Fig. 9.10, it appears that the clustering method is working well.

```
function [score] = homework1( image_name, K )
    score = 0;
    image = imread(image_name);
    rows = size(image, 1);
    cols = size(image, 2);
    pixels = zeros(rows*cols, 3);
    N1 = rows*cols; %This is the number of pixel vectors
    disp(N1);
    for i=1:rows
        for j=1:cols
            pixels((j-1)*rows+i, 1:3) = image(i,j,:);
        end
    end
    [class1, centroid1] = mykmeans(pixels, K);
    converted_image1 = zeros(rows, cols, 3);
    for i=1:rows
        for j=1:cols
            converted_image1(i, j, 1:3) = centroid1(class1((j-1)*rows+i),:);
        end
    end
    centroid1
    converted_image1 = converted_image1 / 255;
    % Plotting the pictures below
    subplot(1,3,1);
    h = imshow(image_name, 'InitialMag',100, 'Border','tight');
    title('Original')

    subplot(1,3,2);
    h = imshow(converted_image1, 'InitialMag',100, 'Border','tight');
    title('K-means')
end
```

Fig. 9.9 Matlab code for k-means clustering

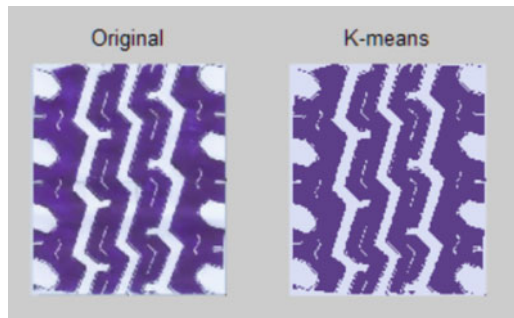


Fig. 9.10 Matlab code for k-means clustering

9.7 Comparison of Results

It was observed that there is a substantial difference between the clustering method estimate and the CATIA evaluation of land sea ratio (63.76 % and 71.22 %). To understand the reasons for this difference, a part of the footprint is cut and using this part the land sea ratio estimate is assessed. That is, a part of Fig. 9.7 is cut and saved as a new file, and then using the new file, the estimate is obtained. Figure 9.11 shows old figure and newly cropped figure from the old figure in one place. The cropped figure is cut in such way that it is one cycle of the design that repeats on the thread. This is understood better by examining Fig. 9.12 in which two copies of the cropped figure are placed one below the other. In this, you can see the design repeating the cropped portion.

The land sea ratio is estimated from the cropped figure and the resulting estimate is 69.27 %. This outcome suggests that selection of the portion of the footprint is crucial in estimating the land sea ratio. Our selection sounds logical because the portion is selected in such a way that one full cycle of the tread pattern is captured in this. Exploring further, the land sea ratio is also estimated from the figure obtained by combining two copies of the cropped figure (this is the figure on the right side of Fig. 9.12). The resulting estimate of the land sea ratio is 69.64 %. It is possible that CATIA's evaluation of land sea ratio itself is biased. This aspect as well as the need for any improvement in the clustering method of estimating land sea ratio are being explored.

9.8 Summary

Land sea ratio is an important tyre parameter that influences the tyre performance. At present it is being assessed using a method that needs the skill of a testing expert and a costly equipment, CATIA. In this project, an attempt is made to find out an alternative method for assessing the land sea ratio. Starting with the usual method

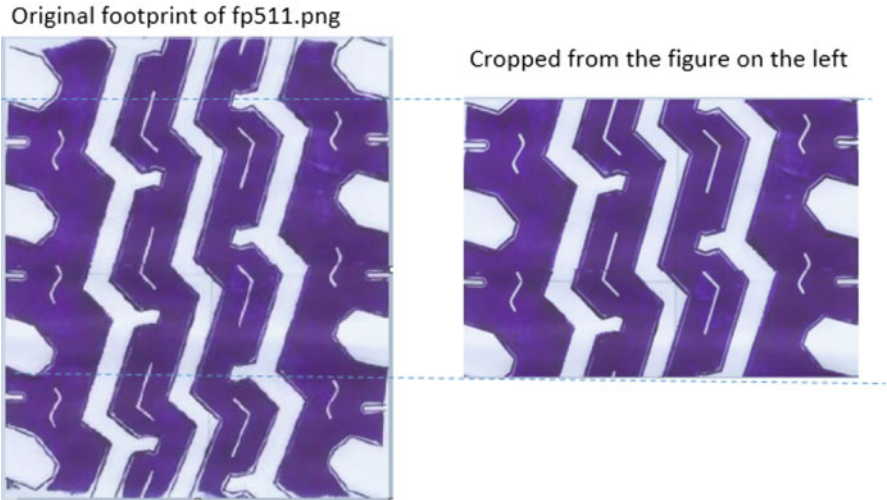


Fig. 9.11 Full and cropped figure of footprint fp511.png. The *horizontal lines* are drawn in such a way that the pattern between the lines on the tread repeats

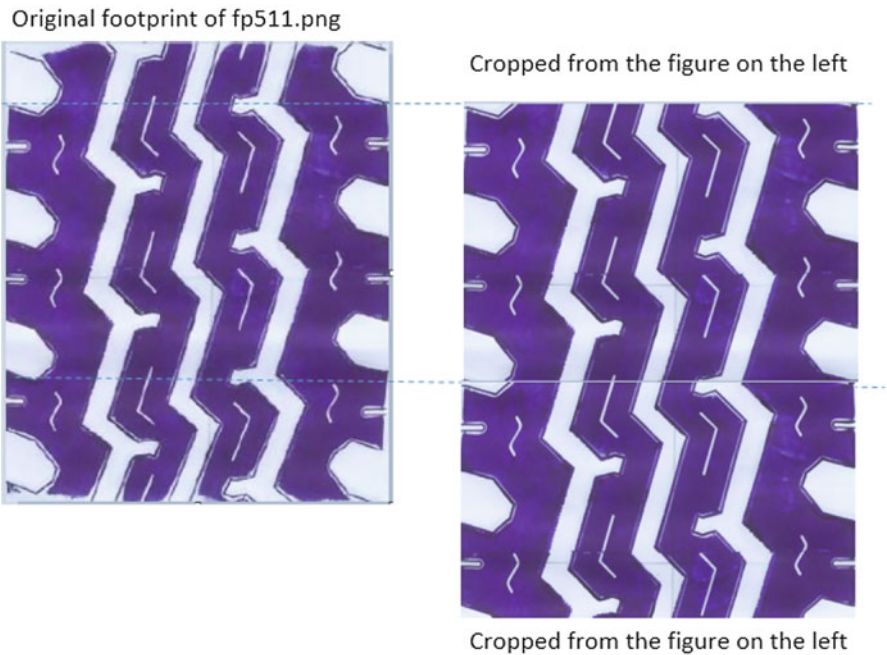


Fig. 9.12 Two copies of the cropped figure placed one below the other showing the repetition of the design

of simulation, the use of principal component analysis and the multivariate clustering method have been explored to assess the land sea ratio. While the results are encouraging, we are yet to find a complete solution for the problem.

Problem: Try clustering with Mahalanobis distance.

Acknowledgment This is to acknowledge that this work was initiated by my student Mr. Ajay Sakarwal who took active role in pursuing this case study. I wish him success in all his endeavours.

