Operations Research for Social Good: A Practitioner's Introduction Using SAS® and Python

Natalia Summerville
Rob Pratt

The correct bibliographic citation for this manual is as follows: Summerville, Natalia and Rob Pratt. 2023. *Operations Research for Social Good: A Practitioner's Introduction Using SAS® and Python.* Cary, NC: SAS Institute Inc.

Operations Research for Social Good: A Practitioner's Introduction Using SAS® and Python

ISBN 978-1-68580-005-5 (Hardcover)
ISBN 978-1-955977-83-8 (Paperback)
ISBN 978-1-955977-84-5 (Web PDF)
ISBN 978-1-955977-85-2 (EPUB)
ISBN 978-1-68580-004-8 (Kindle)

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

October 2023

Contents

About this Book

What Does This Book Cover?

This book's purpose is to showcase Operations Research (OR) methodologies to applications targeted to make this world a better place. This book also provides skills and practical examples to model and solve OR problems with both SAS and Python.

Each use case is a real-life application that has been implemented and proven successful. We solve use cases with both SAS and Python, driving students to learn both programming languages to solve OR problems and giving professors flexibility to choose which technology to focus on in their classes.

This book does not cover operations research theory or optimization algorithms. Instead, it focuses on problem modeling and formulation.

Is This Book for You?

This book is for data scientists, analytics and operations research practitioners, and graduate-level students interested in learning optimization modeling with applied use cases.

What Are the Prerequisites for This Book?

Knowledge of linear algebra (specifically algebraic summation syntax) is needed.

What Should You Know about the Examples?

All examples in the book are formulated with SAS and Python, providing helpful coding syntax to the readers. All applications are based on real-life Data4Good projects.

Software Used to Develop the Book's Content

SAS OPTMODEL and Python/Pyomo.

Example Code and Data

You can access the example code and data for this book by linking to its author page at https://support.sas.com/authors.

SAS OnDemand for Academics

This book is compatible with SAS OnDemand for Academics. If you are using SAS OnDemand for Academics, then begin here: https://www.sas.com/en_us/software/on-demand-for-academics.html.

Where Are the Exercise Solutions?

Selected problem solutions can be found at the end of the book.

We Want to Hear from You

SAS Press books are written by SAS Users for SAS Users. We welcome your participation in their development and your feedback on SAS Press books that you are using. Please visit sas.com/books to do the following:

- Sign up to review a book
- Recommend a topic

- Request information on how to become a SAS Press author
- Provide feedback on a book

Learn more about these authors by visiting their author pages, where you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more:

- https://support.sas.com/summerville
- https://support.sas.com/pratt

About the Authors

Natalia Summerville is the Director of Applied Data Science in the Strategy and Innovation Division at Memorial Sloan Kettering Cancer Center. Her team develops data analytics products to support hospital strategy and innovations in care delivery as well as cutting-edge cancer research. Previously, she led a team of Operations Research and Machine Learning experts at SAS, building analytical engines for customers across industries such as Health Care, Life Sciences, Retail, and Manufacturing. Natalia has been teaching undergrad and grad-level classes in Operations Research, Data Analytics, and Machine Learning since 2005, and she is currently an Adjunct Professor at Duke University. She is deeply passionate about the Data4Good movement and has been collaborating with many non-profit and mission-driven organizations to implement data analytics for social good. She is a board member within the "Pro-Bono Analytics" committee and is part of the "Franz Edelman Award" committee at INFORMS.

Rob Pratt has worked at SAS since 2000 and is a Senior Manager in the Scientific Computing Department in the Analytics R&D Division. He manages a team of developers responsible for the optimization modeling language and solvers for linear, mixed integer linear, quadratic, and conic optimization. He earned a B.S. in Mathematics (with a second major in English) from the University of Dayton and both an M.S. in Mathematics and a Ph.D. in Operations Research from The University of North Carolina at Chapel Hill.

Learn more about these authors by visiting their author pages, where you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more:
http://support.sas.com/summerville
http://support.sas.com/pratt

Chapter 1

Introduction: Mathematical Optimization and the Data4Good Movement

Data4Good is a broad initiative, encompassing many types of analytics implementations for nonprofit organizations and/or organizations with missions that focus on the greater good. Examples of Data4Good projects include humanitarian logistics supporting disaster relief, cancer treatment innovation, equitable access to children's playgrounds, and deforestation forecasting, among many others. Typically, these implementations are performed by data scientists and analytics professionals on a pro bono/volunteer basis due to limited budgets available for analytics within these organizations. Over the last decade, the Data4Good movement has been significantly expanding, motivating more and more analytics professionals to bring their skills to support mission-driven organizations.

