Models and Solution Approaches for Large-scale Multistage Stochastic Programs with Endogenous or/and Exogenous Uncertainty

by

Zuo Zeng

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 12, 2020

Keywords: Multistage Stochastic Programs; Endogenous Uncertainty; Exogenous Uncertainty; solution approaches

Approved by

Selen Cremaschi, B. Redd Professor of Chemical Engineering
Mario R. Eden, Chair and McMillan Professor of Chemical Engineering
Jing Wang, Walt and Virginia Woltosz Professor
Andrew Adamczyk, Assistant Professor
John Siirola, Sandia National Laboratory

Abstract

This dissertation aims to develop both models and solution approaches of multistage stochastic programs (MSSP) with endogenous and/or exogenous uncertainty. My work first develops two discrete-time large-scale nonconvex mixed-integer nonlinear programs (MINLP) models to solve the artificial lift infrastructure-planning problem. The problem is determining: (a) which artificial lift method should be installed, (b) when exactly the artificial lift method should be installed, and uninstalled. The objective is to maximize the net present value (NPV). Using the unique structure of the nonlinear terms, we formulate two equivalent mixed-integer linear programs MILP. Sets of tightening inequalities are used to decrease the solution times for both MINLP and MILP up to two orders of magnitude. We develop the MSSP model of artificial lift infrastructure planning problems by incorporating endogenous uncertainty of production rate, where the objective is to maximize the expected net present value (ENPV). The improvement and difference between the ENPVs of the stochastic and deterministic solutions, i.e., the value of the stochastic solution, is $163,831, a 5% increase from the deterministic solution. Secondly, we propose two alternatives multi-stage stochastic programming formulations to determine the optimal clinical trial plan under uncertainty. Decisions of a clinical trial plan include which clinical trials to start and their start times. Its objective is to maximize the expected net present value of the entire clinical trial plan.

The MSSP model grows quickly and becomes computationally intractable for real-world sized problems due to their space and time complexities with increases in the number of uncertain parameters or increases in the number of realizations for each uncertain parameter. To address the complexity of MSSPs, my work to date has investigated different solution approaches for solving large-scale MSSPs with endogenous and/or exogenous uncertainties.

The first approach develops a generalized Knapsack-problem based Decomposition Algorithm (GKDA) built upon the KDA (Christian and Cremaschi, 2014) to efficiently obtain feasible solutions for large-scale MSSPs with endogenous and/or exogenous uncertainties. The GKDA decomposes the original multi-period MSSP into a series of knapsack problems and solves these problems at appropriate decision points of the planning horizon. The GKDA is applied to obtain feasible solutions for four different planning problems, which include continuous and discrete decision variables, recourse action variables, and both exogenous and endogenous uncertainties.

The second approach develops a general primal-bounding framework based on extending the concepts of expected value solution and the value of the stochastic solution from multistage stochastic programs under exogenous uncertainties. It yields a tight feasible bound, and an implementable solution for multistage stochastic programs under endogenous uncertainties called the absolute expected value solution (AEEV).

The third approach presents a new Lagrangian relaxation for obtaining a valid dual bound for MSSPs under endogenous uncertainties. By exploiting the structure of the MSSP, a tight dual problem is formulated, which reduces the total number of Lagrange multipliers with a modified multiplier updating scheme. The new Lagrangian relaxation is applied to bound instances of artificial lift infrastructure planning problem under uncertain production rate and the clinical trial planning problem under uncertain clinical trial outcomes.

The fourth approach presents a Relaxed Knapsack-problem based decomposition Algorithm (RKDA) to efficiently generate tight dual bounds for large-scale MSSPs of resource-constrained multiple projects scheduling with stochastic task success. Unlike Lagrangian relaxation regarding and relaxing NACs as complicating constraints, the RKDA relaxes the resource constraints and embedding non-anticipativity of planning decisions in its framework.

We develop the decomposition frameworks utilizing both GKDA and Lagrangian decomposition algorithm. The frameworks decompose original MSSP to individual scenario subproblems and are applied to instances of clinical trial planning problems. The primal and dual bounds obtained from GKDA and Lagrangian decomposition are updated iteratively. We also explore different frameworks with dual and primal bounding approaches for solving the MILP of the random forest model, which has similar complicating constraints with two-stage stochastic programs.

Acknowledgments

First and foremost, I would like to thank Professor Selen Cremaschi for being a wonderful advisor and for her persistence guidance, support, patience, and understanding throughout my research work at Auburn University. I have learned so much from her. She has contributed to my development, both personally and professionally.

Many thanks to Dr. Mario Eden, Dr. Jing Wang, Dr. Andrew Adamczyk, Dr. John Siirola, and Dr. Aleksandr Vinel, for their invaluable time and inputs. I would also like to thank my colleagues at Auburn University for our fruitful discussions and for providing such an enjoyable working environment.

Finally, I would like to express my deepest gratitude to my parents (Changfa Zeng and Xiangmei Feng) for their love and support, and all the other friends (Haichao Wu, Wenyang Lu, Xing He) for sending their positive energy throughout my Ph.D.

Table of Contents

LIST OF TABLES

List of Figures

CHAPTER 1

INTRODUCTION

In industry, planning and scheduling with discrete decision process always involved independent uncertainties (exogenous uncertainty) and decision dependent uncertainties (endogenous uncertainties). Some problems involve with endogenous uncertainties have already been explored, e.g., synthesis of process networks with uncertain process yields (Goel and Grossmann, 2007; Tarhan and Grossmann, 2008), R&D pipeline management (Colvin and Maravelias, 2008, 2009, and 2010), the oil field infrastructure planning problem (Goel and Grossmann, 2004); Goel et al., 2006) and artificial lift planning (Zeng and Cremaschi, 2017a). Multistage stochastic programming (MSSP) is an approach for optimization problems with discrete decision processes under uncertainties. MSSP is a scenario-based method and considers decisions and recourse actions in multiple stages. In MSSP formulation, the whole scenarios are generated based on all realizations of uncertain parameters. The decision variables are defined independently for each scenario. The non-anticipativity constraints are introduced to prevent the decisions anticipating the unrealized future outcomes. Unfortunately, these stochastic programming formulations grow exponentially with the increases in the numbers of scenarios and periods and quickly become computationally intractable for real-world sized problems. Thus, the MSSP results in both space and computational complexity: the solver can not generate the model and obtain the optimal solution within a reasonable wall time.

In general, the solution approaches for large-scale MSSPs with endogenous uncertainty rely on heuristic, approximation, and decomposition methods, such as sample average approximation algorithm (Solak et al., 2010), improved Lagrangian decomposition framework (Gupta and Grossmann, 2014), sequential scenario decomposition approach (Apap and Grossmann, 2017), a

branch and bound algorithm (Christian and Cremaschi, 2017) and a rolling-horizon heuristic approach (Colvin and Maravelias, 2009). It has been shown that moderate-size problems can be solved to optimality.

1.1 Objectives

This dissertation will:

(1) Explore and develop new MSSP models under endogenous and/or exogenous uncertainties.

(2) Investigate and develop novel heuristic approaches to address the computational complexity of large-scale MSSPs, and apply heuristic approaches to quickly generate a feasible solution, a primal and dual bound for large-scale MSSP problems.

1.2 Organizations

Chapter 2 provides the literature review on models and existing solution approaches for large-scale MSSPs with endogenous uncertainty. In Chapter 3, two planning problems (artificial lift infrastructure planning problem and pharmaceutical clinical trial planning problem) and their MSSP formulations are introduced. The case studies and results comparison are also introduced in Chapter 3. Section 3.1 presents the deterministic and stochastic mathematical programming models to solve the artificial lift infrastructure planning problems. Section 3.2 proposes two new MSSP models for pharmaceutical clinical trial planning problems. A generalized knapsack-problem based decomposition algorithm (GKDA) is introduced in Chapter 4. Section 4.1 presents a general multistage stochastic programming formulation with both endogenous and exogenous uncertainties. Section 4.2 describes how a general MSSP transformed into a generalized knapsack-problem based decomposition algorithm. Section 4.3 and section 4.4 apply GKDA to both

synthesis of processes networks and artificial lift infrastructure planning problems. The results of case studies are also included in sections 4.3 and 4.4. The absolute expected values solution approach (AEEV) is presented in Chapter 5. Section 5.1 presents different primal bounds generation schemes in MSSP models. A systematic framework to generate the absolute expected value solution (AEEV) is introduced in Section 5.2. The AEEV is applied to instances from three applications in Section 5.3, and their results are presented in Section 5.4. In Chapter 6, two dual bounding approaches are presented to generate tight dual bounds for large-scale MSSPs efficiently. Section 6.1 presents a modified Lagrangian relaxation algorithm (mLR) for formulating a tight dual problem by utilizing the structure of MSSP under endogenous uncertainties and reducing the total number of Lagrange multipliers. Section 6.2 presents a relaxed knapsack-problem based decomposition (RKDA) to solve a relaxed version of the MSSP to optimality. Chapter 7 introduced a framework utilizing GKDA and Lagrangian decomposition to fully decompose the original MSSP into each individual scenario subproblems. In Chapter 8, twenty frameworks with different primal and dual bounding approaches are developed to solve the MILP of random forest, which has similar complicating constraints with two-stage stochastic programs. Overall conclusions and future directions are summarized in Chapter 9.

Note that the parts of contributions made in this dissertation have been published in three journal papers (Zeng and Cremaschi, 2017a, 2018a, 2019a) and six conference papers (Zeng and Cremaschi, 2017b, 2018b, 2018c, 2019b, 2019c, 2020).

CHAPTER 2

BACKGROUND

2.1 The stochastic programming framework

The optimization problems with uncertainties are typical in the chemical process industry. Stochastic programming is a framework for modeling optimization problems under uncertainty. The stochastic program relies on the information of the distribution of the uncertain parameters to find an optimal solution that is feasible for considered scenarios (Birge and Louveaux, 2011). Most optimization problems with discrete planning horizon under uncertainties, by nature, can be modeled as multi-period multistage stochastic programming (MSSP) formulations. MSSP is a scenario-based method that considers decisions and recourse actions in multiple stages. Scenarios in an MSSP model correspond to possible future states of the system characterized by the uncertain parameters, and they are generated using the allowable combinations of the uncertain parameter outcomes.

The uncertainty can be grouped into two broad categories, endogenous and exogenous, where the endogenous uncertainties can further be classified as Type I and Type II. Endogenous and exogenous uncertain parameters in MSSP are defined based on how and when they are realized. Exogenous uncertain parameters are realized at a known fixed time period in the planning horizon irrespective of the values of the decision variables (Birge and Louveaux 2011). For example, demand is generally considered independent of any capacity expansion decisions in process industries, and hence, it is regarded as an exogenous uncertain parameter (Goel and Grossmann, 2006; Tarhan and Grossmann 2008). In contrast, the values of the decision variables impact the realization times (Type II) and/or the distribution of endogenous uncertain parameters (Type I) (Jonsbraten et al., 1998). An example of Type II endogenous uncertain parameter is encountered

in the clinical trial planning problem (Colvin and Maravelias, 2008). Uncertainty in whether a drug successfully completes a clinical trial or not is only realized after the corresponding clinical trial is completed, and the decision to push a drug through a clinical trial does not impact its success or failure. On the other hand, in a facility protection problem, the likelihood of a facility failing to deliver goods or services after a disruptive event is an example of Type I endogenous uncertain parameter where the values of the decision variables impact its realized value, which depends on the level of resources allocated as protection to that facility (Medal et al., 2016).

Multistage stochastic programming models with endogenous uncertain parameters (in the presence or absence of exogenous uncertain parameters) consider the full horizon scenarios directly, where the decision variables are defined independently for each scenario. Then, a set of constraints called non-anticipativity constraints (NACs) are introduced to the model for preventing the anticipation of unrealized future outcomes when making decisions in individual scenarios. Unfortunately, these formulations model the uncertainty at different decision stages and take into account all possible scenarios and their impacts, which results in, typically, the exponential growth of the model with increases in the numbers of scenarios and stages. Therefore, these models quickly become computationally intractable for real-world size problems due to both space and time complexities. In practice, the MSSP formulations for most real-world size problems cannot be generated with currently available memories in most standard workstations, and their optimal solutions cannot be obtained within a reasonable time (Zeng and Cremaschi, 2018a).

The structure of multistage stochastic programming makes it suitable for utilizing solution approaches that decompose the problem using scenarios. Figure 2.1 illustrates this structure for an MSSP model under endogenous uncertainties. The model contains scenario specific constraints, which are essentially the deterministic models of each scenario (Figure 2.1). The variables in these

models are linked to each other with initial and conditional NACs (Figure 2.1). The objective function value of any feasible solution of the original MSSP model provides a primal bound for its optimal solution. The optimum objective function value of a relaxed MSSP model (e.g., which can be obtained by removing one or more of the constraints of the original MSSP) yields a dual bound for the optimum solution of the original MSSP. For large-scale problems with the exponential growth of space and computational complexities, it is necessary and important to find a feasible solution, and hence a primal bound, for the original problem with a relatively tight duality gap in a reasonable time (Boyd and Vandenberghe, 2004).

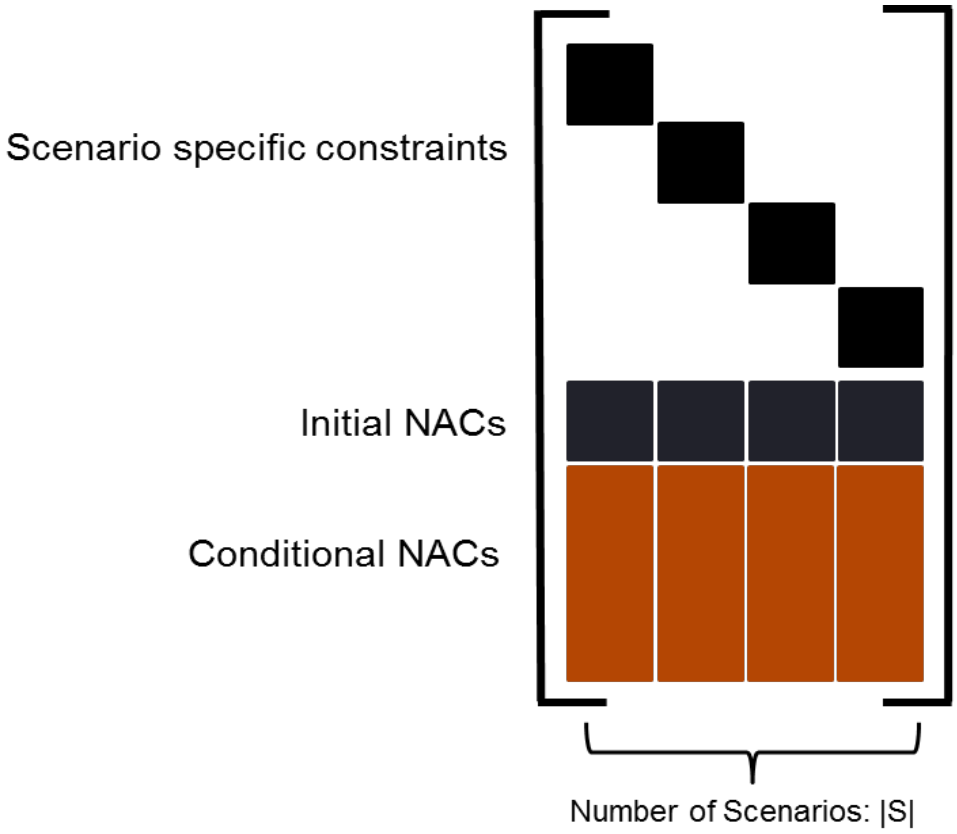

Figure 2.1: Structure of a multistage stochastic programming model under endogenous uncertainties (Zeng et al., 2018)

Most scenario-decomposition algorithms for solving MSSP models under endogenous uncertainties separate the scenario space into different scenario groups and enforce NACs in each

scenario group (Apap and Grossmann, 2017). The scenario-group sub-problems include subsets of scenarios and NACs, and thus are easier to solve than the original MSSP problem. In the extreme case, each individual scenario may be used to form a group, which essentially removes all NACs from the MSSP formulation. The solution of these sub-problems generates a dual bound for the original MSSP. For example, the solutions of individual scenario sub-problems yield the expected value of perfect information (EVPI) and is a valid dual bound (Birge and Louveaux, 2011). These dual-bounding approaches are then generally combined with heuristic primal-bounding approaches in an iterative fashion to solve the MSSP problems (Apap and Grossmann, 2017). A review of MSSP model libraries under Type II endogenous uncertainties is given in Section 2.2. A review of recent literature on the solution approaches for MSSPs under endogenous uncertainties is given in Section 2.3.

2.2 A library of models for multistage stochastic programs under type II endogenous uncertainty

This section describes the different MSSP models with endogenous uncertainties and their optimization problems in the literature. The literature contains 11 MSSP models with type II endogenous uncertainty. There are a great variety of aspects considered when developing and solving MSSP models under endogenous uncertainties. Table 2.1 summarizes the details of 11 MSSP models and provides an overview of the problems, authors, the first model publication years, sources of uncertainties, ways of scenario generations, the maximum scenarios of case solved, and the mathematical programming form.

Table 2.1: MSSP model attributes

Problems of MSSP models	Author	First Year	Uncertainties	Number of sources	Scenarios generation	Maximum scenarios	Problem formulation
Size Problem	Jonsbråten et al.	1998	Production cost	Single	Full (Cartesian)	64	MILP
Oil/gas-field development	Goel & Grossmann	2004	Field parameters	Two	Full (Cartesian)	729	MILP
Process network synthesis	Goel & Grossmann	2006	Yields of processes	Single	Full (Cartesian)	4	MILP
Open Pit Mine	Boland et al.	2008	Geology uncertainty	Single	Arbitrary (Sampled)	50	MILP
Clinical Trial	Colvin and Maravelias	2008	Outcomes of clinical trials	Single	Full (Cartesian)	1024	MILP
R&D portfolio	Solak et al.	2010	Resource requirements and Return level	Two	Unknown	Unknown	MILP
Offshore oilfield infrastructure planning	Gupta and Grossmann	2014	Field parameters	Multiple	Full (Cartesian)	8	MILP
Vehicle routing	Khaligh & MirHassani	2016	Demands of costumers	Single	Arbitrary (Sampled)	15	MINLP to MILP
Artificial Lift	Zeng and Cremaschi	2017	Gas flow rate change ratio	Single	Full sets (Cartesian)	8	MINLP to MILP
Smart grid technology	Giannelos et al.	2018	Level of consumer participation	Single	Full sets (Cartesian)	6	MILP
New technology investment planning	Brianna and Cremaschi	2018	Outcomes of technologies development, yields of processes, and learning curve rate	Multiple	Full sets (Cartesian product)	4608	MINLP to MILP

2.2.1 Size problems

Jonsbråten et al. (1998) was the first author who identified size problems as a class of problems under decision-dependent (endogenous) uncertainties. The framework of size problem is extended by Goel and Grossmann (2006) with disjunctive programming formulation. The size problem arises in the manufacturing design, where a variety of products under different sizes need to be

produced to meet the demands of an application. It is possible to substitute the larger size into a smaller size if the demand for this particular size cannot be met in specific time periods. The single source of endogenous uncertainty is the production cost, where the realization of these uncertainties can only be observed when that size is produced. The full set of scenarios is generated based on the Cartesian product of all outcomes of production costs for sizes. The size problems are formulated as mixed-integer linear programs (MLIP) and solved with up to 64 scenarios by Goel and Grossmann (2006).

2.2.2 Oil/gas-field development problem

Goel and Grossmann (2004) developed MSSP for an offshore gas field development problem. The problem has a set of gas reservoirs, known as fields, well platforms (WPs), production platforms (PPs), and connecting pipelines. The model determines the optimum investment decisions for installations of WPs and PPs, the capacities of these platforms, and pipelines' connections. The model incorporates two sources of endogenous uncertainties as sizes and initial deliverability of the fields, which can only be observed after the investment and operational decisions. The scenarios in the MSSP model is also full scenario set generated using the Cartesian product. The MSSP is formulated as a MILP and solved with up to 729 scenarios. The MSSP model has also been implemented and solved by Goel et al. (2006) with proposed branch-and-bound algorithms and further developed by Tarhan et al. (2009) for gradual realizations of uncertainties.

2.2.3 Process network synthesis

The MSSP model of process network synthesis is developed by Goel and Grossmann (2006). Process network synthesis determines the optimum installation, operation, and expansion

decisions to produce the final product and decide the purchase, sales, and inventory of intermediate and final products. The MSSP model of process network synthesis developed by Goel and Grossmann (2006) incorporates the processes' yield as endogenous uncertainties, where the installation, operation, and expansion decisions impact the observations of uncertainties. The scenarios are also generated by the full Cartesian product of outcomes of all processes under uncertain yields. The MSSP is formulated as a MILP and solved with two uncertain processes (4 scenarios). Tarhan and Grossmann (2008) extended the MSSP model of process network synthesis to incorporate the gradual realizations of uncertainties, where the yields of processes are resolved in two steps based on accumulated years of operation/expansion decisions.

Both oil/gas-field development and process network synthesis problems are implemented and solved with different proposed approaches. Both MSSP models are implemented and solved by Gupta and Grossmann (2011), Gupta and Grossmann (2014), Grossmann et al. (2017), and Apap and Grossmann (2018) with proposed Lagrangian decomposition (LD), sequential scenario decomposition heuristic (SSD), and several theoretical properties such as symmetry and scenario grouping to relax the non-anticipativity constraints are introduced to address the computational complexities of the MSSP models.

2.2.4 Open Pit Mine Production Scheduling (OPMPSP)

Boland et al. (2008) developed the MSSP model for the open-pit mine production scheduling problem (OPMPSP) under uncertain geology. The MSSP model determines the optimal scheduling of the extraction of a mineral deposit broken into several blocks (partitioned into series aggregates) to maximize the net present value (NPV). The endogenous uncertainty is the grade of aggregate, where the observation of uncertainties is controlled by the binary decisions indicating whether extraction began. Instead of generating scenarios using the Cartesian product, where the attribute

values of aggregate are usually partially linked, the scenarios are sampled from a continuous, not discrete, distribution. The MSSP is also formulated as a MILP and solved up to 50 scenarios. Boland et al. (2008) also proposed a lazy-constraints approach for reducing a large number of non-anticipativity constraints.