References

1. Agoston, M.K.: Computer Graphics and Geometric Modeling: Implementation and Algorithms, pp. 300–306. Springer, London (2005)
2. Anderson, D.: Color Spaces in Frame Grabbers: RGB vs. YUV. Retrieved 2008-04-08. http://amik-bsi-tasikmalaya.ptkpt.net/_lain.php?_lain=6960&_en=ENGLISH#cite_ref-2
3. Broadbent, A.D.: A critical review of the development of the CIE1931 RGB color-matching functions. *Color Res. Appl.* **29**(4), 267–272 (2004)
4. Fishman, G.S.: Monte Carlo: Concepts, Algorithms, and Applications. Springer, New York (1995)
5. Flury, B.: Common Principle Components Analysis and Related Multivariate Models. Wiley, New York (1998)
6. Hoffmann, G.: CIELAB Color Space. <http://docs-hoffmann.de/cielab03022003.pdf>
7. Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis, 6th edn. Prentice Hall, Upper Saddle River (2007)
8. Morrison, D.F.: Multivariate Statistical Methods. McGraw Hill, New-York (1990)
9. Muirhead, R.J.: Aspects of Multivariate Statistics. Wiley, New York (1982)
10. Poynton, C.A.: Digital Video and HDTV: Algorithms and Interfaces. Morgan Kaufmann, Amsterdam/Boston (2003)
11. Tkalcic, M., Tasic, J.: Color spaces—perceptual, historical and applicational background. In: 10th IS&T/SID Color Imaging Conference, Scottsdale, vol. 1, 304–308 (2003)

Chapter 10

A Case Study that Lead to New Designs

Abstract This chapter deals with construction of a new design for an experiment that was required for one of the clients, an aluminium alloy foundry in Chennai, India. The experiment was aimed at reducing casting defects. Typically these experiments are carried out using suitable orthogonal arrays as they are most efficient for such experiments. But the experiment required for the study could not be fitted into any of the known orthogonal arrays. This has triggered a quest for finding the most efficient design for the requirement. This quest has lead to construction of a factorial design that is pretty close to an orthogonal array. The ideas for constructing the new design are interesting and the approach could be extended to construction of similar designs for similar requirements. The approach uses integer linear programming formulations for the construction. The main purpose of including this study as a chapter in this book is to highlight the fact that research is an important faculty of a successful consultant.

10.1 Introduction

Design of Experiments is one of the most widely applied subjects for product and process optimization in industry. Most of the experiments fall under the category of factorial experiments, and orthogonal arrays (OAs) are the most frequently used designs [7, 13, 16–18]. This chapter deals with construction of a new design for an experiment that was required for one of the clients, an aluminium alloy foundry in Chennai, India. The experiment was aimed at reducing casting defects. Typically these experiments are carried out using suitable orthogonal arrays as they are most efficient for such experiments. But the experiment required for the study could not be fitted into any of the known orthogonal arrays. This has triggered a quest for finding the most efficient design for the requirement. This quest has lead to an innovative construction of a factorial design that is pretty close to an orthogonal array. The ideas for constructing the new design are interesting and the approach could be

extended to construction of similar designs for similar requirements. The approach uses integer linear programming formulations for the construction. The main purpose of including this study as a chapter in this book is to highlight the fact that research is an important faculty of a successful consultant. A preliminary version of this chapter may be found in [12].

The reader is expected to have the basic knowledge of factorial experiments to understand the contents of this chapter. The reader may refer to [4, 8] and [11] for topics on factorial experiments. However, we shall make the presentation in such a way that this requirement is kept to a minimum. The organization of the chapter is as follows. Section 10.2 introduces the company's problem. Basic concepts of factorial effects, the main effects and interactions, are briefly introduced in this section with the help of the factorial experiment designed for the foundry problem. Section 10.3 presents the details of the search for a design for the problem from the literature and the criteria for optimal designs. In Sect. 10.4, we present the ad hoc method of constructing the design for the company's problem and show how it can be used for constructing design for similar requirements. Section 10.5 presents how the method of construction can be formulated as an integer linear programming problem. The chapter is concluded in Sect. 10.6 with a summary.

10.2 Background

In this section we shall introduce the problem encountered in the aluminium alloy foundry. With the help of the design constructed for this problem, the basic concept of factorial experiments, the main effects and the interactions of the factors are introduced. Further, the original model of the factorial experiments and its reparameterized are discussed.

10.2.1 Problem

In order to optimize the process parameters in the foundry, it was planned to conduct an experiment with the following factors and levels: (i) Bath Temperature (A) at three levels, (ii) Phosphorous Content (B) at three levels, (iii) Charge Ratio (C) at two levels, and (iv) Filtering Method (D) at two levels. For simplicity, we shall denote the levels of factors A and B by 1, 2 and 3 and those of C and D by 1 and 2. A treatment combination (TC) is a combination of the levels of the four factors. Thus, when all these four factors are kept at level 1, we get the TC (1,1,1,1). This TC is shown in Table 10.1 against TC number 1. Supposing A and C are kept at level 1, and B and D are kept at level 2, we get the TC (1,2,1,2). This is shown against TC number 2 in Table 10.1. Similarly, TC (1,3,1,1) is shown against TC number 3 in Table 10.1. Note that there are $36 (= 3^2 \times 2^2)$ TCs in all and 18 of them are shown in the second to fifth columns in Table 10.1. Each time an experiment is

conducted with a TC, a response is observed. For example, it could be the proportion of defective castings produced with a given TC. Designing an experiment means specifying which TC will be experimented with. For example, Table 10.1 suggests that conduct 18 experiments with the 18 TCs listed there. The question in designing the experiment is all about which of the 36 TCs to be experimented with. The final goal of design of experiments is to identify the best treatment combination.

Table 10.1 Design in 18 treatment combinations

TC i	A	B	C	D	Response y_i
1	1	1	1	1	y_1
2	1	2	1	2	y_2
3	1	3	1	1	y_3
4	2	1	1	1	y_4
5	2	2	1	2	y_5
6	2	3	1	1	y_6
7	3	1	1	1	y_7
8	3	2	1	2	y_8
9	3	3	1	2	y_9
10	1	1	2	2	y_{10}
11	1	2	2	1	y_{11}
12	1	3	2	2	y_{12}
13	2	1	2	2	y_{13}
14	2	2	2	1	y_{14}
15	2	3	2	2	y_{15}
16	3	1	2	2	y_{16}
17	3	2	2	1	y_{17}
18	3	3	2	1	y_{18}

10.2.2 Main Effects and Interactions

For the purpose of this chapter, we shall briefly describe main effects and interactions. Suppose we have decided to conduct the experiment with the 18 TCs listed in Table 10.1. Let y_i be the response corresponding to the TC i , $i = 1, 2, \dots, 18$. A **contrast** of the responses is defined as a linear combination of y_i s, say $l'y = \sum l_i y_i$, where $l = (l_1, l_2, \dots, l_{18})^t$ and $y = (y_1, y_2, \dots, y_{18})^t$. We say that $l'y$ is contrast if l is nonzero vector with its sum of coordinates is equal to zero. If $l'y$ is a contrast, then we say that l is the corresponding contrast vector.

For any two level factor, we have a contrast called the main effect of that factor and that is defined by the corresponding contrast vector l . For C , the contrast vector is given by l_C whose first 9 coordinates are 1s and the remaining ones are -1 s. The contrast vector for main effect D is given by $l_D = (1, -1, 1, 1, -1, 1, 1, -1, -1, -1, 1, -1, -1, 1, -1, -1, 1, 1)^t$. Note that l_i , the

coefficient of y_i in the contrast, is 1 if the level of the factor corresponding to y_i is 1, and it is -1 otherwise. This is how a contrast vector for a 2-level factor is defined.