However, most of these applications focus on descriptive/diagnostic analytics, sometimes on predictive analytics, and rarely on prescriptive analytics. Traditionally, only analytically mature organizations built end-to-end prescriptive analytics engines that included optimization models. This is mostly due to the specific (and scarce) mathematical expertise required to properly formulate optimization models that often need PhD-level skills, available data to support these formulations, and established processes to incorporate new decision-making support systems that focus on user adoption and end value. Despite the reduced number of optimization projects in Data4Good (as opposed to descriptive and predictive modeling projects), we are firm believers that optimization tools can be key to help these mission-driven organizations make better decisions and be more efficient in using their very limited resources.

In this book, we introduce optimization modeling concepts that can help any organization be more efficient but with Data4Good applications. All applications discussed in this book come from proven real-life implementations, albeit often simplified for teaching purposes.

We hope that by studying this book, you will not only familiarize yourself with optimization modeling and scripting (in both SAS and Python) but also learn heartwarming applications where optimization can make this world a better place.

Chapter 2

Mathematical Optimization Landscape

Mathematical optimization provides organizations with actionable insights and results that are fundamentally geared toward improving organizational efficiency. This value-driven focus places optimization on top of Prescriptive Analytics, a field that generates the highest competitive advantage to those organizations who decide to use their data to build and implement optimization tools. But before we dive into definitions and specific characteristics of mathematical optimization, let's review the three main Advanced Analytics areas and how they relate to each other.

2.1 Areas of Advanced Analytics

Advanced Analytics is typically classified into three (or four, depending on the source) categories based on their usage and competitive advantage for the organization. These areas are Descriptive/Diagnostic (some authors split these two into separate categories), Predictive, and Prescriptive, as shown in Figure 2.1.