2.2.5 The clinical trial planning problem

Colvin and Maravelias (2008, 2009, and 2010) developed MSSP models for pharmaceutical clinical trial planning problems with clinical trial outcomes as endogenous uncertainties. The model determined the optimum selection of drugs to push through the clinical trial phases and their start time. A set of candidate drugs must pass three clinical trial phases before Food and Drug Administration (FDA) approval. The endogenous uncertainty is the outcomes of clinical trials, and each drug naturally has four discrete outcomes: {PI-Fail, PII-Fail, PIII-Fail, PIII-Pass}. The uncertainties in clinical trial outcomes are gradually realized and controlled by decisions that determine the start of clinical trials. The scenarios are generated by the Cartesian product of outcomes of clinical trials of all drugs. The MSSPs are formulated as MILPs and solved up to 4,608 scenarios by Colvin and Maravelias (2008, 2009, and 2010).

2.2.6 R&D project portfolios

An MSSP model was developed by Solak et al. (2010) for an R&D portfolio planning problem with return level and resource requirements as endogenous uncertainties. The MSSP model determines the optimal allocation of resources for a set of projects. The return level and resource requirements are regarded as two endogenous uncertainties and are gradually realized based on investment level. The uncertainty dependency between projects is considered. The MSSP is formulated as a MILP and solved with up to ten projects. However, the scenario generation and

information of instances are not available. The MSSP model is solved with a proposed Sample Average Approximation (SAA).

2.2.7 Offshore oilfield infrastructure planning

Gupta and Grossmann (2014) developed an MSSP model for offshore oil and gas field infrastructure planning problem under multiple uncertain sources. The MSSP model determines the optimal installation, operation, and drilling decisions of platforms, fields, and wells and their connections. There are multiple endogenous uncertainty sources in fields considered: filed size, oil deliverability, water-oil ratio, and gas-oil ratio. The uncertainties in fields will be realized if a certain number of wells have been drilled. To reduce the total number of scenarios, perfect correlations among the uncertain parameters are assumed, where only a subset of the possible scenarios is generated using the Cartesian product. The original model is formulated as a MINLP and converted to a MILP formulation. The model is solved up to eight scenarios.

2.2.8 Vehicle routing problem

Khaligh and MirHassani (2016) studied an MSSP formulation of a vehicle routing problem with endogenous demand uncertainties realized after the visit decisions. The MSSP model determines the optimal planning of visit and loading. Each customer must be satisfied in exactly one visit, and enough load must be ensured to satisfy customers. In this MSSP model, the customers' demands are regarded as endogenous uncertainties and are revealed as soon as the vehicle arrives at customers. The scenarios are generated randomly with different demands. Among all models, the vehicle routing problem and open-pit mining production scheduling problem are only two of the models that create arbitrary scenario sets instead of using the Cartesian product. The original model is formulated as a MILP and is solved up to 15 scenarios.

2.2.9 Artificial lift infrastructure planning problem (ALIP)

Zeng and Cremaschi (2017a, 2018b) developed MSSP models of artificial lift infrastructure planning problems for horizontal shale producing wells. The MSSP model determines the optimal selection of artificial lift methods (ALM) and its installation and uninstallation time. After installing ALMs, the new production rate is unknown, and the ratio of new product rate to previous production rate is defined as the gas flow rate change ratio, which is the source of the endogenous uncertainty in the model. The parameter values of the gas flow rate change ratio are revealed after the installation decisions of ALMs are made. The scenarios are generated using the Cartesian product of outcomes of all uncertain ALMs' gas flow rate change ratios. The original MSSPs are MINLPs and linearized as MILPs solved with eight scenarios up to a 16-month planning horizon.

2.2.10 Demand-Side Response Planning

Giannelos et al. (2018) proposed an MSSP model of the Demand-Side Response (DSR) planning problem under uncertain consumer participation levels. The MSSP model generates optimal investment decisions in conventional line reinforcements and DSR to maximize the expected net benefit with operational and investment costs. The consumer participation level is revealed depending on investment decisions. The scenarios are generated using the Cartesian product of outcomes of consumer participation levels. The MSSP model is formulated as a MILP and solved up to 6 scenarios. A modified Benders decomposition is also proposed by Giannelos et al. (2018) to address the computational complexity of the problem.

2.2.11 New technology investment planning

Christian and Cremaschi (2018) developed an MSSP model for the new technology investment planning (NTIP) problem. The objective is to determine the technology investment plan that yields

the expected minimum total cost. Demand at each time period should be met either via production or via purchasing at market price. The MSSP model contains multiple sources of endogenous and exogenous uncertainties. There are four endogenous uncertainties associated with each new technology: (1) project abandonment, (2) process yield, (3) learning-by-searching elasticity, and (4) learning-by-doing elasticity. The demand for products is also considered as an exogenous uncertainty source. The full scenario sets are generated using the Cartesian product. The original model is formulated as a MINLP and approximated by a MILP formulation. The model is solved up to 4608 scenarios.

2.3 Solution Approaches of Multistage Stochastic Program with Endogenous Uncertainty

In general, recent literature focuses on developing approximation and decomposition approaches to address the computational complexity of large-scale MSSPs. Goel and Grossmann (2004) proposed an approximation approach, which searched a restricted feasible region of the MSSP to find a "good" solution to the offshore gas field infrastructure planning problem. Their results revealed that "good" solutions obtained by their approximation method yielded significantly improved objective function values when compared to the optimum solution of the deterministic problem. A Lagrangean duality based branch-and-bound algorithm, which is guaranteed to give the optimal solution, was introduced to solve similar planning problems (Goel et al., 2006). The results suggested that the branch-and-bound algorithm achieved significantly better solutions and tighter optimality gaps than the heuristic presented in (Goel et al., 2006). The process synthesis problem with decision-dependent uncertainty in process yields, which are revealed gradually as investments in processes occurred, was modeled as an MSSP model by Tarhan and Grossmann (2008). A duality based branch-and-bound approach was able to solve instances of this problem with 16 scenarios to within a 3% optimality gap. Tarhan et al. (2009)

incorporated nonlinear reservoir models and the gradual realization of uncertainty to the offshore gas field infrastructure planning problem (Goel and Grossmann, 2004) and solved the resulting MSSP models using a duality-based branch-and-bound algorithm. The solutions found by the algorithm were up to 22% better than the ones obtained using an expected value approach, however the authors noted that the solution times were "rather long".

Mercier and Hentenryck (2008) proposed a multi-step anticipatory algorithm for solving pharmaceutical R&D pipeline management problems. The algorithm uses a sample average approximation approach to generate a Markov Decision Process (MDP), solves the resulting MDP, and returns the greedy solution. The algorithm was tested using 12 instances of the R&D pipeline management problem, and yielded solutions that were 10% better than the dynamic programming equivalents for all instances. The authors concluded that the algorithm was, nevertheless, computationally expensive. Colvin and Maravelias (2010) explored a rolling-horizon approximation approach to solve large instances of the pharmaceutical clinical trial planning problem. The approach constructs a relaxed MSSP model by only retaining NACs for the stages of the detailed time block, and implements the solution of the relaxed MSSP model to these stages. Then, the uncertainties associated with the implemented decisions are realized. The approach continues to construct and solve relaxed MSSP models rolling the detailed time block at each iteration until the end of the planning horizon is reached. The authors were able to successfully solve cases with more than 1000 scenarios. Another approach, a branch-and-cut algorithm5, initially constructs and solves a relaxed MSSP model that includes a percentage of the NACs, and then iteratively adds the NACs that are violated. The authors concluded that the algorithm reduces the number of NACs that should be included in the problem formulation significantly and that this

method would be advantageous for any stochastic programming formulation where the majority of constraints are NACs.

Solak et al. (2010) proposed a sample average approximation (SAA) algorithm for solving R&D portfolio optimization problems. The algorithm solved SAA problems to generate candidate solutions and evaluated the quality of the first-stage decision of these solutions using a larger sample set of scenarios. Computational studies included two technology portfolio examples where five and ten projects were considered, and the solutions were within 3.4%-8.3% of the optimal solution.

Vayanos et al. (2011) introduced a novel approach that uses decision rules for approximating MSSP problems under endogenous uncertainties. The approach employs piecewise constant and piecewise linear functions of uncertainties and can handle uncertain parameters with both discrete and continuous distributions. The authors applied the approach to solving the oil infrastructure planning problem, and the approach yielded improved primal bounds with limited time budgets.

A solution strategy for MSSP models with endogenous uncertainty whose deterministic equivalent is a non-convex mixed-integer nonlinear programming (MINLP) model was presented in Tarhan et al. (2013). The strategy utilized global optimization and outer-approximation, and it was used to solve two planning problems, a modified version of the process synthesis problem (Goel and Grossmann, 2007; Tarhan and Grossmann, 2008) and an offshore oilfield planning problem (Tarhan et al., 2009). Gupta and Grossmann (2014) explored improved Lagrangean decomposition frameworks for both process synthesis and oilfield planning problems, which obtained solutions with less than 1% gap from the optimal solution and with fewer iterations. For MSSP models with both endogenous and exogenous uncertainties, Apap and Grossmann (2017) proposed a sequential scenario decomposition (SSD) approach. The approach selects one scenario

from the exogenous uncertainty scenario group at the initial time period, solves a series of sub-problems with endogenous uncertainties to determine binary investment decisions, fixes the binary investment decisions to satisfy the first-time period endogenous and exogenous NACs, and solves the resulting model to obtain a feasible solution. The authors used the SSD approach in conjunction with the Lagrangian relaxation strategy (dual bound) to solve the full-space problem with endogenous and exogenous uncertainties. To test their algorithm, the authors considered the process synthesis problem(Goel and Grossmann, 2007; Tarhan and Grossmann, 2008) and the offshore oilfield planning problem(Tarhan et al., 2009), and they concluded that the SSD approach found high-quality feasible solutions several orders of magnitude quicker than solving the full space model.

CHAPTER 3

TWO PLANNING PROBLEMS OF MULTISTAGE STOCHASTIC PROGRAMS

3.1 Artificial Lift Infrastructure Planning Problem (ALIP)

Shale gas has become a significant source of natural gas in the United States. In 2010, shale gas provided over 20% of U.S. natural gas production compared to its 1% share in 2000. The U.S. Energy Information Administration predicts that 46% of the U.S.' natural gas will be obtained from shale gas in 2035 (Stevens et al., 2012). Although horizontal drilling has been known since the 1950s, recent advancements in equipment combined with hydraulic fracturing techniques allowed the production of natural gas from previously inaccessible sources such as shale formations. These developments considerably increased the quantity of shale gas available for production (Robbins, 2013).

The typical lifetime of a horizontal gas well is shown in Figure 3.1. As can be seen from Figure 3.1, the production from a horizontal well can be categorized into three stages (Bondurant et al. 2007). During the first stage, the well produces naturally. The produced fluids are mostly gas with fine liquid droplets dispersed in it. As the production continues, the liquid flow rate decreases faster compared to the decrease in the gas flow rate. When the gas flow rate drops below a critical value, i.e., below the loading flowrate, liquid accumulates at the bottom of the well, and gas production stops. During this second stage of production, the well should be deliquified using one or more of the ALMs to maintain production. During both first and second stages, the pressure of the reservoir is higher than the wellhead pressure. At the last stage of production, the reservoir pressure decreases below the wellhead pressure, and external energy and mechanical assistance are required to remove the fluids and extend the well lifetime. Because large amounts of fluid are injected to the shale formation during the fracturing process, shale gas wells often require

deliquification to unload the well relatively quickly, generally within their first or second year of production.

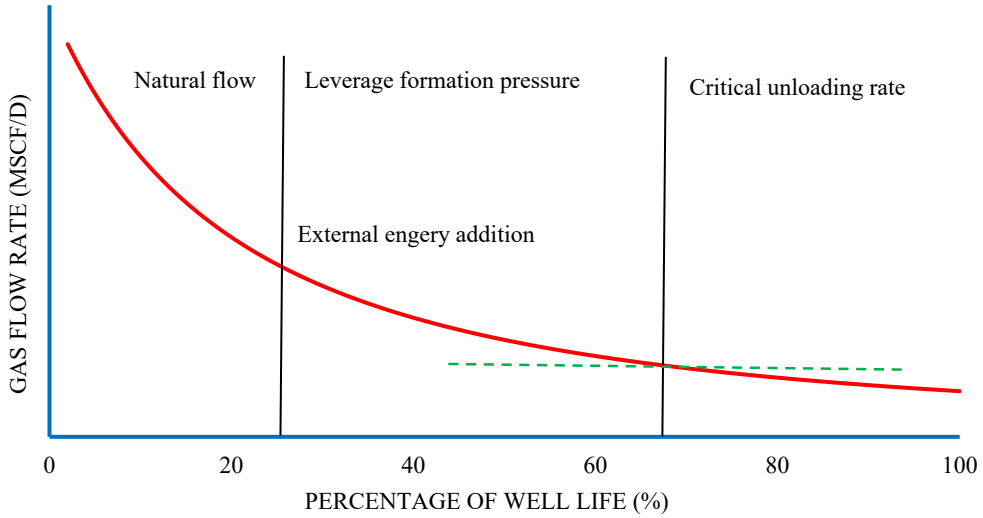

Figure 3.1: Typical lifetime of a horizontal well

Artificial lift methods can be divided into two types: Passive and active. Passive systems, such as velocity strings, plunger lift and foam lift, mainly naturally carry and remove liquid from the wellbore because sufficient energy remains in the reservoir. The active systems, such as sucker rod pump, wellhead compressors and gas lift, add energy to the system, and are generally implemented once the reservoir pressure drops below the wellhead pressure (Oyewole and Lea, 2008).

The correct selection of an ALM is important to the long-term profitability of most producing wells. A poor choice will result in production reduction, and is changing the installed type of ALM is a significant investment significantly. As the production continues, the characteristics of wells such as liquid flow rate, gas flow rate and gas liquid ratio (GLR) change along with the plan life. Thus, a continuous review of the performance of the ALM is required to change the ALM at a proper time. The typical lifetime of a well is around 20-25 years, and hence, multiple ALMs may

be installed in horizontal wells for achieving desirable well production performance after the first stage. Therefore, it is imperative for oil companies to manage their ALM infrastructure planning more effectively to avoid the potential cost of failing methods. To effectively plan the infrastructure of ALM, new systematic optimization models are necessary.

There are many related studies in the Chemical Engineering community dealing with shale gas production. Large-scale production of shale gas has been considered as a major issue in the U.S. energy industry, and optimization of the shale gas process has become one of the most popular research areas (Gao and You, 2017). Cafaro and Grossmann (2016) presented a continuous-time nonlinear programming model and a multi-period mixed-integer programming model for refracturing planning problem in horizontal shale gas wells to determine the refractured wells, the frequency of refracturing and its time schedule. Drouven and Grossmann (2016) developed a planning model for multiple gas wells to decide when, where and how many wells to drill and whether some of the selected wells should be shut-in based on the long-term production and gas quality forecasting. The model also designed the infrastructure such as allocating drilling rigs, completion crews, amount of freshwater, gathering pipelines and compression power. There are several guidelines and expert systems designed for ALM selection in the production of shale gas processes. They mainly use parameters specifying well characteristics, operation envelopes of the ALMs, and economic criteria to recommend suitable ALMs. In an early example, Clegg (1988) recommended using three economic factors, income, operating cost, and capital cost, generally in that order as a basis for ALMs selection for vertical wells. Clegg (1988) noted that the chosen ALM should produce the reserves in a timely fashion with the minimum operating cost. His paper analyzed five different ALMs in terms of production flow rates and operating costs. The first expert system for ALM selection was introduced by Heinze et al. (1989). The expert system considered

a limited set of well characteristics and recommended one of the four ALMs based on these characteristics and ALM design requirements. Clegg et al. (1993) developed a matrix that compared 31 well-related attributes against the operation envelopes of eight ALMs to eliminate the inappropriate ones. With upgrades and development in ALMs, Heinze et al. (1996) have expended the considerations for selection of pumping method and developed the decision tree of selecting ALMs. Ramirez et al. (2000) developed a multi-stage expert system using more detailed well information and ALM operation envelopes. The initial screening using two parameters eliminated the inappropriate ALMs: maximum lift depth and production. Then, the remaining methods were compared in terms of their specific power requirements (number of watts required to lift one barrel of oil per day at the surface condition). The w VS Qg curves has been presented to compare ALMs power consumption related to gas flow rate. Finally, the economic analysis was performed to compare the ALMs in terms of their net present values (NPV). Oyewole and Lea (2008) have evaluated the lowest bottom hole pressure for various ALMs for the gas well. Weatherford (2013) presents the operating limits for the most common ALS used in gas well deliquification. The latest expert system was developed by Valbuena (2015). It first eliminates ALMs using developed guidelines based on the well characteristics, field experience and the operation envelopes of the ALMs. Then, the remaining ALMs are ranked by twenty-four attributes, which are weighted on a scale from 1 to 10 defined by the operators. Finally, an economic evaluation is performed to make the final recommendations. All methods and expert systems, in general, consider the well characteristics at a single starting point in time and assess the suitability of the ALMs for the well at that point in time. Although some of the well characteristics change over the lifetime of the well due to production, they assume that the selected ALM from the starting point stays in place. As time goes, production from wells may be halted under initial decisions.

Therefore, we need to develop a modeling framework for selecting optimal ALMs that can incorporate changing well characteristics, consider the interaction of the ALMs with the well and each other, and produce an artificial lift infrastructure plan for the lifespan of the well.

3.1.1 Deterministic models

3.1.1.1 Problem Definition

The problem addressed in this paper can be stated as follows. We assume that a candidate horizontal shale gas well has been identified for requiring artificial lifts. We are given a set of potential ALMs $i \in I$ for a given horizontal gas well. In the given potential ALMs, the characteristics of a candidate horizontal shale-gas well such as its geometry, operating conditions and production performance to date are given. The costs for installing ALMs are expressed by functions with charges along the month, which include the ALMs' equipment, installation, operating costs and the economical parameters such as gas price, natural gas liquid price. The economic performance is assessed using net present value (NPV) for the MINLP and MILP. The planning horizon is discretized into $|T|$ equal times (in this case, months).

The problem is to determine: (a) which artificial lift method should be installed, (b) when exactly the artificial lift method should be installed and uninstalled. The objective is to maximize the net present value (NPV).

3.1.1.1.1 Assumptions

After an ALM installed, a peak in the production rate was observed. The behavior of the production rate can be characterized based on the value of the decline exponent constant b. D is a parameter called the nominal decline rate, which is defined as the fractional change in rate per unit

time. Generally, this secondary production peak presents a fraction of the previous peak as shown in Figure 3.2,

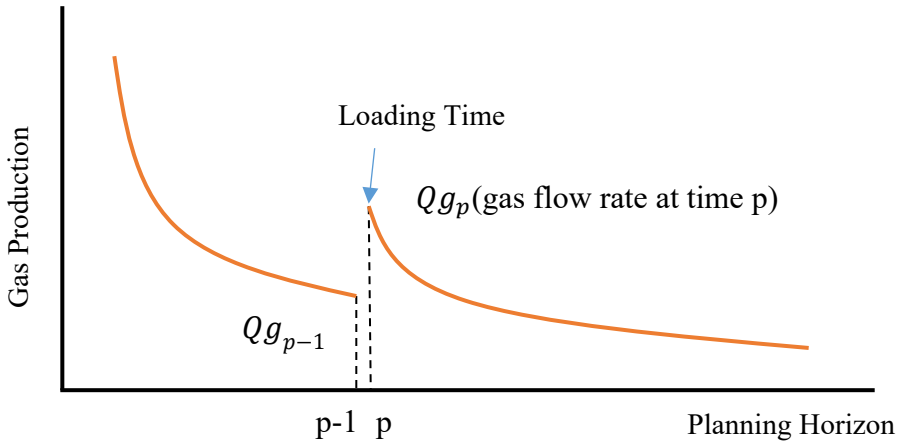

Figure 3.2: Productivity curves of a gas well with artificial lift method installed

The flowrate change ratio of the new production peak after ALM is installed to the previous production rate is unpredictable. The models rely on the assumption that the decline curve functions predict the production rate well, and the flowrate change ratio is fully known. In the current model, we assume that the ALMs do not affect the decline curve parameters, b and D. We take the different average equipment costs of each month for different ALMs disregarding their limited lifetime.

3.1.1.2 MINLP formulations

This section lists the two different MINLP formulations used to model the artificial lift infrastructure planning problem. The total nomenclature can be found in Appendix A1.

3.1.1.2.1 Separate index variables (F0)

In the first MINLP formulation, we use separate index binary variables, $w_{i,p}$ and $z_{i,t}$, to model the problems. The major decision binary variable of the formulation is $w_{i,p}$ which is equal to one

if artificial lift method *i* is installed at the period *p*. We introduce the binary variable $z_{i,t}$ to keep track of whether the artificial lift method is *i* is uninstalled at the period *t*. The logic constraints are defined below:

$$\sum_{i} w_{i,p} \leq 1 \qquad \forall i \in I \;\; \forall p \in T \tag{3.1}$$

$$\sum_{p} w_{i,p} \leq 1 \qquad \forall i \in I \;\; \forall p \in T \tag{3.2}$$

$$\sum_{i} z_{i,t} \leq 1 \qquad \forall i \in I \;\; \forall p \in T \tag{3.3}$$

$$\sum_{t} z_{i,t} \leq 1 \qquad \forall i \in I \;\; \forall p \in T \tag{3.4}$$

$$\sum_{p} w_{i,p} - \sum_{t} z_{i,t} = 0 \qquad \forall i \in I \tag{3.5}$$

$$\sum_{i,t,p} z_{i,t} w_{i,p} = 0 \qquad \forall i \in I, t \leq p \tag{3.6}$$

$$z_{i,t} w_{i,p} + z_{i,k} w_{j,l} \leq 1 \; \forall j \neq i, p+1 \leq l \leq t, t+1 \leq k \leq T \tag{3.7}$$

Equations (3.1) - (3.4) state that each method is restricted to installed and uninstalled at most once, and the number of ALMs, which can be installed and uninstalled at any given month, are limited to a maximum of one. The equations (3.5) and (3.6) ensure that each ALM is uninstalled only once, and the removal happens after its installation time. The last equation (3.7) prevents the overlap of two ALMs.