For a 3-level factor, there are two contrast vectors, called the linear and quadratic contrasts. The linear contrast vector l of a 3-level factor is defined by $l_i = -1$ if y_i corresponds to level 1 of the factor, $l_i = 0$ if y_i corresponds to level 2 of the factor, and $l_i = 1$ if y_i corresponds to level 3 of the factor. The quadratic contrast vector q of a 3-level factor is defined by $q_i = 1$ if y_i corresponds to level 1 or level 3 of the factor and $l_i = -2$ if y_i corresponds to level 2 of the factor. Thus, the linear contrast vector of A is given by $l_A = (-1, -1, -1, 0, 0, 0, 1, 1, 1, -1, -1, -1, 0, 0, 0, 1, 1, 1)^t$ and the quadratic contrast of A is given by $q_A = (1, 1, 1, -2, -2, -2, 1, 1, 1, 1, 1, 1, -2, -2, -2, 1, 1, 1)^t$.

Let p and q be two contrast vectors. Define their product contrast vector as r where the $r_i = p_i q_i$ for all i . The two contrast vectors are said to be orthogonal (to each other) if $\sum p_i q_i = 0$. Orthogonality of p and q means that the *estimates of the corresponding effects* are uncorrelated. Having uncorrelated estimates is a desirable feature of the design. The quantity $\sum p_i q_i$ is called the correlation between the two corresponding effects (see Problem 10.2).

If p and q are two contrast vectors corresponding to two distinct factors, then their product contrast vector is called their interaction contrast vector. If p is the interaction contrast vector of factors C and D , then the contrast $p'y$ is called the interaction between C and D .

Between any two 2-level factors, there is only one interaction. Consider the two factors A and C – a 3-level factor and a 2-level factor. There are two interaction contrast vectors for these two factors – one of them defined by the linear contrast vector of A with contrast vector of D and the other with the quadratic contrast vector of A with the contrast vector of D . Next, between two 3-level factors, say A and B , there will be four interaction contrast vectors – between linear (of A) and linear (of B), between linear (of A) and quadratic (of B), between quadratic (of A) and linear (of B) and between quadratic (of A) and quadratic (of B).

For ease of notation, we shall denote the main effects by A_L, A_Q, B_L, B_Q, C and D , and the interactions by $A_L B_L, A_L B_Q, A_Q B_L, A_Q B_Q, A_L C$, and $A_Q C$.

10.2.3 Model of the Factorial Experiments

The full factorial experiment for the above problem involves 36 different TCs. Any k of these 36 TCs will be referred to as a fraction of size k (k is called the run length). Say that a fraction of size k is regular if 36 is divisible by k . Say that a column of a fraction is *homogeneous* if all the levels in that column appear with same frequency; and say that a fraction is homogeneous if all the columns in the fraction are homogeneous. Throughout this chapter, we confine our attention to regular homogeneous fractions only.

Let y_i denote the response due to i th TC. It is assumed that y_i 's, $1 \leq i \leq k$, are uncorrelated, each with variance σ^2 . If $A_3B_2C_1D_2$ is the i th TC, then under the model assumptions, the expected value of y_i is given by

$$E(y_i) = \mu + a_3 + b_2 + c_1 + d_2 + (ab)_{32} + (ac)_{31} \quad (10.1)$$

Here μ is the general effect and $a, b, c, d, (ab)$ and (ac) denote respective factorial effects at their corresponding levels. Using reparameterizations of these factorial effects, the above model can be remodeled as

$$E(y) = X\beta \text{ and } D(y) = \sigma^2 I. \quad (10.2)$$

Here $y = (y_1, y_2, \dots, y_k)^t$ is the response vector and β is the vector whose first coordinate correspond to μ , and the remaining coordinates correspond to the linear and quadratic effects (main effects and interactions) of the factors that were discussed in previous subsection. The i th row of the *design matrix* X , X_i , corresponds to the i th TC in the experiment and the j th column of X , X_j , corresponds to β_j .

For the foundry's problem in question, β consists of 13 elements, namely – general effect (μ), linear/quadratic components of main effects (A_L, A_Q, B_L, B_Q, C, D), and the linear/quadratic components of interactions ($A_LB_L, A_LB_Q, A_QB_L, A_QB_Q, A_LC, A_QC$). Further, X is a $k \times 13$ matrix consisting of corresponding contrast vectors along with a vector of 1's for μ .

In order that all the main effects and interactions (including μ) be estimable, a necessary condition is that X is of full column rank (see [15]). The least square estimator of β is given by $\hat{\beta} = (X^tX)^{-1}X^ty$ and the dispersion matrix of $\hat{\beta}$, $D(\hat{\beta})$, is $\sigma^2(X^tX)^{-1}$. For simplicity, we shall assume, without loss of generality, that $\sigma^2 = 1$ throughout this chapter.

10.3 Search for Optimal Design

Given the requirements in terms of main effects and interactions to be estimated, one has to first design the experiments, i.e., identify treatment combinations which can lead to estimation of parameters of interest. Since there are several choices of designs, one is interested in using the most efficient design. Usually the efficiency of a design is measured in terms of run length, i.e., the number of TCs, and dispersion matrix of the estimators of parameters of interest.

A factorial experiment is said to be asymmetrical if there are at least two factors for which the number of levels considered for each of these factors is not the same. Designing asymmetrical fractional factorial experiments is relatively more difficult compared to the other category. The problem becomes more complex when the model involves interactions. Many researchers suggested methods for orthogonal plans (see [1, 5, 6, 19, 20]). Orthogonal designs are efficient, but for a fixed run length such designs may not exist. In such cases one has to sacrifice orthogonality in favour of smaller run length. Anderson and Thomas [2] proposed to derive

resolution IV designs by collapsing the levels in foldover designs. Webb [21] developed a number of catalogues for small incomplete experiments where each factor is tried at either two levels or three levels. Box and Draper [3] studied the optimality of designs using $|X^tX|$ criterion (see Sect. 10.3.1). Mitchell [9] proposed DETMAX algorithm for the construction of efficient designs with respect to $|X^tX|$ criterion. Wang and Wu [20] introduced the concept of *near-orthogonality* and produced some construction methods for the main-effect plans. They also enlist a number design layouts with varying run lengths for 2^m3^n factorial experiments.

The foundry problem requires an efficient design to estimate the main effects A, B, C, D and the interactions AB and AC . Failing to get a satisfactory answer to this problem from the literature, we hit upon an ad hoc method for constructing an efficient design for this purpose. It is transparent from the method of construction, that certain main effects/interactions can be estimated orthogonally. We, then, developed a general methodology to convert our ad hoc method into a systematic one.

An interesting aspect of the methodology is that we formulate the problem as an optimization problem. It is shown that constructing efficient designs, under some restrictions, can be formulated as an integer linear programming problem (ILP) with either linear or quadratic objective function.

The scope of this chapter is confined purely to 2^m3^n factorial experiments where it is required to estimate all the main-effects and some of the two-factor interactions.

10.3.1 Criteria for Optimal Design

For the given requirements of estimating the factorial effects, there exist several choices of designs and one is interested in choosing an optimal design. A number of criteria (for optimality) have been developed (see [14]) to compare and construct designs. We shall mention two such criteria here.

- Under the assumption that columns of X are normalized, $|X^tX|$ is a overall measure of efficiency (denoted as D -efficiency) of the design (see [15, 20]). Since the $|X^tX|$ is always less than or equal to product of the diagonal entries of X^tX , $|X^tX|$ attains its maximum when the columns of X are orthogonal. Furthermore, under orthogonality, the estimates of β_i s attain the minimum variance. Therefore the D -efficiency of design is given by

$$D - \text{efficiency} = 100 \times |X^tX|^{\frac{1}{p}} / k, \quad (10.3)$$

where p is the number of parameter to be estimated and k is the run length (see [10]).