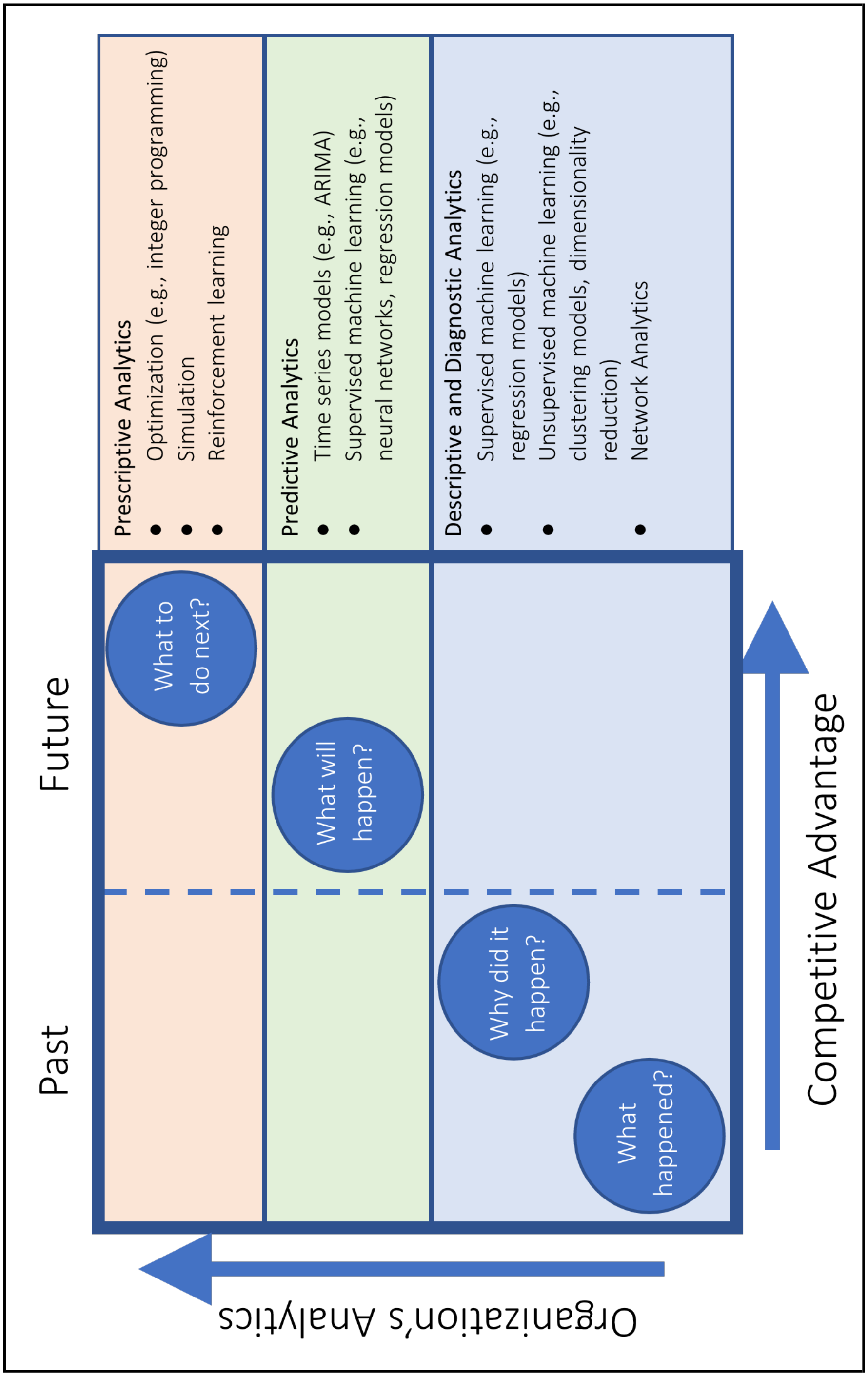

Figure 2.1: Areas of Advanced Analytics

Descriptive/Diagnostic Analytics

Descriptive and/or Diagnostic Analytics focuses on using data analysis to understand what has happened and why it has happened. Besides basic data analysis techniques such as scatter plots and correlation analysis, the most used Advanced Analytics models include:

- Clustering (Unsupervised Machine Learning) to understand groups of observations and their similarities
- Network Analytics to describe patterns in interconnected data
- Regression Analysis (Supervised Machine Learning) to understand causal relationships

For example, we might want to understand the differences between groups of patients based on their molecular characteristics from lab tests using clustering techniques. We might also be interested in identifying the most relevant production settings that influence key quality metrics in wallboard manufacturing using regression models.

Predictive Analytics

Predictive Analytics uses statistical analysis to forecast future states. Besides naïve forecasting techniques such as year-over-year and moving averages, some typical forecasting models are:

- ARIMA models (Time Series Forecasting) to derive historical patterns from past sequential data and predict future observations by using those historical patterns
- Recurrent Neural Networks (Supervised Machine Learning) to predict future states based on previous states and their interactions

For example, time series models would forecast weekly product sales for a specific grocery store or expected daily arrivals for labor and delivery unit in a hospital.

Prescriptive Analytics

Prescriptive Analytics focuses on providing the best possible future action to achieve organizational goals. Besides techniques such as heuristic rule-based approaches and decision analysis, the most typically used Advanced Analytics tools are:

- Optimization, which includes an algebraic representation of the business problem, including relevant goals, key performance indicators (KPIs), rules, and limitations, as well as mathematical algorithms that find the best possible decisions that satisfy those rules while maximizing or minimizing those KPIs
- Simulation models to build a digital system representation, including stochastic distributions of relevant parameters, and to run what-if analysis and evaluate decision options
- Markov Chains that model systems where there are transitions between states according to probabilistic rules

For example, we might need to find the best possible schedule for retail employees, aiming to cover shifts with highest demand while minimizing overtime using mathematical optimization.

The three Advanced Analytics areas are closely related, and typically all of them are required for a successful analytical implementation. For example, very often within optimization models we need to incorporate forecasted demand for a product, or relationships between manufacturing settings and relevant KPIs, which in turn use time series and regression models, respectively.

2.2 Process to Produce an Optimal Solution

To generate an optimal solution via mathematical optimization, four main steps need to happen after thorough data exploration, validation, and predictive model building (if required). These steps are highlighted in Figure 2.2.

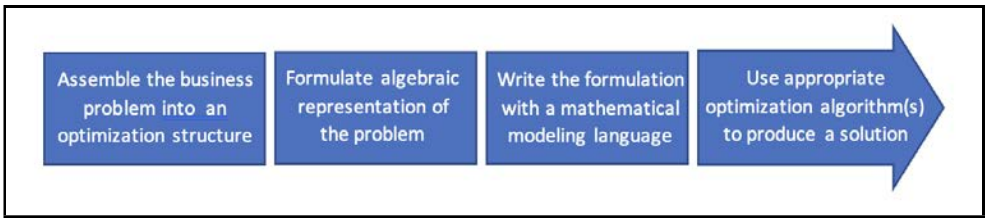

Figure 2.2: Optimization Process

We first need to assemble the problem into an optimization structure, which includes identifying the following components (also presented in Table 2.1):

Table 2.1: Optimization Components

Optimization Component	Definition	Examples
Decision Variables	Controllable actions	Promotion discounts Selection of investment funds Classroom assignment to student groups
Constraints	Rules and limitations	Do not let profit be negative Stay within available budget Each group needs to have a classroom
Objective Function(s)	Goals for key performance indicators	Maximize revenue Minimize final inventory Maximize classroom utilization

- Decision variables are the controllable actions that users can take. For example, a pricing analyst decides how much price promotion discount to allocate to certain products and when.
- Constraints are all the rules and limitations that restrict those decisions. For example, the promotion discounts must not lead to a negative profit across all products.
- Objective functions are specific goals for the KPIs that the organization wants to achieve in this decision-making process. For example, the user might want to maximize revenue.