We apply the modified hyperbolic decline curve function to predict the production flow rates of a horizontal shale-gas well (Fetkovich et al., 1996). These functions are given in Eqns. (3.8) - (3.10) for predicting gas (*Qg*), oil (*Qo*), and natural gas liquids (*Qng*) flowrates, respectively, after the well is loaded,

$$Qg_r = \sum_{i,t,p} z_{i,t} w_{i,p}\, Qg_{p-1} Qrec_i \left(1 + bD(t-r+1)\right)^{-\frac{1}{b}} \tag{3.8}$$

$$Qo_r = \sum_{i,t,p} z_{i,t} w_{i,p}\, Qo_{p-1} Qrec_i \left(1 + bD(t-r+1)\right)^{-\frac{1}{b}} \tag{3.9}$$

$$Qn_r = \sum_{i,t,p} z_{i,t} w_{i,p}\, Qn_{p-1} Qrec_i \left(1 + bD(t-r+1)\right)^{-\frac{1}{b}} \tag{3.10}$$

We assume that the secondary peak is described as: $Qrec_i Qg_{p-1}$. The gas production flow rate for month p, Qg_p, depends on the gas flowrate at month (p-1) (the month prior to the installation of ALM i), Qg_{p-1}, and the flowrate change ratio of the installed method i, $Qrec_i$.

Each ALM has defined design and operational limitations such as the Artificial Lift R&D Council guidelines, previous expert experience (Wittfeld, 2005), and limitations stated in typical attribute tables (e.g., Weatherford, 2013). These limitations are incorporated as technical limitation constraints in the model ((3.11), (3.12), and (3.13)):

$$f_{i,p,r}\left(LFR_r, w_{i,p}\right) \leq 0 \qquad \forall p, r \in T, i \in I \tag{3.11}$$

$$f_{i,t,r}\left(LFR_r, z_{i,t}\right) \leq 0 \qquad \forall \mathrm{t}, r \in T, i \in I \tag{3.12}$$

$$g_{i,p}\left(w_{i,p}\right) \leq 0 \qquad \forall p \in T, i \in I \tag{3.13}$$

The most common limitation is the liquid flow rate ranges under which an ALM can operate, as the liquid flow rate decreases through the whole lifetime. We use the product of $w_{\mathrm{i},p}$ and LFR_r to describe the upper bound of the liquid flow rate each method can operate. We apply $z_{i,t}$ to the lower bound of the liquid flow rate of each method. For example, the Weatherford (2013) attribute table sets the maximum flow rate a plunger lift can operate as 200 barrels per day (BPD). Therefore, one of the constraints in Eq. (3.11) set is Eq. (3.14), where i = 1 for plunger lift,

$$w_{1,p} LFR_r \leq 200 \qquad \forall t, p, r \in T, t \geq r \geq p \tag{3.14}$$

With respect to the low volume lift capabilities of the Jet Pump method, Clegg *et al.* (1993) stated that the minimum flow rate should be around 200 BPD, which stated as Eq. (3.15) where $i = 9$,

$$z_{9,t} 200 \leq LFR_r \qquad \forall t, p, r \in T, t \geq r \geq p \tag{3.15}$$

The remainder of the technical limitation constraints can be expressed using the binary decision variable $w_{i,p}$ or $z_{i,t}$ and relevant well or ALM parameters as stated in Eq. (3.13). For example, Donald et al. (2014) recommend that candidate wells for well head compression should have less than 1100 psi (80 bar) closed-in tubing head pressure ($CITHP$) based on onshore gas wells applications in the Netherlands. The following constraint, Eqns. (3.16), formulate these statements for wellhead compression where $i = 3$,

$$w_{3,p} CITHP \leq 1100 \qquad \forall p \in T \tag{3.16}$$

The objective is to maximize the NPV, so we consider the cost of equipment and installation and the profit from production over the life of the well. The total gross income GI_r of month r is calculated in (3.17),

$$GI_r = \left(Rev_r (1 - RT)(1 - LT) - \sum_{i,t \geq r, p \leq r} \left(Cm_i z_{i,t} w_{i,p} \right) \right) WI \quad \forall r \in T \tag{3.17}$$

Equation (3.17) estimates the gross income in each month r by deducting local taxes, royalties, and operating and maintenance costs. The binary variables $y_{i,t,p}$ become one if ALM i is installed at month p and uninstalled at month t, respectively. The capital cost and the revenue from gas, oil and natural gas liquids sales are calculated in Eq. (3.18) and (3.19),

$$Rev_r = P_g Qg_r + P_o Qo_r + P_n Qn_r \quad \forall r \in T \tag{3.18}$$

$$CC = \sum_{i,t,p} \left(Co_i z_{i,t} w_{i,p}\right) \left(\frac{1}{(1+MARR)^p}\right) \quad \forall t, p \in T, i \in I \tag{3.19}$$

The taxable income TI_r comes from the difference of gross income, GI_r, and depreciation, Dep_r, calculating for every month in Eq.(3.20) - (3.23),

$$TI_r = GI_r - Dep_r \quad \forall r \in T \tag{3.20}$$

$$TI_r \leq x_r M \quad \forall r \in T \tag{3.21}$$

$$TI_r > (x_r - 1)M \quad \forall r \in T \tag{3.22}$$

$$Dep_r = \sum_{i,t \geq r, p \geq r-n+1}^{p \leq r} \left(\frac{Ce_i}{n} z_{i,t} w_{i,p}\right) \quad \forall t, p \in T, i \in I \tag{3.23}$$

The binary variable x_r, which becomes one if taxable income is negative at month r, avoids deducting federal taxes for months with negative taxable income. The straight-line n-year depreciation model used in the NPV calculations is given in Eq.(3.23).

Based on the above definitions, the mixed-integer nonlinear programming model (MINLP) for the artificial lift infrastructure planning problems is as follows:

$$\begin{aligned} max: NPV = \sum_r \left\{ (GI_r - TI_r \cdot x_r \cdot FT) \left(\frac{1}{(1+MARR)^r}\right) \right\} \\ - CC(1-FT) \quad \forall r \in T \end{aligned} \tag{3.24}$$

$$s.t. (3.1) - (3.13), (3.17) - (3.23) \tag{3.25}$$

3.1.1.2.2 Integrated time variables (F1)

Instead of describing the installation month and uninstallation separately, the integrated time variables formulation based on binary variable $y_{i,t,p}$ which is introduced to present the whole

scheduling period including the starting and ending time of certain methods. $y_{i,t,p}$ is equal to 1 if artificial lift method *i* installed at the period p and uninstallation happens at month t. The $y_{i,t,p}$ has relationship relate to $w_{i,p}$ and $z_{i,t}$,

$$y_{i,t,p} = w_{i,p} \cdot z_{i,t} \qquad \forall i \in I \ \ \forall p \in T \tag{3.26}$$

The sequential constraints are defined as below:

$$\sum_{t,p} y_{i,t,p} \leq 1 \qquad \forall t, p, r \in T, i \in I \tag{3.27}$$

$$\sum_{i,t} y_{i,t,p} \leq 1 \qquad \forall t, p, r \in T, i \in I \tag{3.28}$$

$$\sum_{i,p} y_{i,t,p} \leq 1 \qquad \forall t, p, r \in T, i \in I \tag{3.29}$$

$$y_{i,t,p} = 0 \qquad \forall t, p \in T, i \in I, t \leq p \tag{3.30}$$

$$y_{i,t,p} + y_{j,k,l} \leq 1 \qquad \forall j \neq i, p+1 \leq l \leq t, t+1 \leq k \leq T \tag{3.31}$$

Equations (3.27)-(3.31) define the ALM installation and removal plan constraints. Equation (3.27) states that each method can be installed at most once. Equations (3.28) and (3.29) limit the number of ALMs that can be installed at any given month to a maximum of one. Equation (3.30) ensures that an ALM is installed before its removal. The last constraint, Eq. (3.31), prevents the overlap of two ALMs.

Instead of relating the $w_{i,p}$ and the $z_{i,t}$ to the upper and lower bound of liquid flow rate separately with two types of constraints, envelop constraints of the MINLP formulation with integrated time variables are given below. These limitations are incorporated as technical limitation constraints in the model (Eqs. (3.32) and (3.33)):

$$f_{i,t,p,r}\left(LFR_r, y_{i,t,p}\right) \le 0 \qquad \forall \mathrm{t}, p, r \in T, i \in I \tag{3.32}$$

$$g_{i,t,p}\left(y_{i,t,p}\right) \le 0 \qquad \forall \mathrm{t}, p \in T, i \in I \tag{3.33}$$

The details of constraints (3.32) and (3.33) are given in Appendix A2. With the new decision variable approach, we can formulate the problem as the following mixed-integer nonlinear program (F1):

$$max: NPV = \sum\nolimits_r \left\{ (GI_r - TI_r \cdot x_r \cdot FT)\left(\frac{1}{(1+MARR)^r}\right)\right\} - CC(1-FT) \quad \forall r \in T \tag{3.34}$$

$$GI_r = \left(Rev_r(1-RT)(1-LT) - \sum_{i,t\ge r, p\le r} \left(Cm_i y_{i,t,p}\right) \right) WI \quad \forall r \in T \tag{3.35}$$

$$CC = \sum_{i,t,p} \left(Co_i y_{i,t,p}\right)\left(\frac{1}{(1+MARR)^p}\right) \quad \forall t, p \in T, i \in I \tag{3.36}$$

$$Dep_r = \sum_{i,t\ge r, p\ge r-n+1}^{p\le r} \left(\frac{Ce_i}{n} y_{i,t,p}\right) \quad \forall t, p \in T, i \in I \tag{3.37}$$

$$Qg_r = \sum_{i,t,p} y_{i,t,p}\, Qg_{p-1} Qrec_i \left(1 + bD(t-r+1)\right)^{-\frac{1}{b}} \tag{3.38}$$

$$Qo_r = \sum_{i,t,p} y_{i,t,p}\, Qo_{p-1} Qrec_i \left(1 + bD(t-r+1)\right)^{-\frac{1}{b}} \tag{3.39}$$

$$Qn_r = \sum_{i,t,p} y_{i,t,p}\, Qn_{p-1} Qrec_i \left(1 + bD(t-r+1)\right)^{-\frac{1}{b}} \tag{3.40}$$

$$\text{s.t. } (3.18), (3.20) - (3.22), (3.27) - (3.33) \tag{3.41}$$

3.1.1.3 Reformulated MILP formulation of ALIP

The MINLP model (F0 and F1) has nonconvex terms in the decline curve constraints (Eqns. (3.8)-(3.10) and (3.38)-(3.40)), in the objective function (3.24), and in technical limitation constraints ((3.11), (3.12), and (3.32)). They have a special structure and are the multiplications of a binary variable with multiple continuous variables (such as ($z_{i,t}\, w_{i,p}Qg_{p\text{-}1}$), ($w_{i,p}Qg_r$), ($w_{i,p}LFR_r$), and ($TI_r x_r$)) and two binary variables (such as $z_{i,t}w_{i,p}$). We use exact linearization for replacing these nonlinear terms with linear equivalents (Oral et al., 1991)

3.1.1.3.1 Separate index variables MILP (LF0)

The product of $w_{i,p}$ and LFR_r in the constraints can be linearized as follows,

$$LFRw_{i,p,r} + LFRw1_{i,p,r} = LFR_r \tag{3.42}$$

$$LFRw_{i,p,r} \leq w_{i,p} \cdot M_i \tag{3.43}$$

$$LFRy1_{i,t,p,r} \leq (1 - w_{i,p}) \cdot M_i \tag{3.44}$$

$$LFRw_{i,p,r} \geq 0, \qquad LFRw1_{i,p,r} \geq 0 \tag{3.45}$$

where $LFRw_{i,t,p,r}$ and $LFRw1_{i,t,p,r}$ are two new continuous variables, and $LFRw_{i,t,p,r}$ is equivalent to $(w_{i,p} \cdot LFR_r)$. Because each artificial lift method has different liquid flow rate limitations, the upper bound value M_i is different of each artificial lift method *i*.

To linearize the term $z_{i,t}w_{i,p}$, we introduce a new continuous non-negative variable $zw_{i,t,p}$ in the decline curve constraints. Then the product of $z_{i,t}$ and $w_{i,p}$ can be replaced by the term $zw_{i,t.p}$,

$$zw_{i,t.p} \leq w_{i,p} \tag{3.46}$$

$$zw_{i,t.p} \leq z_{i,t} \tag{3.47}$$

$$zw_{i,t.p} \geq w_{i,p} + z_{i,t} - 1 \tag{3.48}$$

$$zw_{i,t.p} \geq 0 \tag{3.49}$$

To remove the nonlinear term in the decline curve constraints and $w_{i,p}Qg_r$ in constraints, we introduce five new continuous variables $wzQg_{i,t,p,r}$, $wzQg1_{i,t,p,r}$, $wQg_{i,p,r}$, $wQg1_{i,p,r}$, and $Qge_{i,p,r}$, and replace Eq. (3.8) with the constraints given in Eqns. (3.50)-(3.58).

$$Qg_r = \sum_{i,t,p} zwQg_{i,t,p,r} \tag{3.50}$$

$$Qg_{i,p,r} = Qg_{p-1}Qrec_i\left(1 + bD(r - p + 1)\right)^{-\frac{1}{b}} \tag{3.51}$$

$$wQg_{i,p,r} + wQg1_{i,p,r} = Qge_{i,p,r} \tag{3.52}$$

$$wQg_{i,p,r} \leq w_{i,p}Qg_0 \tag{3.53}$$

$$wQg1_{i,p,r} \leq (1 - w_{i,p})Qg_0 \tag{3.54}$$

$$wzQg_{i,t,p,r} + wzQg1_{i,t,p,r} = wQg_{i,p,t} \tag{3.55}$$

$$wzQg_{i,t,p,r} \leq z_{i,t}Qg_0 \tag{3.56}$$

$$wzQg1_{i,t,p,r} \leq (1 - z_{i,t})Qg_0 \tag{3.57}$$

$$wQg_{i,t,p,r}, \quad wQg1_{i,t,p,r}, \quad wzQg_{i,t,p,r}, \quad wzQg1_{i,t,p,r} \geq 0 \tag{3.58}$$

We linearize the term $w_{i,p}Qg_{p-1}$ with the introduction of new continuous variable $wQg_{i,p,r}$ and then replace the product of $wQg_{i,p,r}$ and $z_{i,t}$ with term $wzQg_{i,t,p,r}$. Equation (3.50) calculates the gas production flowrate at month r. The variable $Qg_{i,p,r}$ calculated in Eq. (3.51) is the gas flow rate at month r with ALM i that is installed on month p. Equation (3.53) ensures that the variable $wQg_{i,t,p,r}$ becomes zero if ALM i is not operational at month r. Equations (3.55) and (3.56) ensure that the

gas production flowrate calculated before the artificial lift method has been uninstalled. The parameter Qg_0 is set equal to the maximum possible gas flowrate for the problem. Similar way to linearize the term $z_{i,t}\, w_{i,p} Qo_{p\text{-}1}$ and $z_{i,t}\, w_{i,p} Qng_{p\text{-}1}$ with Eq. (3.59) – (3.76),

$$Qo_r = \sum_{i,t,p} zwQo_{i,t,p,r} \tag{3.59}$$

$$Qo_{i,p,r} = Qo_{p-1} Qrec_i \left(1 + bD(r-p+1)\right)^{-\frac{1}{b}} \tag{3.60}$$

$$wQo_{i,p,r} + wQo1_{i,p,r} = Qoe_{i,p,r} \tag{3.61}$$

$$wQo_{i,p,r} \leq w_{i,p} Qo_0 \tag{3.62}$$

$$wQo1_{i,p,r} \leq (1 - w_{i,p}) Qo_0 \tag{3.63}$$

$$wzQo_{i,t,p,r} + wzQo1_{i,t,p,r} = wQo_{i,p,t} \tag{3.64}$$

$$wzQo_{i,t,p,r} \leq z_{i,t} Qo_0 \tag{3.65}$$

$$wzQo1_{i,t,p,r} \leq (1 - z_{i,t}) Qo_0 \tag{3.66}$$

$$wQo_{i,t,p,r}, \quad wQo1_{i,t,p,r}, \quad wzQo_{i,t,p,r}, \quad wzQo1_{i,t,p,r} \geq 0 \tag{3.67}$$

$$Qng_r = \sum_{i,t,p} zwQng_{i,t,p,r} \tag{3.68}$$

$$Qng_{i,p,r} = Qng_{p-1} Qrec_i \left(1 + bD(r-p+1)\right)^{-\frac{1}{b}} \tag{3.69}$$

$$wQng_{i,p,r} + wQng1_{i,p,r} = Qnge_{i,p,r} \tag{3.70}$$

$$wQng_{i,p,r} \leq w_{i,p} Qng_0 \tag{3.71}$$

$$wQng1_{i,p,r} \leq (1 - w_{i,p}) Qng_0 \tag{3.72}$$

$$wzQng_{i,t,p,r} + wzQng1_{i,t,p,r} = wQng_{i,p,t} \tag{3.73}$$

$$wzQng_{i,t,p,r} \leq z_{i,t} Qng_0 \tag{3.74}$$

$$wzQng1_{i,t,p,r} \leq (1 - z_{i,t}) Qng_0 \tag{3.75}$$

$$wQng_{i,t,p,r}, \quad wQng1_{i,t,p,r}, \quad wzQng_{i,t,p,r}, \quad wzQng1_{i,t,p,r} \geq 0 \tag{3.76}$$

To linearize the term ($TI_r \cdot x_r$) in the objective function (3.34), we first introduce two new continuous variables XTI_r which is non-negative and $XTI1_r$ which is negative variable with the following constraints,

$$XTI_r + XTI1_r = TI_r \tag{3.77}$$

$$XTI_r \leq x_r M \tag{3.78}$$

$$XTI1_r \geq -(1 - x_r)\, M \tag{3.79}$$

$$XTI_r \geq 0, XTI1_r \leq 0 \tag{3.80}$$

where constraints (3.77), (3.78), and (3.79) ensure that if x_r equal to zero, the XTI_r should be zero; if x_r is one, $XTI1_r$ should be zero. Combining with constraint (3.80), we can have XTI_r equivalent to the product of x_r and TI_r.

Even though the linearization approach introduces more constraints and variables, it reduces all the number of nonlinear terms in the model (F0) and significantly reduces the computational burden. Therefore, incorporating the above linearizations, we have the following reformulated MILP model (LF0),

$$max: NPV = \sum_r \left\{ (GI_r - XTI_r \cdot FT) \left(\frac{1}{(1 + MARR)^r} \right) \right\} - CC(1 - FT) \tag{3.81}$$

$$s.t.\,(3.1)-(3.7),(3.11)-(3.13),(3.17)-(3.23),and\ (3.42)-(3.80) \tag{3.82}$$

3.1.1.3.2 Integrated time variables MILP (LF1)

Compare with the LF0 formulation, the integrated time variables MILP formulation (LF1) only need to address the nonlinear terms ($y_{i,t,p}Qg_{p\text{-}1}$) and ($y_{i,t,p}LFR_r$). The linearization is similar to the LF0 model. The term ($TI_r x_r$) has already been linearized by (3.77)-(3.80).

The product of $y_{i,t,p}$ and LFR_r in constraints of integrated time variables formulation can be linearized as follows,

$$LFRy_{i,t,p,r} + LFRy1_{i,t,p,r} = LFR_r \tag{3.83}$$

$$LFRy_{i,t,p,r} \le y_{i,t,p,r} \cdot M_i \tag{3.84}$$

$$LFRy1_{i,t,p,r} \le (1 - y_{1,t,p}) \cdot M_i \tag{3.85}$$

$$LFRy_{i,t,p,r} \ge 0, \qquad LFRy1_{i,t,p,r} \ge 0 \tag{3.86}$$

where $LFRy_{i,t,p,r}$ and $LFRy1_{i,t,p,r}$ are two new continuous variables, and $LFRy_{i,t,p,r}$ is equivalent to ($y_{i,t,p} \cdot LFR_r$). In addition, the upper bound value M_i is different of each artificial lift method *i*.

To linearize the term $y_{i,t,p}Qg_{p\text{-}1}$, we introduce three new continuous variables, $yQg_{i,t,p,r}$, $yQg1_{i,t,p,r}$, and $Qge_{i,t,p,r}$, and replace Eq. (3.38) with the constraints given in Eqns. (3.87) - (3.92).

$$Qg_r = \sum_{i,t,p} yQg_{i,t,p,r} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.87}$$

$$Qg_{i,p,r} = Qg_{p-1}Qrc_i\big(1 + bD(r-p+1)\big)^{-\frac{1}{b}} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.88}$$

$$yQg_{i,t,p,r} + yQg1_{i,t,p,r} = Qg_{i,p,r} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.89}$$

$$yQg_{i,t,p,r} \le y_{i,t,p} Qg_0 \qquad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.90}$$

$$yQg1_{i,t,p,r} \le (1 - y_{i,t,p}) Qg_0 \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.91}$$

$$yQg_{i,t,p,r} \ge 0, yQg1_{i,t,p,r} \ge 0 \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.92}$$

Equation (3.87) calculates the gas production flowrate at month *r*. The variable $Qg_{i,p,r}$ calculated in Eq. (3.88) is the gas flow rate at month *r* for ALM *i* that is installed on month *p* and uninstalled on month *t*. Equation (3.90) ensures that the variable $yQg_{i,t,p,r}$ becomes zero if ALM *i* is not operational at month *r*. The parameter Qg_0 is set equal to the maximum possible gas flowrate for the problem. Similarly, linearize the term $y_{i,t,p}Qo_{p\text{-}1}$ and $y_{i,t,p}Qng_{p\text{-}1}$ with Eq. (3.93) – (3.104),

$$Qo_r = \sum_{i,t,p} yQo_{i,t,p,r} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.93}$$

$$Qo_{i,p,r} = Qo_{p-1} Qrc_i \left(1 + bD(r - p + 1)\right)^{-\frac{1}{b}} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.94}$$

$$yQo_{i,t,p,r} + yQo1_{i,t,p,r} = Qo_{i,p,r} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.95}$$

$$yQo_{i,t,p,r} \le y_{i,t,p} Qo_0 \qquad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.96}$$

$$yQo1_{i,t,p,r} \le (1 - y_{i,t,p}) Qo_0 \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.97}$$

$$yQo_{i,t,p,r} \ge 0, yQo1_{i,t,p,r} \ge 0 \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.98}$$

$$Qng_r = \sum_{i,t,p} yQng_{i,t,p,r} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.99}$$

$$Qng_{i,p,r} = Qng_{p-1} Qrc_i \left(1 + bD(r - p + 1)\right)^{-\frac{1}{b}} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.100}$$

$$yQng_{i,t,p,r} + yQng1_{i,t,p,r} = Qng_{i,p,r} \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.101}$$

$$yQng_{i,t,p,r} \le y_{i,t,p} Qng_0 \qquad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.102}$$

$$yQng1_{i,t,p,r} \le (1 - y_{i,t,p}) Qng_0 \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.103}$$

$$yQng_{i,t,p,r} \ge 0, yQng1_{i,t,p,r} \ge 0 \quad \forall t,p,r \in T, i \in I, p \le r \le t \tag{3.104}$$

Therefore, incorporating the above linearizations, we have following reformulated MILP model,

$$max: NPV = \sum\nolimits_r \left\{ (GI_r - XTI_r \cdot FT) \left(\frac{1}{(1 + MARR)^r} \right) \right\} - CC(1 - FT) \quad \forall r \in T \tag{3.105}$$

$$s.t. (3.35) - (3.37), (3.27) - (3.31), (3.77) - (3.80), (3.83) - (3.104) \tag{3.106}$$

3.1.1.4 Valid inequalities based on problem characteristics

The operation envelopes for liquid and gas flowrates of many ALMs are relatively large. For example, Weatherford (2013) indicates that the maximum operating flowrate for rod pumps is 6000 BPD. These envelopes are directly translated to technical limitation constraints. However, as shown in **Figure 3.2.2**, the flow rate will continuously decrease as the well produces. The upper bound will be overestimated. The gap from the upper bound will always be larger than the previous time. The production flowrate cannot be larger than the amount dictated by the decline curve obtained, assuming that the ALM with the highest Qrc_i is installed. Equations (3.107)-(3.109) translate this physical limitation into inequality constraints that bound the liquid, gas, oil and natural-gas-liquids flowrates, respectively.

$$Qg_r \le Qg_1 Max_{i \in I}\{Qrc_i\} \left(1 + bD(r - p + 1)\right)^{-\frac{1}{b}} \quad \forall p,r \in T, i \in I, p \le r \tag{3.107}$$

$$Qo_r \le Qo_1 Max_{i \in I}\{Qrc_i\} \left(1 + bD(r - p + 1)\right)^{-\frac{1}{b}} \quad \forall p,r \in T, i \in I, p \le r \tag{3.108}$$

$$Qng_r \le Qng_1 Max_{i \in I}\{Qrc_i\} \left(1 + bD(r - p + 1)\right)^{-\frac{1}{b}} \quad \forall p,r \in T, i \in I, p \le r \tag{3.109}$$

These equations are added to the F0, LF0, F1 and LF1 to tighten their formulations, and the new models are called TF0, TLF0, TF1 and TLF1.