- Another measure of efficiency is given by

$$I_F = p / (k \sum_{i=1}^p w_i V(\hat{\beta}_i)) \quad (10.4)$$

where k and p are as defined above ($p = 13$ in our problem), w_i 's are some associated weights and $V(\hat{\beta}_i)$ is the variance of $\hat{\beta}_i$ which is the i th diagonal entry of $(X^T X)^{-1}$ (see [21]).

For any 3-level factor F , we shall use the notation l_{F_L} and l_{F_Q} to denote the contrast vectors of the linear and quadratic components of the main effect of F . If G is another 3-level factor, then $F_L G_L$, $F_L G_Q$, $F_Q G_L$, $F_Q G_Q$ are the linear and quadratic components of FG interaction.

For any two contrast vectors u and v , we define the *nonorthogonality* between u and v as $|u^T v|$. If F and G are any main effects or interactions, we define the nonorthogonality between F and G to be nonorthogonality between u and v where u and v are any contrast vectors of F and G respectively.

10.4 Construction Method

In this section, we will describe our method of constructing designs. It should be mentioned that this method of construction produces only regular fractions which are also homogeneous. It is clear that in order to estimate the main effects and interactions specified for the foundry problem, we need at least 13 TCs. Any regular fraction of this $2^2 3^2$ experiment must contain $2^i 3^j$ TCs, $0 \leq i, j \leq 2$. Since the minimum number is 13, we go for the smallest regular fraction with a run length not less than 13.

10.4.1 Construction in 18 Runs

To construct a design in 18 runs, we first look at $2^1 3^2$ full factorial. This layout has 3 columns (see columns (2), (3) and (4) of Table 10.2). Since it is a full factorial, all the main effects and interactions (of A, B, C) are estimable and are orthogonal. We now augment the 3 columns with another 2-level column (column under D) so that we have a layout for the required design. The new column is chosen in such a way that after rearranging the rows and columns, columns under D, B, A form a layout which a full factorial experiment (see second part of Table 10.2). The resulting layout with augmented column yields the required design. Therefore, the main effects of A, B and their interaction are orthogonal to effect of D . Note that one of the choices for column under D is the column under C itself. However, we wish to choose, if possible, the column under D in such a way that l_D is orthogonal to l_C , $l_{A|C}$ and $l_{A|C}$. But it is impossible to get a 2-level homogeneous column for D so that $l_C^T l_D = 0$ (because the run length is 18). Therefore, we try to obtain the column under D so as to minimize nonorthogonality between the main effects of C and D (i.e., minimize $|l_C^T l_D|$). While doing so, we should also try to keep the nonorthogonalities of D with AC interaction as low as possible. Clearly we have an optimization problem at hand. We defer the details of this methodology to Sect. 10.5.

For the foundry problem, we first tried to construct a suitable design using the existing literature on the subject. Consequently, following [2], we constructed a resolution IV design (so as to estimate the interactions). This required 19 treatment combinations. The resulting layout and the corresponding dispersion matrix are given in Tables 10.4 and 10.5 respectively. The D -efficiency and I_F -efficiency of this design are 103.89 % and 80.88 % respectively. Moreover, this design is irregular and nonhomogeneous.

Table 10.4 Resolution IV design in 19 runs

TC. No.	C	D	A	B	TC. No.	C	D	A	B
1	1	1	1	1	11	2	1	2	2
2	1	1	1	2	12	2	2	1	2
3	1	1	1	3	13	2	2	1	3
4	1	1	2	1	14	2	2	2	1
5	1	1	3	1	15	2	2	2	2
6	1	2	1	1	16	2	2	2	3
7	1	2	2	2	17	2	2	3	1
8	1	2	3	3	18	2	2	3	2
9	2	1	1	1	19	2	2	3	3
10	2	1	3	3					

Table 10.5 Dispersion matrix of the design of (in 19 runs) $\times 10^2$

Effect	μ	C	D	A_LC	A_QC	A_L	A_Q	B_L	B_Q	A_LB_L	A_LB_Q	A_QB_L	A_QB_Q
μ	7.67												
C	1.51	6.76											
D	1.51	-1.58	6.76										
A_LC	-2.67	-0.46	-0.46	10.32									
A_QC	-0.55	-0.46	-0.46	-0.39	3.18								
A_L	-1.88	-1.39	-1.39	1.19	-0.40	9.74							
A_Q	-0.48	-0.60	-0.60	-1.31	1.34	-0.50	4.02						
B_L	-1.89	-1.54	-1.54	0.17	0.96	-0.24	1.56	9.87					
B_Q	-0.82	-0.45	-0.45	1.50	-0.61	1.42	-1.26	-0.35	3.64				
A_LB_L	-0.48	0.08	0.08	-1.72	0.07	-0.20	-0.25	-0.03	-0.71	10.70			
A_LB_Q	1.47	0.60	0.60	-0.97	0.02	-2.27	0.48	-0.79	-1.17	0.16	5.54		
A_QB_L	1.48	0.75	0.75	0.05	-1.34	-0.62	-1.58	-2.57	0.60	-0.01	-0.11	6.01	
A_QB_Q	-1.30	-0.39	-0.39	0.75	-0.10	0.76	0.25	0.66	0.50	-0.36	-0.62	-0.51	1.58

10.4.2 Construction in 12 Runs

When there are restrictions on the run length, one often ignores certain higher order interactions so that the run length is reduced. There could be other reasons to ignore higher order interactions such as difficulty in giving practical interpretation to such interactions etc. For example, in [21] it is argued in favour of considering only linear components of interaction involving 3-level factors. Supposing we can sacrifice a highest order interaction in the foundry problem, we can ask the question: Is there a design in 12 runs?

Table 10.6 Construction of a fraction with 12 runs

2 ² .3 full factorial					Additional column					After rearranging rows and columns				
TC. No.	C	D	A	B	TC. No.	B	D	C	A	TC. No.	B	D	C	A
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(6)	(7)	(8)	(9)	(10)
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	2	3	8	1	1	2	2	8	1	1	2	2
3	1	1	3	2	5	1	2	1	2	5	1	2	1	2
4	1	2	1	2	12	1	2	2	3	12	1	2	2	3
5	1	2	2	1	3	2	1	1	3	3	2	1	1	3
6	1	2	3	3	7	2	1	2	1	7	2	1	2	1
7	2	1	1	2	4	2	2	1	1	4	2	2	1	1
8	2	1	2	1	11	2	2	2	2	11	2	2	2	2
9	2	1	3	3	2	3	1	1	2	2	3	1	1	2
10	2	2	1	3	9	3	1	2	3	9	3	1	2	3
11	2	2	2	2	6	3	2	1	3	6	3	2	1	3
12	2	2	3	1	10	3	2	2	1	10	3	2	2	1

Earlier it was observed that a minimum of 13 runs is necessary to estimate all the components of main effects and interactions. We assume that the highest order component of the interaction between A and B , viz., $A_Q B_Q$, is absent. We proceed, as before, to construct the design by first writing down the $2^2 3^1 (= 12)$ full factorial layout (see columns (2)–(4) of Table 10.6).