It is typically helpful to have those components written in a natural language before moving to the next step. For example:

- I need to decide how to price my products (decision variables).
- I want to achieve highest revenue (objective function).
- Margin cannot be negative (constraint).
- Demand for each product is expected to be $100 - 0.2 * Price$ (constraint).

Please notice that the demand equation above (albeit simplified for this example) requires a predictive model that explains the relationship between price and demand.

Once there is clear understanding of the components described above, we need to formulate this problem mathematically, using appropriate algebra. For example:

$$
\begin{aligned}
\text{maximize} \quad & \sum_p Demand_p \times Price_p \\
\text{subject to} \quad & \sum_p (Price_p - Cost_p) \geq 0 \\
& Demand_p = 100 - 0.2 Price_p \quad \text{for all } p \in \text{PRODUCTS} \\
& Price_p \geq 0 \quad \text{for all } p \in \text{PRODUCTS}
\end{aligned}
$$

The next step is expressing the algebraic formulation in a mathematical programming language like SAS OPTMODEL or Python Pyomo that use an intuitive coding syntax to facilitate an easy translation between the math and the code.

The final step is to call an efficient algorithm (sometimes called a solver) to produce an optimal solution. In the example above, the algorithm would return an optimum price for each product that would generate the maximum revenue while making sure profit is nonnegative.

2.3 Types of Optimization Models

Optimization models are classified based on the mathematical characteristics of the algebraic representation of the problem such as linearity in constraints/objective functions and types of decision variables used. Some of the most used optimization models include Linear Programming (LP), Integer Programming (IP), Mixed Integer Linear Programming (MILP), Nonlinear Programming (NLP), and Multicriteria Optimization. This list is not exhaustive. Differences among these types of models are shown in Table 2.2.

Table 2.2: Most Used Types of Optimization

Optimization Component	Linear Programming	Mixed Integer Linear Programming	Nonlinear Programming	Multicriteria Optimization
Decision Variables	Continuous	Continuous, Integer, or Binary	Continuous	Continuous, Integer, or Binary
Constraints	Linear	Linear	Nonlinear	Linear or Nonlinear
Objective Function(s)	One	One	One	More than one

Network optimization models deserve special mention. Many well-studied formulations such as Shortest Path or Traveling Salesman have MILP models. However, their specific structural characteristics enable specialized algorithms to generate solutions much faster than a typical MILP algorithm such as Branch-and-Bound would. Therefore, it is often desirable, when appropriate, to reformulate MILP models as Network models to take advantage of their specific structure and specialized algorithms.

2.4 Optimality and Algorithmic Performance

Some optimization problems can be very complex (in terms of number of decision variables, number of constraints, nonlinearity, and integrality of variables, among others). This complexity translates into potentially long running times for the solution algorithms and sometimes inability to achieve an optimal solution. However, in practice, many times even if a feasible solution is not globally optimal, it can still provide significant business value over the status quo.

2.5 Example Application – Medical Resource Optimization

In March 2020, the COVID-19 global pandemic forced hospitals to reassign most of their resources in order to maintain Emergency Rooms (ERs) and Intensive Care Units (ICUs) ready to support the population's growing need for emergency COVID-19 care and put on hold many of the elective surgeries and procedures that clinics and hospitals offered. SAS and Cleveland Clinic partnered to develop a mathematical model to support decision-making regarding reopening these optional services, considering the capacity limitations on manpower, equipment, and COVID-19 tests, among others.

This problem was formulated as a Multicriteria MILP mathematical model. The components of the model are described as follows.

Decision Variables

Controllable decision variables in this model were defined as the selection of subservices to reopen and the reopening dates.

Constraints

The reopening of subservices and number of patients accepted had to adhere to many constraints, the most significant of which are:

- The capacity of each resource at a facility, their services, and their corresponding subservices cannot be exceeded.
- The utilization of ICU resources at a facility cannot exceed a specified upper limit.
- The number of patients accepted at a facility will never exceed the maximum forecasted demand.
- The total number of daily emergency surgery and patients accepted across all facilities should not exceed the number of daily rapid tests available.
- If a subservice is open at a facility/service-line on a day, it should remain open for the remainder of the horizon (logical condition).

Objective Functions

This model had two objective functions to reboot clinics' cash flow, which in turn enables further support for emergency care:

- Maximize the total revenue
- Maximize the total margin

Revenue and margin are functions of the number of patients accepted over the planning horizon.

More detailed information, as well as the full code, can be found at https://github.com/sassoftware/medical-resource-optimization.

Many other very useful optimization model examples and their SAS code can be found at
https://support.sas.com/rnd/app/examples/ORexamples.html#MPE.