3.1.2 Multistage Stochastic Programming models

1.1.1.1 Problem Statement

In the deterministic models, the flowrate change ratio Qrc_i assumes to be known. The outcomes of Qrc_i after ALM i installed is fixed parameters in MINLP and MILP models. However, the flowrate change ratio in reality is unpredictable and uncertainties in the field operations. The Qrc_i is uncertain, and it is only revealed when ALM i is installed. The uncertainties of flowrate change ratio will significantly impact the operation of ALIP. For example, before installation of Sucker Rod Pump (SRP), the production rate at time period *p-1* is Qg_{p-1}. After installation of SRP, there may be two outcomes, i.e., High production rate at time p (scenario 1), Low production rate at time p (scenario 2). The production in scenario 2 may violate the operating envelope of SRP and results in production halted shown in **Figure 3.2**,

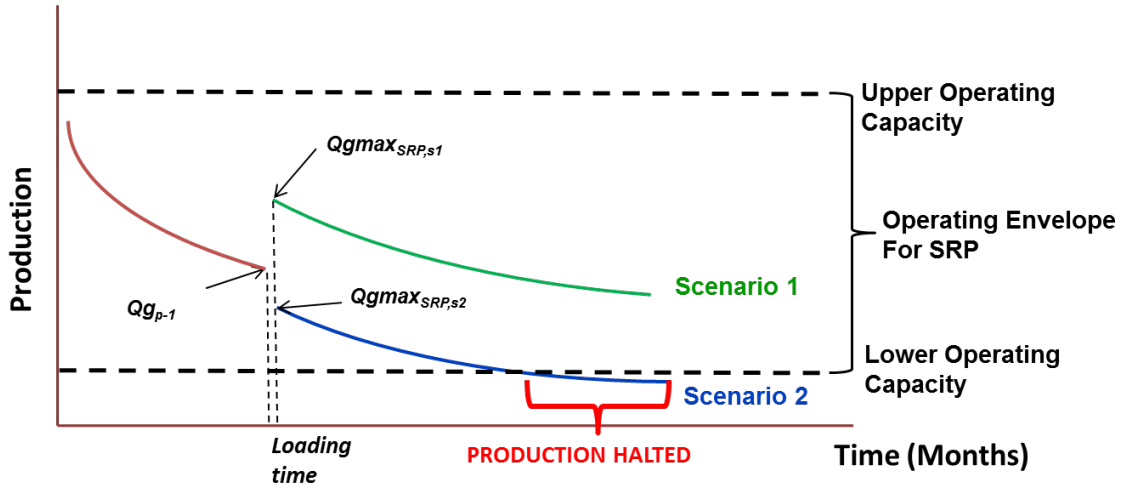

Figure 3.2: Productivity curves of a gas well with artificial lift method installed

Thus, it is necessary and important to incorporate the impact of uncertainty in ALM-dependent production rates and develop a stochastic mixed-integer linear programming model (SMILP).

1.1.1.2 Representation of scenarios

The time horizon is still discretized into time periods, and the probability distributions of uncertain parameters are discrete. Let ξ_i represent the random variable associated with the endogenous uncertain parameter Qrc_i. Assuming ξ_i has two outcomes, i.e., Ω_i = {High (H), Low (L)}, the scenario set, s ϵ S, is generated as the Cartesian product of all possible outcome sets. With nine candidate ALMs, $|S| = 2^9 = 512$. The deterministic equivalent of the stochastic program is obtained by appending the subscript s to all variables to represent their values under scenario s and adding non-anticipativity constraints (NACs) explicitly.

1.1.1.3 The multistage stochastic programming model of ALIP

The deterministic equivalent of the stochastic program is obtained by appending the subscript s to all variables to represent their values under scenario s and adding non-anticipativity constraints (NACs) explicitly.

In TLF1, $w_{i,p}$and $z_{i,t}$ satisfy the logical expression:

$$y_{i,\mathrm{t,p},s} \Leftrightarrow w_{i,\mathrm{p,s}} \wedge z_{i,\mathrm{t,s}} \tag{3.110}$$

Equations (3.111) - (3.114) translate the logical expression given in Eq. (3.110) into constraints. Equation (3.114) prevents ALMs to be uninstalled without installation.

$$y_{i,\mathrm{t,p},s} \geq z_{i,\mathrm{t,s}} + w_{i,\mathrm{p,s}} - 1 \qquad \forall i \in I, t, p \in T, s \in \boldsymbol{S} \tag{3.111}$$

$$z_{i,\mathrm{t,s}} \geq y_{i,\mathrm{t,p},s} \qquad \forall i \in I, t, p \in T, s \in \boldsymbol{S} \tag{3.112}$$

$$w_{i,\mathrm{p,s}} \geq y_{i,\mathrm{t,p},s} \qquad \forall i \in I, t, p \in T, s \in \boldsymbol{S} \tag{3.113}$$

$$\sum_{p} w_{i,p,s} - \sum_{t} z_{i,t,s} = 0 \qquad \forall i \in I, s \in \boldsymbol{S} \tag{3.114}$$

The first set of NACs (3.115) is for p = 2, the installation time of the first ALM. Note that our model assumes the well loaded at the first time period, and the first ALM selected is installed in the second time period. At this stage, all scenarios are indistinguishable because none of the ALMs is installed.

$$w_{i,2,s} = w_{i,2,1} \qquad \forall\, i \in I, s \in \boldsymbol{S}/\{1\} \tag{3.115}$$

We define the subset $\boldsymbol{B} \subset \boldsymbol{S} \times \boldsymbol{S}$ for scenarios s and s' which differ in the outcome of one method**.** The remaining NACs for scenario pairs $(s, s') \in \boldsymbol{B}$ should be active until the differentiating event occurs. After ALM i is uninstalled at period t, the decision variable to install other methods $j \neq i$ after t should be indistinguishable for scenarios s and s'. Thus, the remaining NACs can be expressed with Eq. (3.116), which is converted to Eqns. (3.117) and (3.118).

$$\begin{bmatrix} z_{i,t,s} \\ w_{j,l,s} = w_{j,l,s'} \end{bmatrix} \vee \left[\neg z_{i,t,s}\right] \qquad \forall l \geq t+1, (s, s') \in \boldsymbol{B} \tag{3.116}$$

$$w_{j,l,s} - w_{j,l,s'} \leq 1 - z_{i,t,s} \qquad \forall l \geq t+1, (s, s') \in \boldsymbol{B} \tag{3.117}$$

$$w_{j,l,s} - w_{j,l,s'} \geq z_{i,t,s} - 1 \qquad \forall l \geq t+1, (s, s') \in \boldsymbol{B} \tag{3.118}$$

The deterministic equivalent of the stochastic program is obtained by appending the subscript *s* to all variables to represent their values under scenario *s* and adding non-anticipativity constraints (NACs) explicitly.

To combine the NACs and append the subscript s to all variables in TLF0 and TLF1, two resulting large-scale mixed integer linear programmings are referred to as SMIL0 and SMILP1.

1.1.2 Case Study

We use the deterministic models for planning ALM infrastructure for two case studies of Woodford Shale Horizontal Well: (I) a high liquid flowrate well, and (II) a low liquid flowrate well (Valbuena, 2015)15. The planning horizons were up to 48 months. The mathematical programs are modeled in GAMS 24.6.1 and solved using a Lenovo ThinkStation P900 with Intel 2.3 GHz CPU and 24 GB RAM. The MILP is solved using CPLEX 12.6.3, and the MINLP is solved with SCIP 3.2.

3.1.2.1 Field Application: Woodford Shale Horizontal Well

Case one is a horizontal well with a long lateral section that is hydraulically fractured in eight different stages. The well depth is about 17,180 ft MD (measure depth), and 11,990 ft TVD (true vertical depth). The EOT (End of Tubing) depth is 11,900 ft MD at a well deviation of 60° where the wireline entry guide (WEG) is located. It is completed with a 2 7/8-in. production tubing in a 5 ½-in. casing in a packer-less configuration. The completion string includes an XN nipple at 11,860 ft MD with a well deviation of approximately 60°, and 13 gas lift valves (GLVs) with the shallowest at 1,950 ft and the deepest at 10,700 ft. The well is currently operating with the gas lift using a well-site compressor. The following casing and tubing pressures before the gas lift conversion were measured: casing head pressure (CHP) of 1,200 psi, closed-in casing head pressure (CICHP) of 2,500 psi, flowing tubing head pressure (FTHP) of 300 psi, closed-in tubing head pressure (CITHP) = 1,200 psi. The reservoir pressure and temperature are 6,300 psi and 220° F, respectively. The base productivity information to generate the inflow performance relationship was obtained from a flowing gradient survey, which indicated a gas rate of 900 MscfD, WC(Water Cut) of 96%, GLR of 1,950 scf/STB, and the flowing bottom hole pressure of 2,400 psi.

Case two is a horizontal well with a long lateral. The well depth is about 10,350 ft MD, and 7,100 ft TVD. It is completed with a 2 3/8-in. production tubing in a 5 ½-in. casing. The well is currently operating with a plunger installed at 7,261 ft MD and 68° with a stop collar. The casing and tubing pressures are recorded as casing head pressure (CHP) of 260 psi, closed-in casing head pressure (CICHP) of 284 psi, flowing tubing head pressure of (FTHP) of 130 psi, closed-in tubing head pressure (CITHP) of 154 psi. The reservoir pressure and temperature are 2,500 psi and 190° F, respectively. The production history shows erratic production, one of the typical symptoms of liquid loading. The well has scale problems mainly due to the hard formation water characteristics with total dissolved solids of 128,000 ppm (12.8%). It was assumed that the well dies and the ALS installation will restore the production to the normal production decline. In this case, an incremental gas production of 200 Mscf/D is expected, which represents approximately 6,080 Mscf/Month.

3.1.2.2 Comparison of eight deterministic models

We use the deterministic models for planning ALM infrastructure for two case studies of Woodford Shale Horizontal Well: (I) a high liquid flowrate well, and (II) a low liquid flowrate well (Valbuena, 2015). The planning horizons were up to 48 months. The mathematical programs are modeled in GAMS 24.6.1 and solved using a Lenovo ThinkStation P900 with Intel 2.3 GHz CPU and 24 GB RAM. The MILP is solved using CPLEX 12.6.3, and the MINLP is solved with SCIP 3.2.

3.1.2.2.1 Comparison of eight models on Case I

To compare the different formulations arising from the two different binary variables, we look at the objective function, the computational time, the number of discrete and continuous variables, the number of constraints and the number of nonlinear nonzero terms for different problems

instances. Table 3.1 summarizes the model statistics of F0, F1, LF0, LF1, TF0, TF1, TLF0, and TLF1. The model statistics include the numbers of discrete and continuous variables, the number of constraints, and the number of nonlinear nonzero terms (where applicable). Table 3.1 reveals that models with integrated binary variables (F1, TF1, LF1, TLF1) have more binary, continuous variables, and constraints than the other models. Additionally, the MINLP models, F1, and TF1 have fewer nonlinear nonzero terms than F0 and TF0. By adding valid inequalities, TF0, TF1, TLF0, TLF1 have the same number of discrete and continuous variables and the slightly large number of constraints compared with models F0, F1, LF0, LF1 without tightening constraints. Table 1 also shows that model sizes have an exponential increase with the planning horizon. With a 24-month planning horizon, all models have nearly 4 million constraints, which may require significant computational effort.

Table 3.1: Models Statistics of Case Study I

Month	F0				F1			
	Dis. Var.	Con.Var	Const.	NL.NZ	Dis. Var.	Con.Var	Const.	NL.NZ
12	228	314	249,864	998,014	1,308	1,394	259,828	14,236
18	342	470	1,261,407	5,044,591	2,934	3,062	1,289,592	47,976
24	456	626	3,983,778	15,935,164	5,208	5,378	4,043,612	113,576
Month	TF0				TF1			
	Dis. Var.	Con.Var	Const.	NL.NZ	Dis. Var.	Con.Var	Const.	NL.NZ
12	228	314	249,908	998,014	1,308	1,394	259,872	14,236
18	342	470	1,261,475	5,044,591	2,934	3,062	1,289,660	47,976
24	456	626	3,983,870	15,935,164	5,208	5,378	4,043,704	113,576
Month	LF0				LF1			
	Dis. Var.	Con.Var	Const.	NZ	Dis. Var.	Con.Var	Const.	NZ
12	228	24,598	286,385	597,843	1,308	34,858	297,310	616,422
18	342	73,616	1,371,323	2,820,419	2,934	117,032	1,421,847	2,913,789
24	456	163,242	4,228,037	8,631,763	5,208	276,642	4,363,664	8,889,096
Month	TLF0				TLF1			
	Dis. Var.	Con.Var	Const.	NZ	Dis. Var.	Con.Var	Const.	NZ
12	228	24,598	286,429	597,887	1,308	34,858	297,354	616,466
18	342	73,616	1,371,391	2,820,487	2,934	117,032	1,421,915	2,913,857
24	456	163,242	4,228,129	8,631,855	5,208	276,642	4,363,756	8,889,188

Dis. Var., discrete variables; Con. Var., continuous variable; Const., constraints; NL.NZ, nonlinear nonzeros; NZ, nonzeros.

The artificial lift infrastructure planning problem for Case Study I is solved for planning horizons of 12, 18 and 24 months using the eight models. The maximum CPU time was limited to 10 hours. All the models recommend installing Sucker Rod Pump (SRP) for the overall planning horizon for problems solved to optimality. The NPV and solution times of total eight models are summarized in Table 3.2.

Table 3.2: Comparison of Performance of MILP, TMINLP and TMILP of Case Study I

		Solution time (s)					
				Valid Inequalities			
Month	NPV ($)	LF0	LF1	TF0	TF1	TLF0	TLF1
12	2805092	55	4	122	7	15	1
18	3714696	1916	101	5123	138	246	5
24	4418921	NA	978	14300	741	2356	22

The first approach is solving the original MINLP problem (F0) with its valid constraints model (TF0) by global optimizer SCIP. Its linearization reformulation model (LF0) and its valid constraint formulation (TLF0) are solved by CPLEX solver. The second approach is to solve the MILP problem (F1) with integrated binary variable $y_{i,t,p}$ and its valid constraints model (TF1) with global SCIP solver. Their linearization-reformulated problem (LF1) and (TLF1) are solved with CPLEX. From Table 3.1, we know that F1 has more continuous variables and constraints with fewer nonlinear nonzero terms.

Furthermore, we can see from Table 3.2 that without valid inequalities, only linearization reformulated problem (LF1) can be solved by SCIP global solver within 10 CPU hours. We know from Table 3.1 and Table 3.2 that the reformulated models (TF1, LF1 and TLF1) can obtain

solutions in much shorter time than original models (TF0, LF0, TLF0) with decrease solution time up to two magnitudes. It is because that the product of $w_{i,p}$ and $z_{i,t}$ used in objective function and decline curve constraints contribute to huge nonconvexity and to the problem in large-scale period. Overall, it is harder to solve the MINLP formulation (TF0, TF1) with an increased planning horizon without the linearization formulation. The additional inequality constraints reduce the solution times up to two orders of magnitude both for MINLP and MILP models. An integrated approach with valid inequality constraints MILP model (TLF1) is the best option for an expert because this formulation can be solved within 30 CPU seconds for 24 months planning horizon. The TLF1 has the tightest bound and smallest feasible region among all eight models. This creates the potential to apply to the large longer life span gas wells.

3.1.2.2.2 Results for TLF1 model in Case I and Case II

The planning problems of Case Studies I and II are solved to optimality using the TLF1 formulation for planning horizons up to 48 months. The results are compiled in Table 3.3. The solutions reveal that a single ALM is installed and operated until the end of the planning horizons for all cases considered. For Case Study I, the optimal ALM is the Sucker Rod Pump. For Case Study II, which has a low liquid flowrate, the optimal ALM is Sucker Rod Pump (SRP) if the planning horizon is 12 months, Well Head Compression (WHC) for the planning horizon of 24 months, and Velocity String (VS) for planning horizons of 36 and 48 months. The flowrate change ratio of SRP is higher than WHC and VC, and it has lower operating costs. However, the liquid flowrate drops below the minimum allowable value for SRP for planning horizons longer than 12 months. Because the production is relatively low from this well, the model favors the installation of an ALM that can be operated for the complete planning horizon. For longer planning horizons (36 and 48 months), the VS is preferred over WHC due to its lower operating costs.

Table 3.3: The solution of Case Studies I and II using the TLF1 model

	Months	NPV ($)	Time (s)	Optimal ALM
Case I	12	2805092	1	SRP
	24	4418921	22	SRP
	36	5424537	154	SRP
	48	6093133	129	SRP
Case II	12	129811	0.8	SRP
	24	203164	12	WHC
	36	257223	82	VS
	48	296336	288	VS

3.1.2.3 Computation results for MSSP models

In this section, we compared two MSSP models SMILP0 and SMILP1. The well was hydraulically fractured in eight different stages. The production before the loading indicated a gas rate of 800 thousand standard cubic feet per day (Mscf/D), a liquid rate of 520 barrels per day (BPD) (of which 500 barrels were water and the remaining oil), a water cut of 96%, and a gas-to-liquid ratio of 1,950 standard cubic feet per stock tank barrel (scf/STB). The well is currently operated using gas lift with a well-site compressor (Valbuena, 2015). The case study assumes that the planning was carried out at the first loading prior to the installation of gas lift. For uncertain parameter Qrci, High outcome is assumed to be 20% above its nominal value, and Low outcome 20% below with equal probabilities. The planning horizon is 12 months, which is divided into 12 equal time periods. The mathematical programs are constructed in GAMS 24.6.1, and solved using CPLEX 12.6.3 on a Lenovo ThinkStation P900 with Intel 2.3 GHz CPU and 24 GB RAM. The solutions of SMILP0 and SMILP1 recommend installing SRP throughout the planning horizon for all scenarios. The ENPV is $2,805,092, which is 8.8% higher than the ENPV if gas lift were used. For this problem, SMILP0 and SMILP1 are solved to 0.1% optimality gap in 7039 seconds and 27 seconds, respectively.

3.1.2.3.1 A hypothetical case study

We also constructed a hypothetical case study where the operating envelope of SRP is modified by setting the minimum allowable flow rate to 300 BPD. The planning horizons are 12, 16 and 20 months. The maximum CPU time is limited to 20 hours. Table 3.4 summarizes the results.

Table 3.4: The ENPVs, optimality gaps, and solution times for SMILP

	SMILP0			SMILP1		
Months	ENPV ($)	GAP (%)	Times (s)	ENPV ($)	GAP (%)	Times (s)
12	2,783,894	0.08	4948	2,783,894	0.08	20
16	3,416,506	42.51	72000	3,416,507	0.1	1959
20	3,840,889	49.39	72000	3,916,996	11.89	72000

The SMILP0 cannot be solved to optimality for 16-month and 20-month planning horizons (Table 1), whereas SMILP1 cannot be solved to optimality only for 20-month planning horizon. We hypothesize that the additional binary variable $y_{i,t,p,s}$ helps CPLEX to branch better when solving SMILP1.

The solution of the deterministic model, using the nominal values of Qrc_i, recommends installing and operating SRP throughout the planning horizon. If the deterministic solution were implemented, the ENPV would have been $3,252,676. With the deterministic solution, under scenarios with Low outcome of $Qrc_{,SRP,}$ the production rate violates the operational envelope of SRP in month 13, and the production is halted. The optimal plan obtained as the solution of SMILP2 installs SRP initially for all scenarios. At the 7th time period, the ESP is installed for scenarios with Low outcome of $Qrc_{,SRP,}$ while SRP stays in operation for the rest of the scenarios. The difference between the ENPVs of the stochastic and deterministic solutions, i.e., the value of the stochastic solution, is $163,831, a 5% increase from the deterministic solution.

3.1.3 Conclusion

This section contributes to the new and growing body of literature in PSE on shale-gas production and infrastructure planning from horizontal wells. First, eight deterministic optimization models for artificial lift infrastructure planning of shale gas producing horizontal wells are presented. We start to develop two discrete-time, large-scale, nonconvex, mixed-integer nonlinear programming models (F0 and F1) considering the possibility of multiple different artificial lift methods installed over the lifespan of a well. The original model incorporates technical limitation constraints such as the operational and design limitations of each ALM, previous expert experience and limitations stated in the typical attribute tables. Then, a mixed-integer linear programming model (LF0 and LF1) is developed by linearizing the original MINLP model. A set of valid inequalities are introduced to tighten both MINLP and MILP models (TF0, TF1, TLF0 and TLF1), and they reduced the solution times up to two orders of magnitude. The TLF1 has the best performance among all eight models.