We then augment these 3 columns with another 3 – level column (under B) so that after rearranging the rows, the columns under B , D and C (see second part of Table 10.6) form a $2^2 3^1$ full factorial layout. Here we obtain the column under B by minimizing the nonorthogonality between A and B while maintaining the orthogonality of B with C , D and CD . This, again, is an OR problem and the methodology is described in Sect. 10.5. The dispersion matrix of this design is given in Table 10.7. The D -efficiency and I_F -efficiency of the design are 84.92 % and 54.55 % respectively.

Table 10.7 Dispersion matrix for the design with 12 runs

Effect	μ	A_L	A_Q	B_L	B_Q	$A_L B_L$	$A_L B_Q$	$A_Q B_L$	C	$A_L C$	$A_Q C$	D
μ	0.093											
A_L	0.000	0.139										
A_Q	0.000	0.000	0.046									
B_L	0.000	−0.014	0.014	0.222								
B_Q	0.000	0.014	0.005	0.000	0.074							
$A_L B_L$	0.000	−0.042	0.014	−0.167	0.056	0.667						
$A_L B_Q$	0.019	−0.014	−0.005	0.000	0.037	0.111	0.111					
$A_Q B_L$	0.019	−0.000	0.009	−0.056	−0.019	0.111	0.000	0.111				
C	0.009	−0.014	0.005	−0.056	0.019	0.222	0.056	0.056	0.167			
$A_L C$	0.000	0.000	0.000	0.125	0.042	−0.125	0.042	−0.083	−0.042	0.250		
$A_Q C$	0.000	0.000	0.000	0.042	−0.042	−0.125	−0.042	0.000	−0.042	0.000	0.083	
D	0.009	−0.014	0.005	−0.056	0.019	0.222	0.056	0.056	0.083	−0.042	−0.042	0.167

10.4.3 Construction of Design for $2^3 3^3$ with Interactions in 24 Runs

Consider an experiment with three 3-level factors, A , B , C and three 2-level factors D , E and F . Suppose it is required to estimate the interactions, AB , BC , AD , DE and EF . The minimum run length is 22. We construct a design in 24 runs. As before, we start with a $2^3 3^1$ full factorial layout with the factors A , D , E and F . Next, we augment this layout with two more 3-level columns for B and C in two steps. First we augment the layout with the column for B so that the nonorthogonalities of B with A , AB and AD are minimized and that the columns of B , D , E and F form a full factorial layout. From this, we get the layout for the factors A , B , D , E and F in 24 runs.

Finally, this layout is augmented with another 3-level column for C so that the nonorthogonalities of C with A , B , AB , BC and AD are minimized and that the columns of C , D , E and F form a full factorial layout.

10.5 OR Formulations

In Sect. 10.4.1 we encountered the problem of augmenting a given set of columns with 2- or 3-level column so as to form a design layout. In this section we formulate this problem as an integer programming problem (with linear or quadratic objective function). We first describe the procedure and the formulation, and then illustrate the same with some examples.

10.5.1 Augmenting with 2-Level Columns

Consider the problem of constructing the fourth column discussed in Sect. 10.4.1. It is required to construct a 2-level column which along with columns of A and B forms a full factorial design. In addition to this, the resulting layout (with the columns) should provide estimability of all the main effects and interactions of interest.

The problem of choosing a 2-level column is equivalent to finding a 18×1 contrast vector x so that it is orthogonal to l_{AL} , l_{AQ} , l_{BL} , l_{BQ} , l_{ALBL} , l_{ALBQ} , l_{AQBQ} , and its nonorthogonality with l_C , l_{CAL} , l_{CAQ} is as close to zero as possible. In fact, if the nonorthogonalities ($|x^t l_C|$, $|x^t l_{ALC}|$, $|x^t l_{AQC}|$) are equal to zero, then the resulting design is an orthogonal design with which we can estimate all the main effects and interactions orthogonally. Therefore, we expect that minimizing the nonorthogonality, in some sense, should lead us to estimability of the required main effects and interactions.

Thus the problem of constructing the fourth column (which in turn gives us a design) is an optimization problem. Since we are interested in minimizing

$|x^t l_C|$, $|x^t l_{A_L C}|$ and $|x^t l_{A_Q C}|$ we may tackle the problem in different ways. We shall consider two types of objective functions, $f(x)$, here:

- (i) Take $f(x) = \lambda_1 |x^t l_C| + \lambda_2 |x^t l_{A_L C}| + \lambda_3 |x^t l_{A_Q C}|$, where λ_i s are some positive numbers (weights). Setting any of the λ_i 's to be a very large positive number is equivalent to treating the corresponding component as a constraint. For example, if we set λ_1 to be a large positive number, then we are looking for a solution which satisfies $x^t l_C = 0$. This, in other words, means we are looking for a design which can provide orthogonal estimates for all the main effects while minimizing the nonorthogonality between D and AC
- (ii) Take $f(x) = (x^t l_C)^2 + (x^t l_{A_L C})^2 + (x^t l_{A_Q C})^2$. In this case, the resulting problem is a convex quadratic integer programming problem. Note that $f(x) = x^t P P^t x$ where $P = [l_C, l_{A_L C}, l_{A_Q C}]$.

The complete formulation of the problem with objective function as given in (i) is given below. We use the standard notation a^+, a^- . That is, for any real a , $a^+ = \max(a, 0)$ and $a^- = \max(-a, 0)$; and for any vector x , x^+ and x^- are defined by $(x^+)_i = (x_i)^+$ and $(x^-)_i = (x_i)^-$. Further, we use e for the vector of 1's. The order of e will be clear from the context.

Formulation (F1):

$$\begin{aligned}
 &\text{Minimize } \lambda_1 u + \lambda_2 v + \lambda_3 w \\
 &\text{subject to} \\
 &\quad Mx^+ - Mx^- = 0, \\
 &\quad e^t x^+ = 9, \\
 &\quad x^+ + x^- = e, \\
 &\quad -u \leq l_C^t (x^+ - x^-) \leq u \\
 &\quad -v \leq l_{A_L C}^t (x^+ - x^-) \leq v \\
 &\quad -w \leq l_{A_Q C}^t (x^+ - x^-) \leq w
 \end{aligned}$$

u, v, w, x_i^+ and x_i^- are nonnegative integers, $i = 1, 2, \dots, 18$, where $M = [l_{A_L}, l_{A_Q}, l_{B_L}, l_{B_Q}, l_{A_L B_L}, l_{A_L B_Q}, l_{A_Q B_L}, l_{A_Q B_Q}]^t$, and $\lambda_1, \lambda_2, \lambda_3$ are pre-determined positive numbers.

For any solution u, v, w, x^+, x^- of the above problem, define $x = x^+ - x^-$. Then x is the required contrast vector and the 2-level column is given by $(e + x^+)$. The constraints $x^+ + x^- = e$ and $e^t x^+ = 9$ force the vector x to be a contrast vector.

It may be noted that by taking the objective function $f(x) = (x^t l_C)^2 + (x^t l_{A_L C})^2 + (x^t l_{A_Q C})^2$ in this case, we actually find a design which optimizes the D -efficiency. This is because $|X^t X| = \alpha_1 - \alpha_2 f(x)$, where α_1 and α_2 are some positive constants.

We shall illustrate the above method with another example where it is required to construct a design with 12 TCs in Sect. 10.4.2.

10.5.2 Example with a $2^3 3^1$ Experiment

A design was to be constructed with four factors A, B, C and D with A at three levels, and B, C and D at two levels each. Furthermore, the interaction AB and BC are to be estimated. Here we have $2^3 3^1$ factorial experiment with 24 TCs in all. In order to estimate the main effects and interactions of interest, we need a minimum of 9 TCs. We shall construct a design with 12 TCs using our method. Table 10.8 presents a full factorial experiment for A, B and C . It also presents the corresponding contrast vectors for these factors.

In order to construct the required design, we need to augment the layout of Table 10.8 with another 2-level column as follows: construct a 2-level column for D so that the columns under A, B, D form a $2^3 3^1$ full factorial experiment (by doing this we ensure that AB interaction can be estimated orthogonally) and the nonorthogonalities of D with C and BC are minimized.

Table 10.8 Layout for A, B, C and the contrast vectors

TC. No.	A	B	C	l_{A_L}	l_{A_Q}	l_B	$l_{A_L B}$	$l_{A_Q B}$	l_C	l_{BC}
1	1	1	1	-1	1	-1	1	-1	-1	1
2	1	1	2	-1	1	-1	1	-1	1	-1
3	1	2	1	-1	1	1	-1	1	-1	-1
4	1	2	2	-1	1	1	-1	1	1	1
5	2	1	1	0	-2	-1	0	2	-1	1
6	2	1	2	0	-2	-1	0	2	1	-1
7	2	2	1	0	-2	1	0	-2	-1	-1
8	2	2	2	0	-2	1	0	-2	1	1
9	3	1	1	1	1	-1	-1	-1	-1	1
10	3	1	2	1	1	-1	-1	-1	1	-1
11	3	2	1	1	1	1	1	1	-1	-1
12	3	2	2	1	1	1	1	1	1	1

The mathematical formulation of this problem is given by:

$$\begin{aligned}
 &\text{Minimize } 100u + v \\
 &\text{subject to} \\
 &\quad Mx^+ - Mx^- = 0, \\
 &\quad e^t x^+ = 6, \\
 &\quad x^+ + x^- = e, \\
 &\quad -u \leq x^t l_C \leq u, \\
 &\quad -v \leq x^t l_{BC} \leq v,
 \end{aligned}$$

u, v, x_i^+ and x_i^- are nonnegative integers, $i = 1, 2, \dots, 12$, where

$$M = [l_{A_L} \ l_{A_Q} \ l_B \ l_{A_L B} \ l_{A_Q B} \ l_C]^t.$$

The coefficient of u in the objective function is chosen to be 100 so that if there is a feasible solution x ($= x^+ - x^-$) with $x^t l_C = 0$, then such a solution will emerge

as an optimal solution, which in turn, leads to estimability of all the main-effects orthogonally.

The above problem was solved using the LINGO package and the solution is given by

$$x = (-1, 1, 1, -1, 1, -1, -1, 1, 1, -1, -1, 1)^t.$$

Hence the column for D is $(1, 2, 2, 1, 2, 1, 1, 2, 2, 1, 1, 2)^t$. The dispersion matrix of the design is given in Table 10.9. It can be seen from the dispersion matrix that all the main effects can be estimated orthogonally.

Table 10.9 Dispersion matrix

Effect	μ	A_L	A_Q	B	$A_L B$	$A_Q B$	C	BC	D
μ	0.083								
A_L	0	0.125							
A_Q	0	0	0.042						
B	0	0	0	0.083					
$A_L B$	0	0	0	0	0.125				
$A_Q B$	0	0	0	0	0	0.042			
C	0	0	0	0	0	0	0.083		
BC	0	0	0	0	0	0	0	0.094	
D	0	0	0	0	0	0	0	0.031	0.094

D -Efficiency = 105.22 %

I_F -Efficiency = 97.30 %

10.5.3 Augmenting with 3-Level Columns

Consider the problem of constructing a 3-level column discussed in Sect. 10.4.2. Given the columns of A , C and D , the problem is to construct a 3-level column for B so that its three levels appear with the same frequency and the contrast vectors of B (linear and quadratic) are orthogonal to C , D and CD . In addition, the column should be chosen in such a way that it minimizes the nonorthogonalities of B with A , AB and AC .

Constructing a homogeneous column for B is equivalent to constructing the linear (and hence quadratic) contrast vector with 0s, 1s and -1 s so that they appear with the same frequency. So, for the problem in question, we must construct a (linear) contrast vector x with four 1s, four 0s and four -1 s satisfying the orthogonality constraints with l_C , l_D and l_{CD} . This means $x (= l_{B_L})$ and the resulting quadratic component of B (l_{B_Q}) must be orthogonal to l_C , l_D and l_{CD} . Observe that since x is the linear contrast vector of B (l_{B_L}), the quadratic contrast vector of B is given by $l_{B_Q} = 3(x^+ + x^-) - 2e$. Thus, in order that the columns of B , C , D form a full factorial design, we must have

$$M^t [l_{B_L} \ l_{B_Q}] = 0, \text{ where } M = [l_C \ l_D \ l_C \bullet l_D].$$

Note that the \bullet product between l_C and l_D is nothing but l_{CD} (for any two vectors u and v of same order, $u \bullet v$ is the vector w whose i th coordinate is $w_i = u_i v_i$). Since $l_{B_L} = x^+ - x^-$ and $l_{B_Q} = 3(x^+ + x^-) - 2e$, we have

$$M^t[x^+ - x^-, 3(x^+ + x^-) - 2e] = [M^t(x^+ - x^-), 3M^t(x^+ + x^-)].$$

Hence the constraints reduce to

$$\begin{bmatrix} M^t & -M^t \\ M^t & M^t \end{bmatrix} \begin{bmatrix} x^+ \\ x^- \end{bmatrix} = 0.$$

Next, consider the objective function. We wish to choose x so that the nonorthogonalities of B with A (i.e., $|l'_{A_L} l_{B_L}|$, $|l'_{A_Q} l_{B_L}|$, $|l'_{A_L} l_{B_Q}|$ etc.), the nonorthogonalities of B with AB interaction (i.e., $|l'_{B_L}(l_{A_L} \bullet l_{B_L})|$, $|l'_{B_Q}(l_{A_L} \bullet l_{B_L})|$, $|l'_{B_L}(l_{A_L} \bullet l_{B_Q})|$, etc.) and the nonorthogonalities of B with AC interaction (i.e., $|l'_{B_L}(l_{A_L} \bullet l_C)|$ etc.) are minimized. An interesting observation is that all these nonorthogonalities are linear functions of x^+ and x^- . To see this, first observe that for any vectors p , q , u , and v (of same order),

- (i) $p^t(u \bullet v) = u^t(p \bullet v)$,
- (ii) $(p + q) \bullet (u + v) = p \bullet q + q \bullet u + p \bullet v + q \bullet v$.

So

$$\begin{aligned} l'_{B_L}(l_{A_L} \bullet l_{B_L}) &= l'_{A_L}(l_{B_L} \bullet l_{B_L}) \\ &= l'_{A_L}((x^+ - x^-) \bullet (x^+ - x^-)) \\ &= l'_{A_L}(x^+ \bullet x^+ - x^+ \bullet x^- - x^- \bullet x^+ + x^- \bullet x^-) \\ &= l'_{A_L}(x^+ + x^-) \end{aligned}$$

(as $x^+ \bullet x^- = x^- \bullet x^+ = 0$ and $x^+ \bullet x^+ = x^+$, $x^- \bullet x^- = x^-$)

and

$$\begin{aligned} l'_{B_Q}(l_{A_L} \bullet l_{B_L}) &= l'_{A_L}(l_{B_Q} \bullet l_{B_L}) \\ &= l'_{A_L}((3(x^+ + x^-) - 2e) \bullet (x^+ - x^-)) \\ &= l'_{A_L}(x^+ - x^-) \\ &= l'_{A_L} l_{B_L}. \end{aligned}$$

Similarly we can show that other nonorthogonalities are also linear functions of x^+ and x^- .