Additionally, we presented two stochastic models for generating ALIPs for shale gas producing horizontal wells. Both models incorporate the stochastic ALM-dependent well production, which is an endogenous uncertainty. The first model, SMILP1, has two decision variables, which track the installation and uninstallation times of each ALM separately. In the second model, SMILP2, a new binary variable that integrates both installation and uninstallation times of each ALM is introduced, and the production rate and planning constraints are expressed using it. The additional binary variable in SMILP2 reduced the solution times up to two orders of magnitude.

3.2 Pharmaceutical Clinical Trial Planning Problem

The pharmaceutical industry is a global business with over one trillion U.S. dollars per year market with extensive supply chains throughout the world (PhRMA, 2016). A potential drug identified at discovery stage has to go through pre-clinical testing, generally laboratory and animal model studies, before applying for approval by regulatory bodies, such as The Food and Drug Administration (FDA) in USA. The goal of these laboratory and animal model studies is to understand how the drug works and assess its safety. The potential drugs that successfully complete pre-clinical trials enter clinical trial phase. Clinical trials aim to demonstrate the safety and efficacy of the potential drug and are designed with and carried out under strict guidelines and supervision of regulatory bodies. If a drug successfully completes the clinical trials and is approved by the regulatory bodies, the drug is manufactured and distributed to the market. Pharmaceutical manufacturers are under pressure to improve the efficiency of the pharmaceutical R&D pipeline, partially because the patent protections of a number of significant brand-name drugs will soon expire (IMS, 2013).

Scheduling and planning of clinical trials is one of the efficient ways to reduce the cost of developing new drugs. There are three phases in clinical trials. The goal of Phase I clinical trials is to assess the safety and dosage of the drug and to understand how it is metabolized in the body. The lengths of Phase I clinical trials are several months. Approximately 70% of drugs will move to the next phase. Phase II clinical trials are used to evaluate the drug's effectiveness and short-term side effects on a limited number of target patient volunteers. Phase II may take from several months to two years. Approximately 33% of drugs pass Phase II clinical trials and move to the next stage. Phase III clinical trials aim to assess the benefit-risk ratio of the drug using a large number of target patient volunteers (in the order of thousands). Phase III takes one to four years to complete and approximately 25–30% of drugs pass Phase III (FDA, 2017). Clinical trial planning

is complicated due to its highly stochastic nature: pharmaceutical companies do not know which drugs will successfully complete clinical trials a priori. The outcomes of clinical trials significantly influence drug development plan, the investments and overall profits. Clinical trial planning is a series of trade-offs between maximizing expected economic returns and minimizing risk and maintaining a diverse portfolio of drugs under the limited drug development dollars available. The plan tries to accomplish these goals by selecting which potential drugs to push through the clinical trial pipeline and when to start the clinical trials of selected drugs. This is a challenging problem due to its strong combinatorial character and impact of uncertainty.

There have been a number of studies that introduced stochastic programing models for solving pharmaceutical clinical trial planning problem. Here, we limit our review to the ones that explicitly incorporated the impact of clinical trial outcome uncertainty. Gatica et al. (2003) developed a model that integrates the production planning and investment strategy simultaneously in pharmaceutical industries considering the impact of uncertainty in the outcome of clinical trials. The case studies considered in the paper included only two phases of the clinical trials with three products yielding 64 scenarios. However, if all stages of clinical trials were considered, the size of the resulting model would have become prohibitively large and would require heuristic approaches to solve. Colvin and Maravelias (2008) developed a MSSP model for clinical trial planning and included the impact of endogenous clinical-trial outcome uncertainty. In a later study, they exploited the structure of the problem to reduce the number of scenarios and extended their model to account for resource planning by introducing outsourcing decisions (Colvin and Maravelias, 2008). Colvin and Maravelias (2010) introduced a number of theoretical properties, which reduce the problem size and tighten the formulation, and developed a novel branch and cut algorithm to solve the resulting problem efficiently. Sundaramoorthy et al. (2012) proposed a stochastic

programming formulation that integrates the capacity planning and clinical trial planning and that takes into account uncertainty in the outcomes of clinical trials. The proposed formulation was solved for problems with a total of 256 scenarios in 1127 CPUs. The solution time increases significantly for problems with more than 256 scenarios.

Multistage stochastic programs and their deterministic equivalent forms quickly become computationally intractable because the number of scenarios and the corresponding decision trees grow exponentially as the number of uncertain parameters increases. Furthermore, for discrete-time MSSPs, the problem size increases rapidly especially in numbers of variables and NACs with increases in the length of the planning horizon. Thus, solving any large-scale MSSP model, especially with endogenous uncertainty, requires significant computational effort. Most recent work on MSSPs with endogenous uncertainty focuses on developing new algorithms or decomposition frameworks for solving these problems.

Most of these algorithms decompose the original problem into smaller sub-problems to generate a feasible solution. No recent literature studied the impact of different decision variable definitions and corresponding sequencing and non-anticipativity constraints on the size and solution times of the resulting MSSP formulations constructed for clinical trial planning.

Motivated by the above, we propose two new MSSP formulations, CM1 and CM2, for pharmaceutical clinical trial planning problem. The first formulation, CM1, employs two decision variables, which separately track the start and end time points of clinical trials. The second formulation, CM2, introduces an additional binary variable, which tracks both start and end time points of a clinical trial. We applied both formulations to solve 42 instances of clinical trial planning problem (Christian and Cremaschi, 2017). For comparison, all instances were also solved using the MSSP formulation of (Colvin and Maravelias, 2008), which will be referred to as CM3

here. In the last section, the MSSP formulation with reduction of scenarios in (Colvin and Maravelias, 2010), referred to CM4 is also compared with CM1 and CM2. All instances were solved to 0.1% optimality gap using ILOG CPLEX Optimization Studio (Version 12.6.3.0, IBM, Armonk, NY, USA). The results reveal that CM1 and CM2 were consistently solved up to two orders of magnitude faster than CM3 for all cases. A closer look at the branching trees indicates that the optimum was obtained with fewer branches for CM1 or CM2 than CM3.

3.2.1 Problem Statement

Most of these algorithms decompose the original problem into smaller sub-problems to generate a feasible solution. No recent literature studied the impact of different decision variable definitions and corresponding sequencing and non-anticipativity constraints on the size and solution times of the resulting MSSP formulations constructed for clinical trial planning.

Following the problem definition of (Colvin and Maravelias, 2008) and nomenclature given in Appendix B1, the problem addressed in this paper can be stated as follows. Givens are

(1) A set of candidate drugs ($i \in I$) that should go through a set of clinical trials ($j \in J = \{PI, PII, PIII\}$),

(2) The length of the planning horizon, which is discretized into equal time periods $p = 1, 2, 3 \dots T$ (period p starts at time $p-1$ and ends at time p), $t = 1, 2, 3 \dots T + max_{ij}(\tau_{ij}\ \forall i \in I, j \in J)$ (period t starts at time $t-1$ and ends at time t),

(3) Cost C_{ij}, resource requirements(s) ρ_{ijr} and duration τ_{ij} of each clinical trial,

(4) Potential revenue of each drug if it successfully completes all clinical trials, rev_i^{max}. We assume that the patent life of a drug begins to shrink once it starts its first clinical trial (i.e., $j = 1$). Losses are represented by two penalty terms: γ_i^D (loss of market) and γ_i^L (loss of patent life).

The problem determines: (a) which clinical trials to start; and (b) when to start the selected clinical trials. The objective is to maximize the expected net present value (ENPV) of the pipeline.

3.2.1.1 Scenario Representation

The source of the uncertainty for this problem stems from the outcomes of the clinical trials for each drug. A drug can either pass (P) a clinical trial or fail (F) it. Each candidate drug must successfully complete $|J|$ clinical trials before any revenue associated with that drug can be realized. Because a drug will drop out of the pipeline (i.e., will not continue to the subsequent clinical trials) when it fails a clinical trial, a single uncertain parameter can be used to represent the outcomes of the clinical trials for a drug and will have a total of $|J+1|$ outcomes. Let the clinical-trial outcome uncertainty of drug i be defined by the parameter Ω_i, then the outcome space of this parameter with three clinical trials ($J = \{PI, PII, PIII\}$) can be reduced to $\{PI-F, PII-F, PIII-F, PIII-P\}$. For example, the outcome $PII-F$ for drug i means drug i successfully completed (i.e., passed) clinical trial PI but failed clinical trial PII.

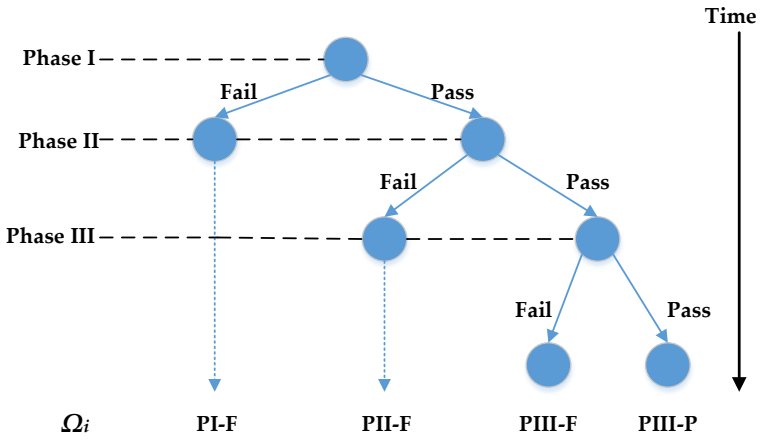

Figure 3.2.2: The outcome space of uncertain parameter associated with drug i, Ω_i

The scenarios for the MSSP model are constructed as the Cartesian product of uncertain parameter outcomes. The total number of scenarios for a clinical trial planning problem with $|I|$ drugs is then equal to $|S| = 4^{|I|}$, where S is the scenario set. Assuming that the probabilities of outcomes are independent, the probability for each scenario s can be calculated by $P_s = P(\Omega_1 \cap \Omega_2 \cap \dots \Omega_{|I|}) = P(\Omega_1) \times P(\Omega_2) \times \dots P(\Omega_{|I|}) = \prod_i P(\Omega_i)$. For a more detailed discussion of scenario representation, we refer the reader to (Colvin and Maravelias, 2008).

3.2.2 Two Binary Variables Model (CM1)

The first formulation, which is referred to as CM1, uses two binary decision variables, $X_{i,j,p,s}$ and $Y_{i,j,t,s}$. The first binary variable, $X_{i,j,p,s}$Tracks when a drug starts a clinical trial. Formally, it is equal to one if drug i starts clinical trial j at time period p in scenario s. The second binary variable, $Y_{i,j,t,s}$, tracks when a clinical trial is completed and is equal to one if drug i completes clinical trial j at time period t in scenario s.

3.2.2.1 Scheduling and Resource Constraints

Equations (3.119) - (3.124) state that each clinical trial j for drug i is started at most once (3.119), that it ends at most once (3.120) and that all clinical trials that have been started should be completed (3.121). Equation (3.122) ensures that as soon as the start time of each clinical trial is determined, its end time is also known. Constraint in Equation (3.123) states that clinical trial j for drug i cannot be started before the drug completes previous clinical trial $j-1$. Equation (3.124) limits the total resource utilized by active clinical trials at any time period to the maximum level for each resource type r.

$$\sum_t Y_{i,j,t,s} \leq 1 \quad \forall i,j,s \tag{3.119}$$

$$\sum_{p} X_{i,j,p,s} \leq 1 \quad \forall i,j,s \tag{3.120}$$

$$\sum_{p} X_{i,j,p,s} - \sum_{t} Y_{i,j,t,s} = 0 \quad \forall i,j,s \tag{3.121}$$

$$Y_{i,j,p+\tau_{i,j},s} = X_{i,j,p,s} \quad \forall i,j,p,s \tag{3.122}$$

$$\sum_{p}^{t\prime} X_{i,j,p,s} \leq \sum_{t}^{t\prime} Y_{i,j-1,t,s} \quad \forall i,j>1,t',s \tag{3.123}$$

$$\sum_{i}\sum_{j}\sum_{t'>p-\tau_{i,j}}^{t'\leq p} \rho_{i,j,r} X_{i,j,t',s} \leq {\rho_r}^{max} \ \forall r,p,s \tag{3.124}$$

3.2.2.2 Non-Anticipativity Constraints

The outcomes of clinical trials are the source of uncertainty for this problem. The scenarios differ in the outcomes of certain (drug, clinical trials) pairs (i,j). The first set of NACs is for p = 1, i.e., the first time period. At this stage, all scenarios are indistinguishable, i.e., none of the drugs completed any clinical trials.

$$X_{i,1,1,1} = X_{i,1,1,s} \ \forall i,s \tag{3.125}$$

We define the subset $\boldsymbol{B}$ of $S \times S$ as scenarios s and s', which are distinguishable in the outcome of one (drug, clinical trial) pair $(i^{s,s'}, j^{s,s'})$. The non-anticipativity constraints (NACs) should be enforced for scenarios $(s,s') \in \boldsymbol{B}$ until the differentiating event occurs, i.e. (drug, clinical trial) pair $(i^{s,s'}, j^{s,s'})$ is completed. The NACs for $(s,s') \in \boldsymbol{B}$ should be active until the differentiating event, i.e., until the drug $i^{s,s'}$ completes clinical trial $j^{s,s'}$. Thus, the remaining NACs can equivalently be written as:

$$\left[\begin{matrix}\neg \sum_{t'\leq p} Y_{i^{s,s'},j^{s,s'},t',s} \\ X_{i,j,p,s} = X_{i,j,p,s'}\end{matrix}\right] \vee \left[\sum_{t'\leq p} Y_{i^{s,s'},j^{s,s'},t',s}\right] \ \forall i,j,p \geq 1, (s,s') \in \boldsymbol{B} \tag{3.126}$$

$$-\sum_{t'\le p} Y_{i^{s,s'},j^{s,s'},t',s} \le X_{i,j,p,s} - X_{i,j,p,s'} \le \sum_{t'\le p} Y_{i^{s,s'},j^{s,s'},t',s} \quad \forall i,j,p \ge 1, (s,s') \in \boldsymbol{B} \tag{3.127}$$

3.2.2.3 Objective function

The objective is to maximize the expected net present value ($ENPV$), which has three components: current revenue Rv_s, future revenue $FRev_s$ and costs Cst_s. The current revenue for scenario s represents the revenue from drugs that have successfully completed all clinical trials within the planning horizon and it is calculated using Equations (3.129). In Equation (3.129), the parameter rev_i^{max} is the potential revenue of drug i once it successfully completes all clinical trials. The revenue is reduced using the penalty parameters ${\gamma_i}^D$ and ${\gamma_i}^L$ for reduced patent life and market share (Colvin and Maravelias, 2008). In Equation (3.130), a binary variable, $D_{i,j,p,s}$, is introduced to determine the cases where drug i successfully completed trial $j-1$ yet it did not start the subsequent clinical trial j.

$$Max: ENPV = \sum_s p_s(Rv_s + FRev_s - Cst_s) \tag{3.128}$$

$$Rv_s = \sum_i \sum_p \left\{ {rev_i}^{max} X_{i,PIII,p,s} - {\gamma_i}^D \sum_{j=PII,PIII} D_{i,j,p,s} - {\gamma_i}^L \left(t + \tau_{i,PIII}\right) X_{i,PIII,p,s} \right\} \forall s \tag{3.129}$$

$$D_{i,j,p,s} = -X_{i,j,1,s} + \sum_{t'>\tau_{i,j-1}}^{t'\le p} X_{i,j-1,t'-\tau_{i,j-1},s} - \sum_{t'}^{t'\le p} X_{i,j,t',s} \quad \forall i,s,j \in \{PII, PIII\}, p \tag{3.130}$$

The future revenue assesses the potential revenue from the clinical trials that have not been completed in the planning horizon and it follows the definition of Colvin and Maravelias (2008). Equations (3.131)-(3.132) are used to calculate future revenue:

$$FRev_s = \sum_i \sum_{j \neq PI} {rev_{i,j}}^{open} f_{i,j} \, D_{i,j,|T|,s} + \sum_i {rev_{i,PI}}^{open} f_{i,PI} D_{i,PI,p,s} + \sum_d \sum_{j \in \{PI,PII\}} \sum_{p > |T| - \tau_{d,j}} {rev_{i,j,t}}^{run} f_{i,j+1} X_{i,j,p,s} \;\; \forall s \tag{3.131}$$

$$D_{i,PI,s} = 1 - \sum_{t'}^{t' \leq |T|} X_{i,PI,t',s} \;\; \forall i, s, j \in \{PII, PIII\} \tag{3.132}$$

The parameter ${rev_{i,j}}^{open}$ in Equation (3.131) is the potential revenue from drug i whose j-th clinical trial has not been started in the planning horizon despite its previous trial $(j-1)$ being completed and it is calculated using Equation (3.132). The parameter ${rev_{i,j,p}}^{run}$ ((3.134)) is the potential revenue of drug i whose clinical trial j will be completed beyond the end of the planning horizon. The parameter $f_{i,j}$ (3.135) represents the fraction of the revenue that would be realized by completing all remaining trials at the end of the planning horizon, and it is used to favor pushing the drugs through the clinical trial pipeline towards the end of the planning horizon (Colvin and Maravelias, 2008). The total cost for scenario *s* is calculated by discounting the total cost incurred due to all completed or running clinical trials by the time discounting factor cd_p shown in Equation (3.136). The discounting factor is calculated for each time period via Equation (3.137), where n_p is the interest rate for period p.

$${rev_{i,j}}^{open} = {rev_i}^{max} - {\gamma_i}^L \left(|T| + \sum_{j' \geq j} \tau_{i,j'} \right) \forall i, j \tag{3.133}$$

$${rev_{i,j,p}}^{run} = {rev_i}^{max} - {\gamma_i}^L \left(p + \sum_{j' \geq j} \tau_{i,j'} \right) \forall i, j, p \tag{3.134}$$

$$f_{i,j} = 0.9 \left[\frac{{rev_i}^{max} - {\gamma_i}^L |T| - \sum_{j' \geq j} C_{i,j}}{{rev_i}^{max} - {\gamma_i}^L |T|} \right] \forall i, j \tag{3.135}$$

$$Cst_s = \sum_{i,j,p} cd_t C_{i,j} X_{i,j,p,s} \ \forall s \tag{3.136}$$

$$cd_p = 1 - \ n_p(t-1) \ \forall p \tag{3.137}$$

The deterministic equivalent of the first multistage stochastic programming formulation (CM1), a large-scale mixed integer linear program (MILP), is then given by:

$$Max: ENPV \tag{3.138}$$

$$X_{i,j,p,s} \in \{0,1\}, Y_{i,j,t,s} \in \{0,1\} \tag{3.139}$$

$$s.t. Constraints \ (3.119) - (3.125), (3.127), (3.129) - (3.137) \tag{3.140}$$

3.2.3 Integrated Time Binary Variable Model (CM2)

The second formulation, which is referred to as CM2, introduces one more binary decision variable, $W_{i,j,t,p,s}$, similar to (Zeng and Cremaschi, 2017a), which tracks both start and end time periods of clinical trial j of drug i. This decision variable is equal to one if (drug, clinical trial) pair (i, j) is started at time period p and ends at time period t. The new binary variable satisfies the following relationship:

$$W_{i,\mathrm{j},\mathrm{t},\mathrm{p},s} \ \Leftrightarrow X_{i,j,p,s} \wedge Y_{i,j,t,s} \tag{3.141}$$

Equations (3.142)-(3.144) translate the logical expression given in Equation (3.141) into constraints.

$$W_{i,\mathrm{j},\mathrm{t},\mathrm{p},s} \ \geq X_{i,j,p,s} + Y_{i,j,t,s} - 1 \ \forall i, j, t, p, \mathrm{s} \tag{3.142}$$

$$X_{i,j,p,s} \geq \ W_{i,\mathrm{j},\mathrm{t},\mathrm{p},s} \ \forall i, j, t, p, \mathrm{s} \tag{3.143}$$

$$Y_{i,j,t,s} \geq W_{i,\mathrm{j},\mathrm{t},\mathrm{p},s} \ \forall i, j, t, p, \mathrm{s} \tag{3.144}$$

With the additional binary variable, we introduce one more constraint: Equation (3.145), which states that each clinical trial can only be started and completed at most once. The original sequencing constraint Equation (3.121) is replaced by Equation (3.146):

$$\sum_{p} W_{i,j,p+\tau_{i,j},p,s} \leq 1 \quad \forall i,j,s \tag{3.145}$$

$$\sum_{t} Y_{i,j,t,s} = \sum_{p} W_{i,j,t+\tau_{i,j},p,s} \quad \forall i,j,s \tag{3.146}$$

Part of scheduling and resource constraints (Equations (3.119),(3.123) and (3.124)), NACs (Equations (3.125) and (3.127)) and the objective function of CM1 are retained in CM2. The new binary variable, $W_{i,j,t,p,s}$, replaces the binary variable $X_{i,j,p,s}$ in Equations ((3.129)–(3.132) and (3.136). The deterministic equivalent of this formulation is also a large scale MILP:

$$Max: ENPV = \sum_{s} p_s(Rv_s + FRev_s - Cst_s) \tag{3.147}$$

$$Rv_s = \sum_{i}\sum_{p}\Bigg\{ {rev_i}^{max} W_{i,PIII,p+\tau_{i,PIII},p,s} - {\gamma_i}^{D} \sum_{j=PII,PIII} D_{i,j,p,s} - {\gamma_i}^{L}(p + \tau_{i,PIII}) W_{i,PIII,p+\tau_{i,PIII},p,s} \Bigg\} \quad \forall s \tag{3.148}$$

$$D_{i,j,p,s} = -W_{i,j,1+\tau_{i,j},1,s} + \sum_{t'>\tau_{i,j-1}}^{t'\leq p} W_{i,j-1,t',t'-\tau_{i,j-1},s} - \sum_{t'}^{t'\leq p} W_{i,j,t'+\tau_{i,j},t',s} \tag{3.149}$$

$$FRev_s = \sum_{i}\sum_{j\neq PI} {rev_{i,j}}^{open} f_{i,j}\, D_{i,j,|T|,s} + \sum_{i} {rev_{i,PI}}^{open} f_{i,PI} D_{i,PI,p,s} + \sum_{d}\sum_{j\in\{PI,PII\}}\sum_{p>|T|-\tau_{d,j}} {rev_{i,j,t}}^{run} f_{i,j+1} W_{i,j,p+\tau_{i,j},p,s} \quad \forall s \tag{3.150}$$

$$D_{i,PI,s} = 1 - \sum_{t'}^{t' \leq |T|} W_{i,PI,t'+\tau_{i,j},t',s} \ \forall i, s \tag{3.151}$$

$$Cst_s = \sum_{i,j,p} cd_t C_{i,j} W_{i,j,p+\tau_{i,j},p,s} \ \forall s \tag{3.152}$$

$$X_{i,j,p,s}, Y_{i,j,t,s}, W_{i,\mathrm{j},\mathrm{t},\mathrm{p},s} \in \{0,1\} \tag{3.153}$$

$$s.t. Constraints\ (3.119), (3.123) - (3.125), (3.127), (3.133) - (3.135), (3.142) - (3.146) \tag{3.154}$$

3.2.4 Model of Colvin and Maravelias (2008) (CM3)

An MSSP model for the clinical trial planning problem was developed by Colvin and Maravelias (2008), and we compared the solution times of CM1 and CM2 to this formulation. In this paper, we refer to the MSSP model of Colvin and Maravelias (2008) as CM3. The entire formulation of CM3 is given in Appendix B2 for completeness. Here, we provide a brief overview. The model CM3 uses binary variable $X_{i,j,p,s}$ to track the start time of clinical trial j of drug i, similar to CM1. The model also defines two more continuous variables, $V_{i,j,p,s}$ and $Z_{i,j,p,s}$, bounded between zero and one. The continuous variable $V_{i,j,p,s}$ is defined such that it is equal to one if drug i completes clinical trial j by the beginning of time period t in scenario s. For example, if the (drug, clinical trial) pair (i, j) is completed at time period p in scenario s, the variable $V_{i,j,t\prime,s}$ will be equal to one for all $t' \geq p$. The third continuous variable, $Z_{i,j,p,s}$, becomes one if drug i completes clinical trial $j-1$ by time period p and has not started clinical trial j.