For simplicity, we take the objective function as $|l'_{A_L} l_{B_L}| + |l'_{A_Q} l_{B_L}| + |l'_{A_L} l_{B_Q}|$. The complete formulation is given below.

Formulation (F2):

$$\begin{aligned}
& \text{Minimize } u + v + w \\
& \text{subject to} \\
& \begin{bmatrix} M^t & -M^t \\ M^t & M^t \end{bmatrix} \begin{bmatrix} x^+ \\ x^- \end{bmatrix} = 0, \\
& e^t x^+ = 4, \\
& e^t x^- = 4, \\
& x^+ + x^- \leq e, \\
& -u \leq l_{AL}^t l_{BL} \leq u, \\
& -v \leq l_{AQ}^t l_{BL} \leq v, \\
& -u \leq l_{AL}^t l_{BQ} \leq w,
\end{aligned}$$

u, v, w, x_i^+, x_i^- are nonnegative integers, $i = 1, 2, \dots, 12$.

The problem was solved using the LINGO package. The resulting layout and the corresponding dispersion matrix are given in Sect. 10.4.2. It must be mentioned that when we included the $A_Q B_Q$ component in the objective function, the program execution terminated with the conclusion that the problem was nonoptimal/infeasible.

10.5.4 Enlisting All Solutions

Consider the problem of finding all columns (of 2 or 3-levels) that are orthogonal to a given set of columns. It is possible to solve this problem iteratively using the integer programming formulations. Wang and Wu [19, 20] constructed a number of orthogonal and nearly orthogonal main effect plans for $2^m 3^n$ factorial experiments. In [20] it was shown that there are 16 2-level columns orthogonal to 3 given columns (one 3-level and two 2-level) – all having the same level in the first coordinate – and enlist all the 16 2-level columns (see Section 2, page 411 of [20] for details).

We illustrate our methodology to enlist all columns orthogonal to a given set of columns. Consider problem **F1** formulated in Sect. 10.5.1. Suppose (x^+, x^-) is a feasible solution of **F1** (u, v, w are not mentioned here as they are determined by x^+ and x^-). Define $\alpha = \{i : x_i^+ > 0\}$ and $S = \{(z^+, z^-) : (z^+, z^-) \text{ is a feasible solution to } \mathbf{F1}\}$. Augment the constraints of **F1** with $\sum_{i \in \alpha} x_i^+ \leq 8$. Let $S^* = \{(z^+, z^-) : (z^+, z^-) \text{ is a feasible solution to the augmented problem}\}$. Note that $S^* = S \setminus \{(x^+, x^-)\}$. Thus by finding a feasible solution to the augmented problem, we get a new feasible solution to **F1**. By repeating the above process (by adding a new constraint in every iteration) we can generate all feasible solutions to **F1**. Solving the Wang and Wu's problem mentioned in the previous paragraph (with the additional constraint $x_i^+ = 1$ to ensure that all columns have one in the first row), we find that there are exactly 16 distinct solutions (columns) and in the 17th iteration, the augmented problem becomes infeasible.

10.6 Summary

Factorial Experiments are extensively used in industrial experimentation and other disciplines. The experiments often use asymmetrical factorial experiments. Constructing efficient designs for asymmetrical factorial experiments is, in general, a complex problem. In this article, we have considered $2^m 3^n$ factorial experiments with interactions. We have presented a methodology for construction of nearly orthogonal designs which are regular and homogenous. By formulating the problem of construction as an integer programming problem, we have shown that the designs can be constructed using OR packages. This has been demonstrated with examples. We have remarked that by minimizing the nonorthogonality, we expect that the resulting design will lead to estimability of the main effects and interactions of interest. We conclude this chapter with the following open question: Is there a bound on the nonorthogonal objective functions considered in the formulations which will ensure nonsingularity of $X^t X$ (in other words, estimability of parameters of interest)?

Problems

10.1. Write down the contrast vectors of all main effects and interactions for the design given in Table 10.1. For the response vector $y = (5, 8, 0, 9, 2, 7, 1, 1, 5, 8, 3, 6, 2, 0, 1, 0, 2, 5)$, find out all the main effects and interactions, both linear and quadratic. Divide the effects into small and large as per your assessment.

10.2. For the design given in Table 10.1, find out which effects are uncorrelated. For the correlated effects, find out the correlations.

References

1. Addelman, S.: Orthogonal main-effect plans for asymmetric factorial experiments. *Technometrics* **4**, 21–46 (1962)
2. Anderson, D.A., Thomas, A.M.: Resolution IV fractional factorial design for the general asymmetric factorial. *Technometrics Commun. Stat. Theory Methods* **8**, 931–943 (1979)
3. Box, M.J., Draper, N.R.: Factorial designs, the $|X^t X|$ criterion, and some related matters. *Technometrics* **13**, 731–742 (1971)
4. Box, G.E., Hunter, W.G., Hunter, J.S.: *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd edn. Wiley, Hoboken (2005)
5. Chakravarti, I.M.: Fractional replication in asymmetric factorial designs and partially balanced arrays. *Sankhya* **17**, 143–164 (1956)
6. Cheng, C.S.: Some orthogonal main-effect plans for asymmetric factorials. *Technometrics* **31**, 475–4477 (1989)
7. Hedayat, A.S., Sloane, N.J.A., Stufken, J.: *Orthogonal Arrays, Theory and Applications*. Springer, New York (1999)
8. Jeff Wu, C.F., Hamada, M.S.: *Experiments: Planning, Analysis, and Optimization*. Wiley, Hoboken (2009)

9. Mitchell, T.J.: An algorithm for construction of D -optimal designs. *Technometrics* **16**, 203–210 (1974)
10. Mitchell, T.J.: Computer construction of “ D -optimal” first order designs. *Technometrics* **16**, 211–220 (1974)
11. Montgomery, D.C.: *Design and Analysis of Experiments*. Wiley, Singapore (2013)
12. Murthy, G.S.R., Manna, D.K.: Optimization in $2^m 3^n$ factorial experiments. *Algorithm. Oper. Res.* **7**, 71–82 (2013)
13. Prasad, C.R.: *Industrial Applications of Orthogonal Array Designs in Quality Case Studies* Edited by Prasad. C. R. Indian Statistical Institute, Kolkata (1989)
14. Raghavarao, D.: *Constructions and Combinatorial Problems in Design of Experiments*. Wiley, New York (1971)
15. Rao, C.R.: *Linear Statistical Inference and Its Applications*. Wiley, New York (1974)
16. Taguchi, G.: *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*. UNIPUB, White Plains (1987)
17. Taguchi, G.: *Introduction to Quality Engineering*. Asian Productivity Organization. UNIPUB, White Plains (1991)
18. Taguchi, G., Wu, Y.: *Introduction to Off-Line Quality Control*. Central Japan Quality Control Association, Nagoya (1980)
19. Wang, J.C., Wu, C.F.J.: An approach to the construction of asymmetrical orthogonal arrays. *JASA* **86**, 450–456 (1991)
20. Wang, J.C., Wu, C.F.J.: Nearly orthogonal arrays with mixed levels and small runs. *Technometrics* **34**, 409–422 (1992)
21. Webb, S.R.: Nearly orthogonal arrays with mixed levels and small runs. Small incomplete factorial experiment design for two- and three-level factors. *Technometrics* **13**, 243–256 (1971)