A comparison of the decision variable values of the three formulations for an example is shown in Figure 3.2.3. In this example, we assume that the first clinical trial (PI) of a drug D1 takes two time periods, and it is started on time period one. Then, binary variables $X_{D1,PI,1,s}$ in CM1, CM2

and CM3 would all be equal to one. In CM1 and CM2, binary variables $Y_{D1,PI,3,s}$ would be equal to one because drug D1 would complete clinical trial PI at time three. Furthermore, in CM2, the binary variable $W_{D1,PI,3,1,s}$ would be equal to one, which indicates that the clinical trial (D1, PI) starts at time one and ends at time three. In CM3, the variable $V_{D1,PI,p,s}$ would be equal to one for all $p \geq 3$ and the variable $Z_{D1,PII,3,s}$ would be equal to one, because the clinical trial (D1, PI) is completed while the trial (D1, PII) have not been started yet.

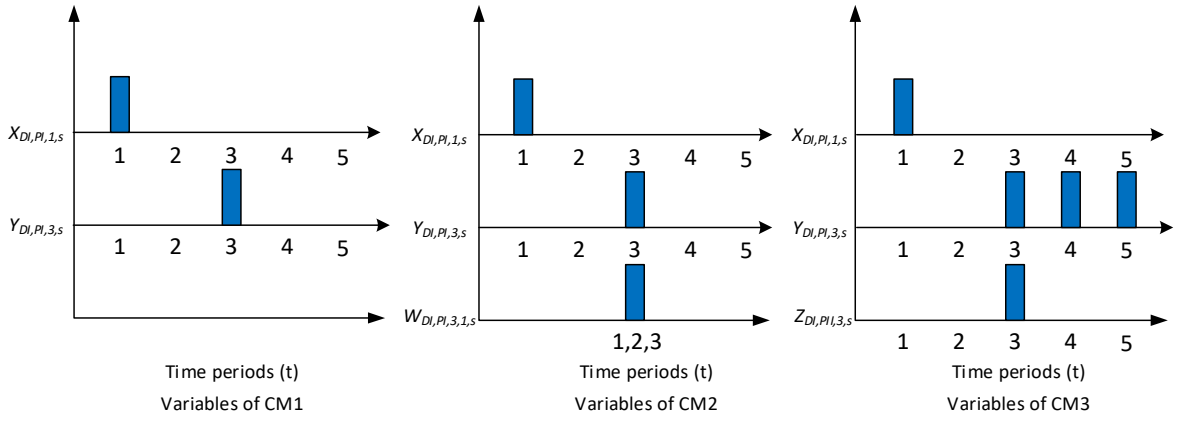

Figure 3.2.3: Comparison of scheduling variables in CM1, CM2 and CM3 formulations

3.2.5 Model of Colvin and Maravelias (2010) (CM4)

In 2010, Colvin and Maravelias proposed models by reduction of decision variables and scenario specific constraints, where remain the same definition of decision variables. A slightly revised version of the MSSP formulation by Colvin and Maravelias (2010) is given in Equation (3.155)-(3.164). The decision variable $X_{i,j,t,s}$ is equal to one if clinical trial *j* of drug *i* is started at time *t* in scenario *s*. Equations (3.155)-(3.158) calculate the ENPV. Equations (3.159)-(3.161) are sequencing constraints. Per Colvin and Maravelias (2010), a clinical trial *j'* of drug *i* cannot be started at time *t* in scenario *s*, (1) if drug *i* fails a prerequisite trial (*j*<*j'*) in scenario *s*, and (2) if $t \leq \sum_{j=1}^{j'-1} \tau_{ij}$. We define set **N** to include trial *j'* of drug *i* that cannot be started at time *t* in scenario

s, and set decision variables associated with these (drug, clinical trial) pairs for these time periods and scenarios equal to zero by Eq. (3.161). Equation (3.162) enforces the resource limitations. The NACs are given in Eq. 10 for the first time period, and Eq. 11 for t>1. For Eq. 11, we define the subset **B** of **S**×**S** as scenarios *s* and s', which are distinguishable in the outcome of one (drug, clinical trial) pair (is,s',js,s').

$$Max: ENPV = \sum_{s} p_s(Rev_s + FRev_s - Cst_s) \tag{3.155}$$

$$\begin{aligned} Rev_s = \sum_i \sum_t \Bigg\{ & rev_i{}^{max} X_{i,PIII,t,s} \\ & - \gamma_i{}^{D} \sum_{j=PII,PIII} (-X_{i,j,1,s} + \sum_{t' > \tau_{i,j-1}}^{t' \le t} X_{i,j-1,t'-\tau_{i,j-1},s} - \sum_{t'}^{t' \le t} X_{i,j,t',s}) \\ & - \gamma_i{}^{L}(t + \tau_{i,PIII}) X_{i,PIII,t,s} \Bigg\} \quad \forall s \end{aligned} \tag{3.156}$$

$$\begin{aligned} FRev_s = & \sum_i \sum_{j \neq PI} rev_{i,j}^{open} f_{i,j} \, (-X_{i,j,1,s} + \sum_{t' > \tau_{i,j-1}}^{t' \le |T|} X_{i,j-1,t'-\tau_{i,j-1},s} - \sum_{t'}^{t' \le |T|} X_{i,j,t',s}) \\ & + \sum_i rev_{i,PI}^{open} f_{i,PI} (1 - \sum_{t'}^{t' \le |T|} X_{i,PI,t',s}) \\ & + \sum_i \sum_{j \in \{PI,PII\}} \sum_{t > |T| - \tau_{i,j}} rev_{i,j,t}^{run} f_{i,j+1} X_{i,j,t,s} \quad \forall s \end{aligned} \tag{3.157}$$

$$Cst_s = \sum_{i,j,t} cd_t C_{i,j} X_{i,j,t,s} \quad \forall s \tag{3.158}$$

$$\sum_t X_{i,j,t,s} \le 1 \qquad \forall i, j, s \tag{3.159}$$

$$\sum_{t'}^{t} X_{i,j,t',s} \le \sum_{t'}^{t - \tau_{i,j-1}} X_{i,j-1,t',s} \qquad \forall i, j > 1, t', s \tag{3.160}$$

$$X_{i,j,t,s} = 0 \quad \forall (i,j,t,s) \in \boldsymbol{N} \tag{3.161}$$

$$\sum_{i}\sum_{j}\sum_{t'>t-\tau_{i,j}}^{t'\leq t} \rho_{i,j,r} X_{i,j,t',s} \leq {\rho_r}^{max} \quad \forall r,t,s \tag{3.162}$$

$$X_{i,j,1,1} = X_{i,j,1,s} \quad \forall i,j,s \tag{3.163}$$

$$\left[\neg \sum_{\substack{t'\leq p-\tau_{i^{s,s\prime},j^{s,s\prime}} \\ X_{i,j,t,s} = X_{i,j,t,s\prime}}} Y_{i^{s,s\prime},j^{s,s\prime},t\prime,s}\right] \vee \left[\sum_{t'\leq t-\tau_{i^{s,s\prime},j^{s,s\prime}}} Y_{i^{s,s\prime},j^{s,s\prime},t\prime,s}\right] \quad \forall i,j,t \geq 1, (s,s') \in \boldsymbol{B} \tag{3.164}$$

3.2.6 Computational Experiments, Results and Discussion

In this section, we apply three formulations to solve 42 different instances of the clinical trial planning problem. All models were implemented in Pyomo and solved using ILOG CPLEX Optimization Studio (Version 12.6.3.0, IBM, Armonk, NY, USA) on a standard node of Auburn University Hopper Cluster. The node has 20 cores with E5-2660 2.60 GHz processors and 128 GB of memory. We compare the sizes and solution times for CM1, CM2 and CM3.

3.2.6.1 Clinical Trial Plan for a Small Example

In this section, we present the details of the clinical trial plan generated for a small case study. The case study includes three potential drugs that must complete three clinical trials. The planning horizon is 36 months, and it is divided into 12 equal time periods. The remaining parameters of this case study can be found in Table C2 of Appendix C. The optimum ENPV is \$1189 M, and all three formulations (CM1, CM2 and CM3) yield the optimum in 7, 5 and 10 CPUs. **Figure 3.2.4** presents the optimum decision tree, i.e., the clinical trial plan for this case study.

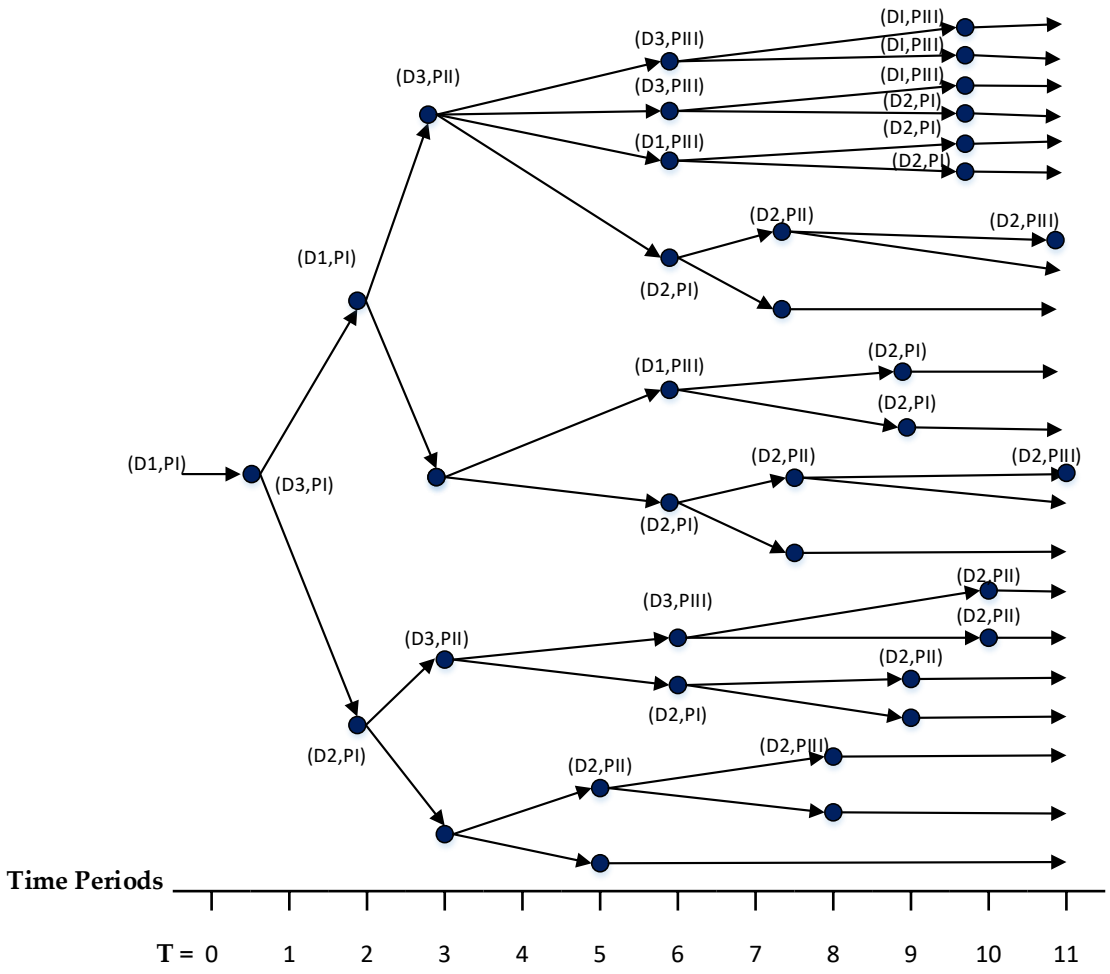

Figure 3.2.4: Multi-stage decisions in the optimal solution

At $t = 0$, all scenarios are indistinguishable, and hence, the solution recommends starting clinical trial PI of drug D1 for all scenarios. The duration of clinical trial PI of drug D1 is two time periods and at $t = 1$, its outcome is not realized. Therefore, all scenarios are still indistinguishable and the decisions to carry (drug, trail) pair (D3, PI) are recommended for all scenarios. At $t = 2$, the (D1, PI) is completed leading to two subsets of scenarios: (a) scenarios that (D1, PI) passes and scenarios that (D1, PI) fail. For scenarios in the group (a), the recommends starting clinical trial PII for drug D1, whereas for group (b), the plan recommends starting (D2, PI) in (b). At $t = 3$, when the outcome of (D3, PI) is realized for scenarios sets in (a) which passes (D1, PI) and

starts (D1, PII), the solution recommends waiting rather than starting the trial (D2, PI) immediately. This is due to the resource limitation. We cannot start the (D2, PII) if D1 is successful, which causes penalties of reduced active patent life for D2 being idle in the pipeline.

3.2.6.2 Clinical Trial Planning—Base Case Studies

We used five instances of the pharmaceutical clinical trial planning problem originally presented in (Christian and Cremaschi, 2017) as base case studies. These instances have two, three, four, five and six potential drugs that must complete clinical trials before they can be marketed for revenue. The parameters of these five base case studies can be found in Appendix C, Tables C1 – C5.

As expected, all three formulations yielded the same optimal ENPV and clinical trial plan for each instance. The optimal ENPVs for these case studies are given in the second column of Table 3.5. The remaining columns in Table 3.5 summarize the number of variables and constraints for each formulation. Table 3.5 reveals that CM1 has the fewest number of variables and constraints while CM2 has the most for each instance and that all models yielded very large MILPs for the six-drug case with more than one million variables and four million constraints. That is one of reasons that six-drug case required considerable computational effort, more than 30 CPU hours, to solve to 0.1% optimality gap. The change in computational times (in CPUs) with the number of drugs for each formulation is plotted using logarithmic scale in **Figure 3.2.5**. The solution times of CM3 were the longest and although the solution times of CM1 and CM2 were similar, those of CM2 were shorter for the five-drug and six-drug instances. For example, CM2 only took 8,114 CPUs to solve the six-drug case, while CM1 12,991 CPUs and CM3 108,746 CPUs. Although CM2 has the most variables and constraints for all instances (Table 3.5), CM2 takes shorter to solve than CM3 in each instance. It is worth noting that the model generation time of Pyomo for

all cases were relatively small and are negligible compared to the solution times. For example, for the six-drug case, the solution time for CM3 was 108,746 CPUs while the generation time was only 18 CPUs.

Table 3.5: Problem sizes for five instances of CM1, CM2 and CM3

Cases	ENPV ($M)	CM1		CM2		CM3	
		Variables	Constraints	Variables	Constraints	Variables	Constraints
2-Drug	1104	586	952	838	1492	613	1051
3-Drug	1189	17,281	44,065	24,193	64,801	21,249	49,761
4-Drug	1697	48,385	138,497	66,817	193,793	57,345	150,529
5-Drug	2082	239,617	802,305	454,657	1,078,785	284,673	862,721
6-Drug	2450	1,142,785	4,386,817	1,585,153	5,713,921	1,359,873	4,677,633

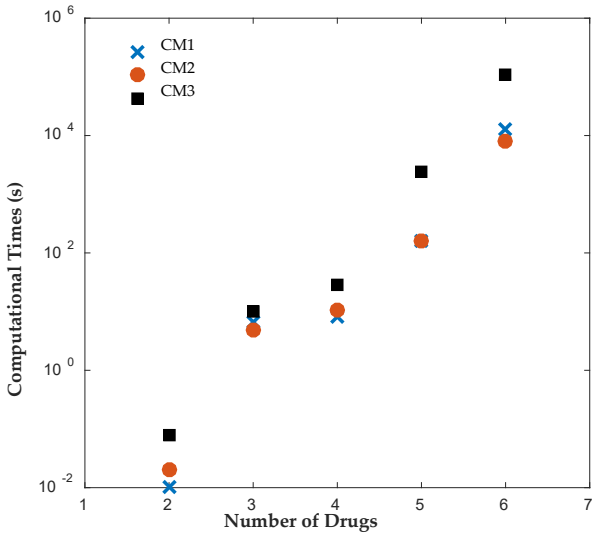

Figure 3.2.5: Change in solution times of CM1, CM2 and CM3 with number of drugs

3.2.6.3 Sensitivity of Solution Times to Problem Parameters and Sizes

In this section, we investigate the sensitivity of the solution times of all formulations to the changes in the problem parameters and size. For these studies, we used the five-drug problem from the previous section as the base case. This problem has five candidate drugs, three clinical trials, a

six-period planning horizon and two limiting resources. The parameters of the base case five-drug problem are given in Table C4 of Appendix C.

3.2.6.3.1 Sensitivity of Solution Times to Problem Parameters

The parameters of the clinical trial planning problem are resource availability, clinical trial costs, maximum revenues, patent-life-loss penalty, market-share-loss penalty and clinical trial durations (Table 3.6). To study the impact of overall resource availability, we construct four problems with varying degrees of overall resource availabilities similar to (Christian and Cremaschi, 2017): (1) fully constrained; (2) 40 percent unconstrained; (3) 70 percent unconstrained; and (4) unconstrained. The base case problem is assumed to be fully constrained. The unconstrained case provides enough resources to allow all drugs to complete all clinical trials simultaneously without any delay. The 40 percent-unconstrained case is generated by increasing the available resource of the fully constrained case by 40 percent. In the 70 percent unconstrained case, this increase is 70%. The values of trial cost, max revenue of each drug, active patent-life-loss penalty and market-share-loss penalty are all perturbed by ±10% and ±25%. To investigate the sensitivity of clinical trial durations, we extend the length of all clinical trials by one and two time periods.

Table 3.6: Descriptions of case studies used to study the impact of problem parameters on solution times of CM1, CM2 and CM3

Variation	Case Name	Variation Value
Resource availability	5-Drug	-
	5-Drug-RA40%	+40%
	5-Drug-RA70%	+70%
	5-Drug-RA100%	+100%
Trial cost	5-Drug-TC75%	−25%
	5-Drug-TC90%	−10%
	5-Drug-TC110%	+10%
	5-Drug-TC125%	25%

Maximum revenue	5-Drug-RV75%	−25%
	5-Drug-RV90%	−10%
	5-Drug-RV110%	+10%
	5-Drug-RV125%	+25%
Active patent life loss penalty	5-Drug-GL75%	−25%
	5-Drug-GL90%	−10%
	5-Drug-GL110%	+10%
	5-Drug-GL125%	+25%
Market share loss penalty	5-Drug-GD75%	−25%
	5-Drug-GD90%	−10%
	5-Drug-GD110%	+10%
	5-Drug-GD125%	+25%
Trial Duration	5-Drug-TL1	+1
	5-Drug-TL2	+2

Table 3.6 summarizes the five-drug clinical trial planning problem instances generated using the explained perturbations. A plus sign (+) in Variation Value column refers to an increase in magnitude and a minus sign (−) refers to a decrease. The number next to the sign indicates the magnitude of the increase/decrease. The case names provide descriptive information. For example, case name 5-Drug-RA70% corresponds to five-drug base-case problem with 70 percent unconstrained resources and 5-Drug-RA100% refers to the five-drug case without any resource constraints. Similarly, case name 5-Drug-RV75% corresponds to the five-drug base case problem whose maximum revenue from each drug are decreased by 25% yielding 75% of maximum revenue for each drug. In Table 3.6, resource availability, trial cost, max revenue of each drug, active patent-life-loss penalty, market-share-loss penalty and clinical trial duration are labeled as RA, TC, RV, GL, GD and TL, respectively. All changes in parameter values except resources availability did not significantly impact solution times, and their results are summarized in Table A6.

The resource availability had a significant impact on solution times. Thus, we extended our sensitivity analysis and included three-, four- and six-drug instances. The results are summarized in Table 3.7 and they reveal that the solution times increase as the problems become more resource

constrained. Similar to the results for the base case studies, the solution times of CM1 and CM2 are up to two orders of magnitude shorter than that of CM3. The difference in solution times becomes more pronounced for instances that are larger and more tightly resource-constrained. The fully unconstrained resource case is equivalent to removing all resource constraints, which will allow completing clinical trials of all potential drugs simultaneously yielding a relaxed problem. Therefore, the fully unconstrained cases take a few CPUs to solve. It is worth noting that the solutions and ENPVs of relaxed problems with fully unconstrained resources (such as 3-Drug-R100%, 4-Drug-R100%, 5-Drug-R100% and 6-Drug-R100%) are valid upper bounds for all cases with resource constraints. As the problem becomes more resource-constrained, it becomes more difficult to solve requiring longer solution times for all three formulations. As an example, for the six-drug problem, the solution times of CM1, CM2 and CM3 are more than two orders of magnitude longer for the original problem (6-Drug in Table 3.7) than the fully unconstrained case (6-Drug-R100%). Therefore, for significantly large cases such as the six-drug one, which requires more than 22 CPU hours to solve to 0.1% optimality gap using model CM3, if the resource constraints were relaxed, valid upper bounds within 3% can be generated very quickly, e.g. within one CPU hour for the six-drug case.

Table 3.7: Resource constraint perturbation results for CM1, CM2 and CM3

Cases	ENPV	Solution Time (CPUs)		
		CM1	CM2	CM3
3-Drug	1189	7	5	10
3-Drug-R40%	1211	5	4	18
3-Drug-R70%	1220	2	2	5
3-Drug-R100%	1221	2	2	4
4-Drug	1697	8	11	28
4-Drug-R40%	1719	3	4	8
4-Drug-R70%	1721	3	4	6
4-Drug-R100%	1721	3	4	5
5-Drug	2083	161	158	2372

5-Drug-R40%	2123	37	43	324
5-Drug-R70%	2128	25	32	153
5-Drug-R100%	2128	59	72	318
6-Drug	2450	12,991	8114	108,746
6-Drug-R40%	2510	649	688.5	9786
6-Drug-R70%	2517	297	334	3629
6-Drug-R100%	2517	282	3234	3348

3.2.6.3.2 Sensitivity of Solution Times to Problem Sizes

Varying the number of trials, the number of resources and the length of planning horizon changes the size of the resulting problems for all formulations. The number of trials and the length of planning horizon change the number of variables and constraints. The number of resources change the number of constraints. For example, by doubling the length of planning horizon, all binary variables will be doubled and so will the related constraints.

We use the five-drug base case problem to study the impact of these parameters on the resulting problem sizes and solution times. Table 3.8 summarizes the considered variations, i.e. the parameter names along with the magnitude of the changes made to generate the case studies. In Table 3.8, each case study is associated with a unique case number for easy identification in graphs and tables. Case 9 in Table 3.8 refers to the original five-drug base case problem. A plus/minus sign (+/−) in Magnitude of Change column of Table 3.8 refers to an increase/decrease in magnitude. The number next to the sign indicates the magnitude of the increase/decrease.

Table 3.8: Size perturbation cases and file names with variation value

Case Number	Variation	Magnitude of Change
1	Number of trials	+1
2		+2
3	Number of resources	+1
4		+2
5		+3
6		+4
7	Length of planning horizon	−2
8	Original 5-Drug Case	+2

The solution times of CM1, CM2 and CM3 are plotted for all cases of Table 3.8 in Figure 3.2.6a. Similar to the results of previous section, the solution times of CM1 and CM2 are consistently shorter than CM3 for all cases with different problem sizes. The root relaxation solution times of CM1, CM2 and CM3 (Figure 3.2.6b) reveals that both CM1 and CM2 yield the root relaxation solution considerably faster than CM3 in all cases. For example, in Case 8, ILOG CPLEX Optimization Studio (Version 12.6.3.0, IBM, Armonk, NY, USA) took 843, 896 and 4059 CPUs to solve the root node relaxation problems for CM1, CM2 and CM3, respectively. The comparison of initial gaps of CM1, CM2 and CM3 are shown in Figure 3.2.6c. The CM1 and CM2 yield much tighter initial gaps than CM3. The comparison of number of branches required to solve the problems to 0.1% optimality gap (Figure 3.2.6d) reveals that most cases (1, 2, 3, 6 and 7) are solved in root node. For cases requiring branching, CM1 and CM2 required considerably fewer nodes than CM3. For example, in Case 8, CM1, CM2 and CM3 require 2085, 1031 and 7184 nodes. These results indicate that CM1 and CM2 require shorter times for branching than CM3, which in turn results in shorter solution times.

From the parameters that change the size of the problem, the planning horizon has the most significant impact on the problem sizes and solution times (Figure 3.2.6, Cases 7 and 8). For example, the solution time for Case 7, which has five candidate drugs, three clinical trials, a four-period planning horizon and two limiting resources, are only 22 CPUs for CM3. After extending the planning horizon to eight time periods, the solution time increased to 785,530 CPUs, a four orders of magnitude increase with only extending the planning horizon by four time periods. For discrete-time MSSPs, the problem size increases rapidly especially in numbers of variables and

NACs with increases in the length of planning horizon and this translates into longer solution times.

The number of trials has relatively small influence on the performance of models (Cases 1 and 2). The number of resources had the least impact as varying the number of resources did not influence the initial gaps and had very little impact on solution times (Cases 3-6 in Figure 3.2.6).

To investigate the strength of the relationship between solution times (in CPUs) and root relaxation solution times (in CPUs), initial gap and number of branches, we plotted solution times against these variables and calculated the corresponding Pearson correlation coefficient. The solution times (s) versus root relaxation solution times (s) are plotted in Figure 3.2.7. The corresponding Pearson correlation coefficients are 1.0, 1.0 and 0.8 for CM1, CM2 and CM3. The values of the correlation coefficient reveal relatively strong positive relationship between solution times and root relaxation solution times. Similar plots for the initial gap and number of branches are given in Figures 7 and 8, respectively. The correlation coefficients for initial gap are 0.8, 0.8 and 0.7 for CM1, CM2 and CM3, a weaker relationship than the root relaxation solution times. The correlation coefficients between solution times and number of branches are 1.0, 1.0 and 1.0 for CM1, CM2 and CM3. Note that the correlation coefficient calculations for number of branches ignore the cases where the solution was obtained at the root node. All correlation coefficients are positive, which indicates (as expected) an increase in root relaxation solution time, initial gap or number of branches increases the solution times. The correlation coefficients for root relaxation solution times and number of branches are larger than the ones for initial gap suggesting a stronger positive correlation with the solution time for these variables.

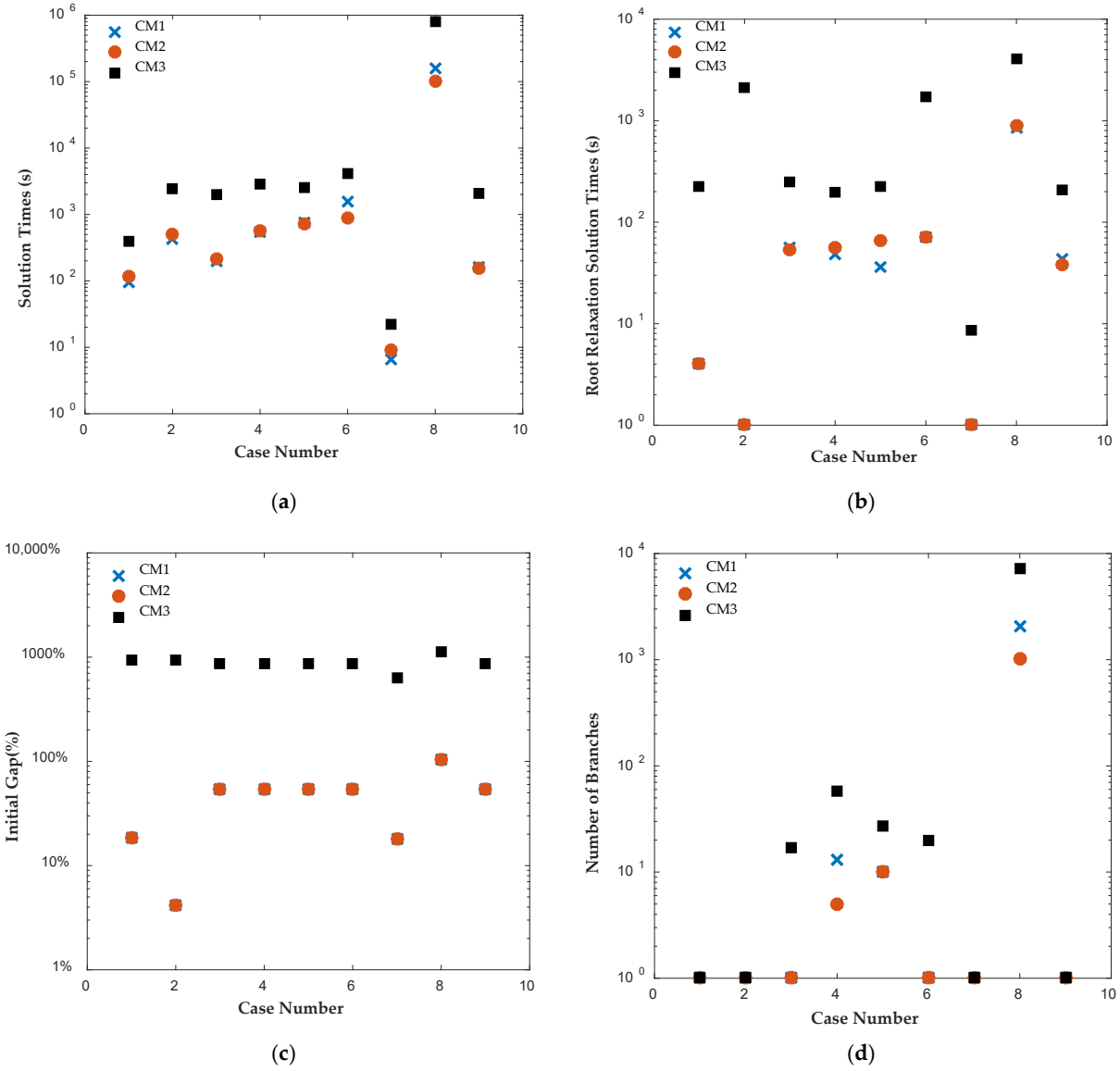

Figure 3.2.6: Comparison of (a) solution times, (b) root relaxation solution times, (c) initial gaps, and (d) number of branches for CM1, CM2 and CM3

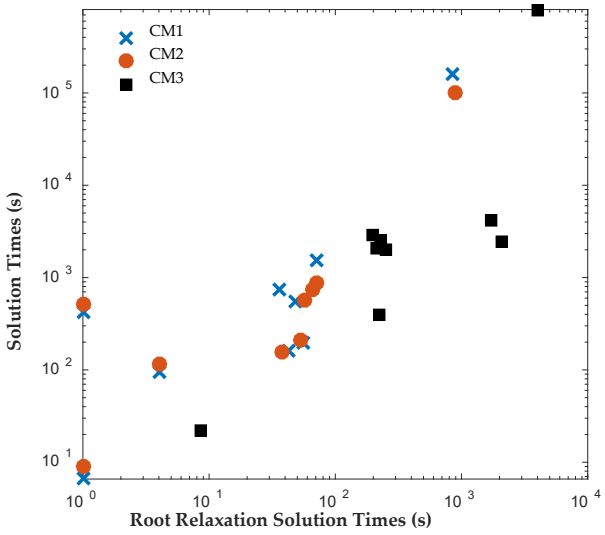

Figure 3.2.7: Change in solution times with root relaxation solution time for CM1, CM2 and CM3. The strong linear correlation is evident from this plot.

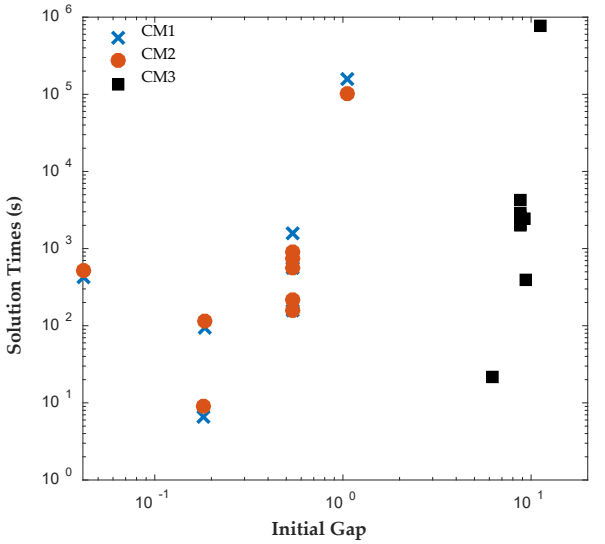

Figure 3.2.8: Change in solution times with changes in initial gaps of CM1, CM2 and CM3. Initial gaps of CM1 and CM2 are consistently smaller than CM3.

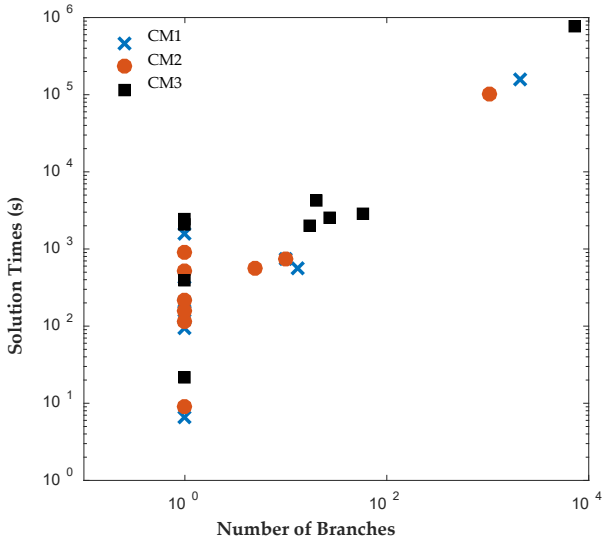

Figure 3.2.9: Change in solution times with the number of branches needed to solve CM1, CM2 and CM3. Please note most problems are solved at the root node, and when branching is required to solve the problem, the linear dependency of solution time to number of branches required to solve the problem is evident from this figure.

By comparing three models of all case studies, the results reveal that both CM1 and CM2 perform better than CM3. The proposed binary variables in CM1 and CM2 contribute to shorter root relaxation solution times and generate tighter initial gaps. In addition, most problems with all formulations were solved at the root node. For instances where ILOG CPLEX Optimization Studio (Version 12.6.3.0, IBM, Armonk, NY, USA) branched, CM1 and CM2 consistently required fewer branches than CM3, which suggests that the binary variables in CM1 and CM2 provided ILOG CPLEX Optimization Studio (Version 12.6.3.0, IBM, Armonk, NY, USA) with more efficient branching variables. The correlations of root relaxation solution time, initial gap and number of branches with solution times revealed that all three have strong positive relationships to solution times. When the root relaxation solution time, initial gap and number of branches of a problem instance increase, so does the solution time. For all formulations, we also found that the root

relaxation solution time and number of branches have stronger impact on solution time than initial gap.

3.2.7 Conclusion and Future

This section presented two new MSSP formulations, CM1 and CM2, for pharmaceutical clinical trial planning problems. The first formulation, CM1, uses two binary variables to track the beginning and the end of clinical trials. The second formulation, CM2, introduces an additional binary variable that tracks both the beginning and the end of clinical trials. We compare the sizes and solution times of CM1 and CM2 to each other and to CM3, an MSSP presented in (Colvin and Maravelias, 2008) for clinical trial planning problem, for different instances.

The results reveal that both CM1 and CM2 provide tighter formulations than CM3 and are solved faster by ILOG CPLEX Optimization Studio (Version 12.6.3.0, IBM, Armonk, NY, USA). The proposed binary variables in CM1 and CM2 contribute to shorter root relaxation solution times and generate much tighter initial gaps. It is worth noting that for all formulations, most problems were solved at the root node. For instances with branches, CM1 and CM2 consistently required fewer branches than CM3, which suggest that the binary variables in CM1 and CM2 provided ILOG CPLEX Optimization Studio (Version 12.6.3.0, IBM, Armonk, NY, USA) a more efficient branching variable. The correlation coefficients between the solution time and root relaxation solution time, initial gap and number of branches, respectively, revealed that all three have strong positive relationship to solution times. When the root relaxation solution time, initial gap and number of branches of a case increase, so does its solution time. For all formulations, the correlations suggest that the root relaxation solution time and number of branches have stronger impact on solution time than the initial gap. We also investigated the sensitivity of solution times and problem sizes to the parameters of the clinical trial planning problem. The results revealed that

the resource constraints and length of planning horizon contribute most to the computational complexity of this problem. It is worth noting that ILOG CPLEX Optimization Studio (Version 12.6.3.0, IBM, Armonk, NY, USA) was not able to solve any of the formulations for a seven-drug case study. The future work will focus on developing decomposition approaches to generate valid upper and lower bounds quickly to solve large instances of the clinical trial planning problem.

The outcomes of clinical trials of different drugs are assumed to be independent. This assumption may not be valid when the pipeline contains drugs that are being developed for treating similar conditions or that have similar formulations and characteristics. In such cases, the outcomes of the events for these drugs may not be independent, i.e. may be correlated. Then, the decision to start any clinical trial of these drugs will not only affect the timing of realizations of their endogenous uncertain parameters (clinical trial outcomes), but also affect the probability distributions of these uncertain parameters. For instance, let's assume that drugs D1 and D2 have similar formulations and their outcomes are correlated and that the solution recommends starting drug trial pair (D1, PI) at period one. When the outcome of (D1, PI) is realized, the probability distribution of outcome space for D2 would be different if drug D1 passes PI or fails it. This will make the resulting problem a MSSP with both Type I and Type II endogenous uncertainty, which is a more difficult formulation to solve with limited approaches available to address it. We left this as future study.

CHAPTER 4

GENERALIZED KNAPSACK-PROBLEM BASED DECOMPOSITION ALGORITHM (GKDA)

In this chapter, we present a generalized Knapsack-problem based Decomposition Algorithm (GKDA) to obtain feasible solutions for MSSP models with endogenous and/or exogenous uncertainties. Instead of enumerating all uncertain parameter space as scenarios, the GKDA decomposes the original multi-period MSSP into a series of knapsack problems and solves these problems at appropriate decision points of the planning horizon. We applied GKDA to obtain feasible solutions for four planning problems, which include continuous and/or discrete decision variables, and endogenous and/or exogenous uncertain parameters.

4.1 The general multistage stochastic programming formulation with endogenous and exogenous uncertainties

A general formulation of the MSSP under endogenous and exogenous uncertainties, MSSP$_{EX}$, is given in Equations (4.1) – (4.2). The presented model MSSP$_{EX}$ is adapted from the original model of (Goel and Grossman, 2006). The planning horizon for this multistage stochastic program is defined as $t \in \{1, 2, 3, .., |T|\}$. The uncertain parameter θ_i is associated with the endogenous uncertainty per source $i \epsilon I$. The model has binary state variables denoted by $b_{i,t}^s$, continuous state variables represented as $y_{i,t}^s$, and recourse-action decision variables defined as γ_t^s. The realizable values of uncertain parameter θ_i are represented using a finite set $\{\theta_i^1, \theta_i^2, \dots, \theta_i^{R_i}\}$. The exogenous uncertain parameter realized at time t is denoted by ξ_t and it has a finite set of possible realizations $\{\xi_t^1, \xi_t^2, \dots, \xi_t^R\}$. The full scenario set, $\boldsymbol{S}$, is constructed using Cartesian product and it is defined as $\boldsymbol{S} := \{\times_{t \in T}\, \xi_t\} \times \{\times_{i \in I}\, \theta_i\}$. We define an exogenous scenario pair set $(t, s, s') \in S_X$ for scenarios s and s' that are indistinguishable at time period t in terms of the realization of the exogenous

uncertain parameter, i.e., $S_X \coloneqq \{(t,s,s'): \xi_t^s = \xi_t^{s'}, t \in T\}$, and an endogenous scenario pair set $(s,s') \in S_E$ for scenarios s and s' that are indistinguishable in terms of the realization of the endogenous uncertain parameter, i.e., $S_E \coloneqq \{(s,s'): \theta_i^s = \theta_i^{s'}, i \in I\}$.

MSSP$_{EX}$: MSSP formulation with endogenous and exogenous uncertainty

$$\max \sum_s p_s G_s\left(b_{i,t}^s, y_{i,t}^s, \gamma_t^s, \theta_i^s, \xi_t^s\right) \tag{4.1}$$

$$g_{i,t,s}\left(b_{i,t}^s, y_{i,t}^s, \gamma_t^s, \theta_i^s, \xi_t^s\right) \le 0 \quad \forall i \in I, t \in T, s \in \boldsymbol{S} \tag{4.2}$$

$$h_{i,t,s}\left(b_{i,t}^s, y_{i,t}^s, \gamma_t^s, \theta_i^s, \xi_t^s\right) = 0 \quad \forall i \in I, t \in T, s \in \boldsymbol{S} \tag{4.3}$$

$$b_{i,1}^s = b_{i,1}^{s'} \qquad \forall i \in I, s, s' \in \boldsymbol{S} \tag{4.4}$$

$$y_{i,1}^s = y_{i,1}^{s'} \qquad \forall i \in I, s, s' \in \boldsymbol{S} \tag{4.5}$$

$$b_{i,t}^s = b_{i,t}^{s'} \qquad \forall i \in I, t \in T, (t,s,s') \in \boldsymbol{S_X} \tag{4.6}$$

$$y_{i,t}^s = y_{i,t}^{s'} \qquad \forall i \in I, t \in T, (t,s,s') \in \boldsymbol{S_X} \tag{4.7}$$

$$\gamma_t^s = \gamma_t^{s'} \qquad \forall t \in T, (t,s,s') \in \boldsymbol{S_X} \tag{4.8}$$

$$\begin{bmatrix} Z_t^{s,s'} \\ b_{i,t}^s = b_{i,t}^{s'} \\ y_{i,t}^s = y_{i,t}^{s'} \\ \gamma_t^s = \gamma_t^{s'} \end{bmatrix} \vee \left[\neg Z_t^{s,s'}\right] \qquad \forall (s,s') \in \boldsymbol{S_E}, \forall t > 1 \tag{4.9}$$

$$Z_t^{s,s'} \Leftrightarrow H\left(b_{i,1}^s, b_{i,2}^s, \dots, b_{i,t}^s, y_{i,1}^s, y_{i,2}^s, \dots, y_{i,t}^s\right) \qquad \forall (s,s') \in \boldsymbol{S_E}, \forall t > 1 \tag{4.10}$$

$$b_{i,1}^s \in \{0,1\} \quad \forall i \in I, s \in \boldsymbol{S} \tag{4.11}$$

$$Z_t^{s,s'} \in \{0,1\} \quad \forall s, s' \in \boldsymbol{S}, \forall t \in T \tag{4.12}$$

$$y_{i,t}^{s}, \gamma_{t}^{s} \in \mathbb{R} \quad \forall s \in \boldsymbol{S}, \forall t \in T \tag{4.13}$$

The MSSP$_{EX}$ consists of five parts:

(1) The objective function, (4.1), minimizes the expected value of the total cost associated with stage decision variables $b_{i,t}^{s}$, $y_{i,t}^{s}$, and recourse-action decision variables γ_{t}^{s}, where p_s is the probability of scenario s and the function $G_s\left(b_{i,t}^{s}, y_{i,t}^{s}, \gamma_{t}^{s}, \theta_{i}^{s}, \xi_{t}^{s}\right)$ calculates the total cost for scenario s.