Index

- adjustment of knives, 126
- agents associates, 139
- amino acids, 163
- analysis and results, 11
- analysis of past data, 156
- analysis of the reference product, 169
- AP problem, 26
- aplication, 28
- application of tool to live data, 149
- application to the real problem, 86
- approach, 6, 142
- April data, 186
- arbitrary censor, 187
- assignment problem, 25
- assumed prices, 70
- asymmetric TSP, 128
- asymmetrical factorial experiment, 253
- atomic weight, 163
- Augmenting with 2-level columns, 259
- Augmenting with 3-level columns, 262
- ayacut, 216
- background, 162
- bare pcb, 19
- basic assumptions, 143
- basic objects, 80
- blanking operation, 2
- blanking sheets, 4
- brightness, 196
- bulk, 196
- bulk density, 110
- bursting strength, 196
- business process outsourcing, 138
- caliper, 94
- call volumes, 143
- calls answered on a normal day, 156
- CATIA measurement, 235
- cello tape problem, 99, 126
- censoring, 187
- central drugs laboratory, 38
- changeover time, 127
- child reels, 95
- clinical trials, 37
- cluster analysis, 174
- clustering technique, 242
- color spaces, 236
- combined analysis, 177
- company's product, 168
- company's solution, 153
- Company's Test Procedure, 185
- comparison of results, 245
- complete formulation, 68
- complete solution, 131

- complexity, 69
- complexity of generating cutting patterns, 133
- component correlations, 175
- computer programs, 134
- conclusion of peptide study, 168
- constraints of the staffing problem, 144
- Construction in 12 runs, 257
- Construction in 18 runs, 255
- Construction Method, 255
- contact patch, 232, 233
- contrast, 251
- contrast vector, 251
- contribution, 54
- core pipe, 94
- cost elements, 57
- cross direction, 94
- crucial observation, 25
- Cutting Patterns, 99
- cutting plan, 5
- Cutting Plans, 100
- cutting stock problem, 98
- decision making problem, 24
- decision support system, 86, 137
- decision variable, 60
- decision variables, 143
- deckle, 97
- deckle optimization problem, 112
- deckle width, 97
- definition, 65
- dendrogram, 174
- density function, 188
- description of PCBSOft, 31
- description of the problem, 39
- design matrix, 253
- determination, 162
- diameter for sheet orders, 134
- discounts, 54
- dispatch plan, 51
- dispersion matrix, 253
- distance between cutting patterns, 127
- distance matrix, 128
- DNA sequence, 163
- drug authority, 37
- dynamic review, 155
- earliest due date, 143
- effect of reduction, 77
- effective freight, 59
- efficiency measure, 133
- efficient design, 253
- evolution of ideas, 164
- excise duty, 54
- expenditure coefficients, 73
- extended model, 71
- extension of ideas to general problems, 166
- features, 147
- feeder allocation, 22
- feeder location, 21
- first planning date, 143
- first solution using simulation, 234
- footprint, 232
- Formulation, 100
- formulation, 9, 24, 55, 144, 219
- formulation of stage-1 problem, 145
- formulation of stage-2 problem, 146
- Formulation 2, 75
- generation of new cutting patterns, 7
- grain any way, GAW, 94
- grain long way, GLW, 94
- grain short way, GSW, 94
- group constraint, 225
- GSM, 94
- handling charges, 54
- handling priorities, 225
- hazard function, 188
- healthcare industry, 139
- hepatitis, 37
- Heuristic, 102
- heuristic, 28
- high-density polyethylene, 53
- homogenous fraction, 252
- HSB color space, 237

- impact of clustering method, 243
- in vitro, 38
- in vivo, 38
- information technology enabled services, 138
- input screen, 149
- inputs, 56
- interesting observation, 241
- irrigation, 216
- irrigation system, 217

- J SPL, Jindal Steel and Power Limited, 2

- LAB color space, 239
- land sea ratio, 232, 233
- last date of planning, 143
- left censor, 187
- level constraint, 224
- LINGO Model, 80
- load default capacities, 148

- machine direction, 94
- Mahalanobis distance, 198
- main effect, 251
- Main Effects and Interactions, 251
- main screen, 147
- management decisions, 54
- matched-pair problem, 46
- maximum number of calls per day, 148
- Method 1, 102
- Method 2, 104, 108
- miniature version, 60
- minimax problem, 144
- minimum cost network flow problem, 226
- minimum flow constraints, 223
- Model of the Factorial Experiments, 252
- model simplification, 69
- moisture content, 196
- molecular weight, 163
- Monte Carlo simulation, 192

- new cutting patterns, 9
- new formulation, 73
- new plant constraint, 67
- nonparametric method, 45
- normal working days, 156
- notation, 56
- number censored, 185
- number of reels for sheet order, 134

- objectives, 5, 55, 141, 218
- observation for yield, 8
- offcut, 128
- One-dimensional cutting stock problem, 100
- opacity, 196
- opening calls, 144
- operations research, 51
- optimal sequence for deckle problem, 130
- OR Formulations, 259
- Ordinary Portland cement, 52

- parametric method, 43
- past data analysis, 80
- Pennar Delta, 216
- permissible options, 65
- pharmaceutical, 37
- plan for Sep'2010, 153
- planning at winder, 108
- planning horizon, 56
- plant loading, 51, 78
- polypeptide, 162, 163
- Pooling orders, 112
- Pozzolanic Portland cement, 52
- press shop unit, 2
- principal component analysis, 172
- principal component approach, 239
- printed circuit board, 19
- probability plot, 188
- problem background, 139
- problem description, 139, 233
- Problem Instance 1, 103
- Problem Instance 2, 104
- problem size reduction, 74
- problems, 87
- process, 2, 20

- process control, 37
- process optimization, 37
- product combination, 53
- product demand constraint, 68
- product quality rating, 196
- production capacities, 57
- production of PCBs, 23
- production restrictions, 108
- production schedule, 5

- question of uniqueness, 165

- radial tyres, 231
- reduced problem, 75, 83
- reel orders, 94
- reference product, 168
- regression approach, 40
- regular fraction, 252
- relative potency, 40
- reliability function, 188
- reservoirs, 218
- resolution of cycling, 228
- resources, 155
- results, 80
- results and conclusions, 46
- RGB color space, 237
- right censor, 187
- rolling resistance, 232
- run length, 252

- SAIL, 2
- sample problem, 108
- screeplot, 177
- sequencing cutting patterns, 124, 128
- service level agreement, 138
- sheet orders, 94
- sheet-metal, 2
- sheeter facilities, 134
- sheeters, 97
- siping, 233
- size reduction, 69
- software development, 30, 225
- software solution, 12
- software tool, 146
- solution to the client's problem, 167

- special constraints, 29
- Stage-1 optimal solution, 154
- Stage-2 optimal solution, 154
- stalling, 228
- statement of the problem, 163
- statistical analysis, 40
- stochastic comparison, 42
- stockist, 59
- stockists, 51
- Stokist Point Offices, 52
- stopping distance, 232
- structure of contribution, 72
- Sulphate Resisting Portland cement, 52
- summary, 14, 33, 87, 159, 178, 229, 245
- symmetric TSP, 128

- technical discussions, 9
- tensile strength, 196
- time censored, 185
- transport capacity constraint, 67
- transportation capacities, 59
- travelling salesman problem, 26, 128
- tread of tyre, 233
- treatment combination, 250
- TSP, 128
- TSP problem, 26
- turn around time, 138
- tyres, 231
- tyres manufacturing company, 231

- unknown prices, 70
- utilization percentage, 148

- vaccine, 38
- variable cost, 54

- width of sheet order, 95
- winder, 95
- winder problem, 102
- World Health Organization, 38

- XYZ color space, 239
- YUV color space, 238