(2) Scenario specific constraints are represented via (4.2) for inequality and via (4.3) for equality constraints. These constraints may be used to model resource availabilities, sequencing constraints, material flow constraints, and capacity constraints for each scenario and time period.

(3) Initial non-anticipativity constraints (NACs) given in Eqns. (4.4) - (4.5) ensure that the values of the decision variables associated with both endogenous and exogenous uncertain parameters are identical at the first time period.

(4) The NACs for exogenous uncertain parameters are presented in Eqns. (4.6)-(4.8). They are written for each time period for which scenario pairs (s, s') remain indistinguishable due to exogenous uncertainty. The corresponding sets of indistinguishable exogenous scenarios pairs is given by the set $\mathbf{S_X}$.

(5) Conditional NACs, disjunction in (4.9), are added to the MSSP formulation for preventing anticipative decisions before two scenarios become distinguishable due to a realization of an endogenous uncertain parameter. The value of the indicator variable, $Z_{t}^{s,s'}$, in the disjunction is determined by Equation (4.10). If $Z_{t}^{s,s'}$ is True, the scenarios pairs $(s, s') \in S_E$ remain indistinguishable, and hence, the NACs for these scenarios are enforced.

In MSSP$_{EX}$, if functions $G_s(.)$, $g_{i,t,s}(.)$ and $h_{i,t,s}(.)$ are linear, the deterministic equivalent of MSSP$_{EX}$ can be formulated as a mixed-integer linear programing (MILP) model; otherwise the deterministic equivalent is a mixed-integer nonlinear programing (MINLP) model.

4.2 The generalized knapsack-problem based decomposition algorithm

The generalized Knapsack-problem based Decomposition Algorithm (GKDA) decomposes time periods and scenarios of a MSSP model into a series of continuous or 0-1 Knapsack Sub-Problems (KSPs). The algorithm generates and solves these problems based on the realizations of uncertainty along the planning horizon. The general form of each individual KSP is given in Eqns. (4.14)-(4.17).

KSP: *General form of an individual knapsack sub-problem*

$$max: -G^{est}(b_i, y_i, \gamma, E[\theta_i], E[\xi]) \tag{4.14}$$

$$g_i(b_i, y_i, \gamma, E[\theta_i], E[\xi]) \leq 0 \quad \forall i \in E \tag{4.15}$$

$$h_i(b_i, y_i, \gamma, E[\theta_i], E[\xi]) = 0 \quad \forall i \in E \tag{4.16}$$

$$b_i \in \{0, 1\}, y_i, \gamma \epsilon R, \quad \forall i \in E \tag{4.17}$$

The objective function and scenario specific constraints ((4.1)-(4.3)) of MSSP$_{EX}$ are used to construct the objective function and constraints of the KSP (Eqns. (4.14)-(4.17)). The objective function of the KSP maximizes the total value of the packed items. The decision variables ($b_{i,t}^s$, $y_{i,t}^s$, and γ_t^s) of MSSP$_{EX}$ are represented by indivisible items (b_i), which has a binary decision variable, and divisible items (y_i and γ), which are continuous decision variables in KSP. The GKDA uses the concept of eligible item set E, where items $e \in E$ are created by enumerating all allowable decisions at time period t in MSSP$_{EX}$. The allowable decisions at time period t are determined

using the state of the system at time period t, the NACs, the logic of the sequencing constraints (if present in MSSP$_{EX}$) and the bounds (if present in MSSP$_{EX}$). The state of the system at time period t is defined using the decision variable values and the uncertainty realizations up to time period $(t-1)$.

The objective function $\sum_s p_s G_s\left(b_{i,t}^s, y_{i,t}^s, \gamma_t^s, \theta_i^s, \xi_t^s\right)$ (Eq. (4.15)) of MSSP$_{EX}$ is estimated in each KSP by $-G^{est}(b_i, y_i, \gamma, E[\theta_i], E[\xi])$ via taking the expected values of uncertain parameters θ_i^s and ξ_t^s as $E[\theta_i]$ and $E[\xi]$. Generally, we directly transfer the coefficients of objective function associated with item $e \in E$ as item values, V_e, in each KSPs. There are two conditions, for which this approach does not accurately approximate the value of the eligible items in KSPs. We introduce two approaches to improve item values, which are discussed in next Section.

The inequality constraints $g_{i,t,s}$ and equality constraints $h_{i,t,s}$ in Eqns. (4.2) – (4.3) of MSSP$_{EX}$ are directly transformed to g_i and h_i, i.e., weight constraints, in KSPs by removing both t and s indices. The coefficients of the decision variables in the equality and inequality constraints (Eqns. (4.2) – (4.3)) of MSSP$_{EX}$ define the item weights W_e, $e \in E$. It should be noted that the equations corresponding to sequencing constraints and bounds that were used for determining the allowable items are not included in weight constraints of the KSPs.

The GKDA is outlined in Figure 4.2.1, where the n^{th} KSP at time t is denoted by KSP$_{n,t}$. In the first KSP generated, both exogenous and endogenous uncertain parameter values are represented by their respective expected values ($E[\theta_i]$ and $E[\xi]$) as there has been no uncertainty realizations. The GKDA constructs a single ($n = 1$) KSP at the root node ($t = 1$), KSP$_{1,1}$,. The eligible item set (E) for KSP$_{1,1}$ is constructed by enforcing the initial time period NACs on decision variables of MSSP$_{EX}$ at $t = 1$. The corresponding item values and weights are estimated using the

appropriate coefficients and the expected values of the uncertain parameters. Then, the GKDA sets N_t, which is equal to the number of KSPs that should be constructed at time t, equal to one.

The optimal solution of root-node $\text{KSP}_{1,1}$ is used to determine the values of the decision variables at the first-time period of the planning horizon. The stage decision variables are set equal to their optimal solution values in $\text{KSP}_{1,1}$. The endogenous uncertainty associated with stage decision variables and exogenous uncertainty that are resolved at the first-time period are realized. If the scenario specific constraints of MSSP_{EX} ($g_{i,t,s}(.)$ and $h_{i,t,s}(.)$) do not contain any uncertain parameters, the values of the recourse-action decision variables determined based on the optimum solution of the KSP are guaranteed feasible. For those problems, the recourse-action decision variables are set equal to their optimal solution values in $\text{KSP}_{1,1}$. On the other hand, if the constraints $g_{i,t,s}(.)$ and $h_{i,t,s}(.)$ include uncertain parameters, the values of the recourse-action decision variables are calculated using the stage decision variable values determined by the optimum solution of the KSP for each unique realization of uncertainty.

$t := 1$

calculate $\{V_i, W_i, E\}$ for root-node KSP

construct $\text{KSP}_{1,t}$

$N_t := 1$

repeat

$n := 1$

for $n \leq N_t$ **do**

$B_{n,t}$ = solution ($KSP_{n,t}$)

calculate recourse actions γ_t based on the realization of θ_i or ξ in $B_{n,t}$

calculate N_{t+1} based on the realization of θ_i or ξ in $B_{n,t}$

update $\{V_i, W_i, E\}_{n,t+1}$

construct $n \in N_{t+1}$ $KSP_{n,t+1}$ at time t+1 based on the realization of θ_i or ξ in $B_{n,t}$

n = n + 1

$t = t + 1$

until $E = \emptyset$ or $t = |T|$ (the end of planning horizon)

Figure 4.2.1: A general algorithm for GKDA

The new child KSPs for the next time period are generated based on the realized uncertainties, i.e. a new KSP is generated for each realization. The number of new KSPs generated (N_{t+1}) for the next time period depends on the number of realizations. For example, suppose that the optimal solution of $\text{KSP}_{1,1}$ set a binary stage decision variable associated with an endogenous uncertain parameter equal to one, and further suppose that the endogenous uncertain parameter has two outcomes (H and L), which will be realized when binary stage decision variable becomes one. For

sake of simplicity, assume that there are no realizations for the exogenous uncertain parameters at $t = 1$. The GKDA then generates two new child KSPs, $\text{KSP}_{1,2}$, $\text{KSP}_{2,2}$, estimates and updates item eligible items set E, item values V_i and weights W_i based on the realizations, which are H and L for the example. For endogenous uncertainties, the GKDA generates child KSPs when the outcomes of selected items realized. For exogenous uncertain parameters, the GKDA generates child KSPs at their realization times, which are known a priori.

After the construction of child KSPs, the GKDA increments the time, and repeats the process of generating child KSPs, solving the KSPs, realizing the uncertainties, and determining recourse actions. The process continues until either the eligible item sets are empty or the end of the planning horizon is reached.

The GKDA is guaranteed to yield a feasible solution as it satisfies all constraints of the original MSSP_{EX}. The KSPs combined with recourse-action determination scheme ensures that scenario specific constraints are satisfied. The NACs of the MSSP_{EX} are satisfied by how the child KSPs are generated. Therefore, at termination, the GKDA provides a feasible solution and a primal bound for the MSSP_{EX}. The quality of the GKDA solution depends on how accurate item values estimate the impact of the decisions associated with these items on the MSSP_{EX} objective function value. Next section discusses how item values are calculated.

4.2.1 Calculation of item values in knapsack sub-problems

An item value in a KSP approximates the contribution of the decision variable associated with that item to the objective function value of MSSP_{EX} at the stage and time period the KSP was generated. The straightforward approach for calculating an item value is to use the MSSP_{EX} objective function coefficients of the decision variable associated with the item. Suppose the objective function of MSSP_{EX} is given in Eq. (4.18).

$$min \sum_{s} p_s \sum_{i} \sum_{t} \left[(V_{1,i,t}+V^s_{1,i,t}) b^s_{i,t} + (V_{2,i,t}+V^s_{2,i,t}) y^s_{i,t} + (V_{3,t}+V^s_{3,t}) \gamma^s_t \right] \quad (4.18)$$

In Eq. , the decisions variables are $(b^s_{i,t}, y^s_{i,t}, \gamma^s_t)$, and there are both deterministic $(V_{1,i,t}, V_{2,i,t}, V_{3,t})$ and uncertain $(V^s_{1,i,t}, V^s_{2,i,t}, V^s_{3,t})$ coefficients. Using the straightforward approach, the objective function of a KSP generated at time period t is given by Eq. (4.19).

$$min \sum_{i} \left(V_{1,i,t} + E\left[V^s_{1,i,t}|D\right]\right) b_i + \left(V_{2,i,t} + E\left[V^s_{2,i,t}|D\right]\right) y_i + (V_{3,t} + E\left[V^s_{3,t}|D\right]) \gamma \quad (4.19)$$

The deterministic coefficients at time period t $(V_{1,i,t}, V_{2,i,t}, V_{3,t})$ contribute directly to the item values. To estimate the contributions of the uncertain coefficients $(V^s_{1,i,t}, V^s_{2,i,t}, V^s_{3,t})$, we calculate their conditional expectations at time period t where $D = \left\{b^s_{i,u}, y^s_{i,u}, \gamma^s_u, u \leq t-1\right\}$. In other words, the expected values are calculated based on the decision variable values up to the previous time period and associated endogenous and exogenous uncertain parameter realizations.

Under certain conditions, the straightforward approach for generating item values may not accurately approximate the expected contribution of some decision variables to the $MSSP_{EX}$ objective function value in KSPs. In such cases, the GKDA may yield a poor feasible solution. We define two such conditions and propose different approaches to approximate their item values.

CONDITION 1

IF *gains/revenues are only realized at certain stages or are only associated with certain decision variables,* ***THEN*** *generate the item value using the maximum potential gain approach (MPGA).*

CONDITION 2

IF *an item represents a capital investment decision variable in* $MSSP_{EX}$, ***THEN*** *generate the item value using annual capital charge.*

CONDITION 1 considers the class of planning problems where the decision maker cannot claim revenue or gains until a certain stage or until a certain outcome is realized in the planning

horizon. For example, in financial planning and control problem (Asset-liability management, Birge and Louveaux, 2011) the decision maker sequentially makes investment decisions at every stage, however the gains are only associated with last state variables. This problem satisfied CONDITION 1, where all investment decisions implicitly contribute to the overall gain. The clinical trial planning problem also satisfied CONDITION 1, where the revenue associated with a drug is only realized once the drug successfully completes the last (i.e., third) clinical trial. With the maximum potential gain approach (MPGA), we calculate an item's value based on the maximum expected contribution of the associated decision to the MSSP objective function value considering the remaining planning horizon. Suppose the objective function of MSSP$_{EX}$ is given as Eq. (4.20).

$$min \sum_{s} p_s \sum_{i} \left\{ -V_{1,i,|T|} b^s_{i,|\mathrm{T}|} - V_{2,i,|T|} y^s_{i,|\mathrm{T}|} - V_{3,|T|} \gamma^s_{|T|} \right\} \tag{4.20}$$

In Eq. (4.20), the decisions variables are $\left(b^s_{i,t}, y^s_{i,t}, \gamma^s_t\right)$, and only at end of planning horizon, the revenue is realized based on the decision variable values of $\left(b^s_{i,|\mathrm{T}|}, y^s_{i,|\mathrm{T}|}, \gamma^s_{|\mathrm{T}|}\right)$. Suppose that there are sequencing constraints in the scenario-specific constraint set that restrict the values of $\left(b^s_{i,|\mathrm{T}|}, y^s_{i,|\mathrm{T}|}, \gamma^s_{|\mathrm{T}|}\right)$ on the values of previous decision variables $\left\{b^s_{i,t}, y^s_{i,t}, \gamma^s_t | t \leq |T| - 1\right\}$. Then, this problem satisfies CONDITION 1. The objective function of a KSP generated at time period t for this problem using the MPGA is given in Eq. (4.21), and the item values are calculated based on the maximum expected contribution of each item to the objective function value.

$$max \sum_{i} E\left[V_{1,i,|T|}\right] b_i + \mathrm{E}\left[V_{2,i,|T|}\right] y_i + \mathrm{E}\left[V_{3,|T|}\right] \gamma \tag{4.21}$$

In CONDITION 2, the problem involves capital investment decisions with revenue streams associated with these investments in the future. For example, in the planning of offshore gas field developments, the decision to invest in a dedicated well platform at a field for producing gas can

be made at any time period, and the revenue associated with the produced gas will contribute to the overall profit for the rest of the planning horizon. Similar to CONDITION 1, due to the decomposition of time periods in GKDA, the KSPs cannot consider the impact of future revenue streams associated with the capital investment decisions. For these problems, the item costs (or negative values) associated with capital investment decision variables are calculated by converting the capital investment to the annual capital charge for the remaining planning horizon.

4.3 Application of GKDA to four planning problems

We applied GKDA to four different planning problems: new technology investment planning, clinical trial planning, process network synthesis and artificial lift infrastructure planning. For each problem, the following subsections present in detail how to construct KSPs, to determine eligible item lists, objective functions and weight constraints, and to implement the algorithm.

4.3.1 New technology investment planning (NTIP) problem

Givens for the new technology investment planning (NTIP) problem is a fixed length planning horizon divided into equal time periods [$t \in T$], a set of chemical processing technologies [$i \in I$], a set of chemicals [$n \in N$], and a set of technology maturity stages [$sg \in \{\text{Laboratory(L)}, \text{Pilot Plant(PP)}, \text{Commercialization(C)}\}$]. The objective is to determine the technology investment plan that yields the expected minimum total cost. Demand at each time period should be met either via production or via purchasing at market price.

The yield of a technology is not known until the technology reaches the commercialization stage. To reach a stage, the installed capacity of a technology must reach the minimum capacity threshold for that stage. A technology may be abandoned at the completion of any stage. Following Colvin and Maravelias (2008), we define the realizable values of uncertain parameters associated

with project abandonment with set $\Psi_i = \{L(A), PP(A), PP(P)\}$, where $PP(A)$ corresponds to the outcome that technology i is abandoned at the pilot plant stage, and $PP(P)$ corresponds to the outcome that technology i reaches the commercialization stage.

There are four uncertain parameters associated with each new technology: (1) project abandonment (ψ_i), (2) process yield (χ_i), (3) learning-by-searching elasticity (α_i), and (4) learning-by-doing elasticity (β_i). The learning elasticities define how the capacity expansion cost may change with investments in a two-factor learning curve. Demand for products is uncertain for time periods beyond the first, and there is an uncertain parameter ($D_{n,t}$) representing the demand of product n at time t. Hence, the NTIP problem contains endogenous (technology development, yield, and learning elasticities) and exogenous (demand) uncertainties.

The NTIP problem has one binary state decision variable, the technology maturity stage ($Y_{i,sg,t,s}$), and two continuous state variables, the level of capacity expansion ($X_{i,t,s}$), and the level of research expenditure ($RDX_{i,t,s}$) for technology i at time period t in scenario s. The recourse actions include production rate ($M_{i,n,t,s}$) of material n using technology i at time period t in scenario s. The MSSP model for the NTIP problem, whose deterministic equivalent is an MINLP model, was developed by Christian and Cremaschi (2018). It is provided in Appendix E1 and E2 along with its nomenclature for completeness.

The first step of the GKDA is the identification of eligible items for the root-node KSP. The items for the NTIP problem include all current-stage decision variables and recourse actions, the technology maturity stage ($Y_{i,sg}$), the level of capacity expansion (X_i) and of research expenditure (RDX_i) for each technology, and the production rate of each material using each technology ($M_{i,n}$). In KSPs, the variable $Y_{i,sg}$ is a binary decision variable, while the variables X_i, RDX_i, and $M_{i,n}$ are continuous decision variables. The item costs (which can be seen as the negative of item values)

are the cost of capacity expansion for technology i (CC_i), and purchasing price of material n ($MCst_n$). The KSP objective function is constructed using the straightforward approach, and the general form of the KSPs for the NTIP problem is given in Eqns. (4.22) – (4.30).

$$max: -\sum_{n} MCst_n \cdot \left(E[D_n] - \sum_{i} \gamma_{i,n,PR} \cdot E[\chi_i] \cdot M_{i,n} \right) - \sum_{i} E[CC_i] \cdot X_i - \sum_{i} RDX_i \tag{4.22}$$

$$E[CC_i] = \sum_{a_i \in \alpha_i} \sum_{b_i \epsilon \beta_i} P(a_i, b_i) \cdot CCO_i \cdot \left(\frac{CX_{i,t-1} + X_i}{CX_{i,0}} \right)^{b_i} \cdot \left(\frac{RD_{i,t-1} + RDX_i}{RD_{i,0}} \right)^{a_i} \tag{4.23}$$

$$M_{i,n} \leq Y_{i,3}(CX_{i,t-1} + X_i) E\left[\psi_i\right] \tag{4.24}$$

$$CX_{i,t-1} + X_i \geq CX_{i,sg}^{min} \cdot Y_{i,sg} \tag{4.25}$$

$$CX_{i,t-1} + X_i \leq \sum_{sg<3} \left(CX_{i,sg}^{min} - CX_{i,sg+1}^{min} \right) \cdot Y_{i,sg} \tag{4.26}$$

$$Y_{i,sg} \leq Y_{i,sg-1} \tag{4.27}$$

$$Y_{i,sg} \leq Y_{i,sg,t-1} \tag{4.28}$$

$$X_i \leq X_i^{max} \tag{4.29}$$

$$RDX_i \leq RDX_i^{max} \tag{4.30}$$

.

In Eq. (4.22), the parameter $\gamma_{i,n}$ is the ratio of the stoichiometric coefficient of the chemical n and the primary reactant/feedstock PR for technology i multiplied by the ratio of the molecular weights. The values of expected demand $E[D_n]$ and expected yield $E[\chi_i]$ are straight forward to calculate. For calculating $E[CC_i]$, the values of the outcomes depend on the value of the variable X_i. As such, the equation is incorporated as a constraint to the KSPs (Eq. (4.23)). In Eq. (4.23). , $P(a_i, b_i)$ is the joint distribution of uncertain parameters $a_i \in \alpha_i$ (learning-by-searching elasticity)

and $b_i \in \beta_i$ (learning-by-doing elasticity), and $CX_{i,t-1}$ and $RD_{i,t-1}$ represent the cumulative installed capacity and research expenditure at the previous time period. The GKDA calculates these values as the sums of X_i and RDX_i, respectively, for all parent KSPs. The scenario specific constraints of the NTIP problem are included in KSPs as Eqns. (4.24) – (4.30). Similar to $CX_{i,t-1}$, the value of $Y_{i,sg,t-1}$ is calculated based on the solutions of the parent KSPs. The resulting KSPs are MINLP models.

After solving the root node KSP, the GKDA determines N_{t+1} based on the realized uncertain parameters associated with the decisions made in the root node KSP and constructs the corresponding child KSPs according to the realizations. Because the constraints $g_{i,t,s}(.)$ and $h_{i,t,s}(.)$ of the NTIP problem MSSP include uncertain parameters, the GKDA determines the values of the recourse-action decision variables using the state decision variable values determined by the optimum solution of the KSP for each unique realization of uncertainty for the realized parameters. The GKDA repeats the process until the end of the planning horizon.

4.3.2 The clinical trial planning problem

The goal of the clinical trial planning problem is to identify the schedule of clinical trials for a set of candidate drugs, $i \in I = \{1,2, \dots, |I|\}$ that maximizes the expected net present value (ENPV) of the clinical trial pipeline over a planning horizon ($t \in T$). A drug has to complete three clinical trials in order ($j \in J = \{PI, PII, PIII\}$), and it can either pass (P) or fail (F) them. The binary decision variable $x_{i,j,t,s}$ is equal to one if drug i clinical trial j is started at time period t under scenario s, and zero otherwise. The outcomes of the clinical trials are uncertain and decision-dependent because the outcome of a clinical trial is realized once the trial is completed. The clinical

trial outcomes are defined with a set $\Theta_i = \{PI(F), PII(F), PIII(F), PIII(P)\}$, where $PIII(P)$ corresponds to the outcome that drug i passes trials PI,PII, and $PIII$.

Clinical trial j of drug i has a known cost C_{ij}, required resources ρ_{ijr} (where $r \in R = \{1,2,\dots,|R|\}$), and fixed duration τ_{ij}. After successfully completing all clinical trials, the revenue from drug i is realized. Potential revenue is defined using the maximum revenue, rev_i^{max}, associated with drug i. The patent life of a drug is assumed to starts shrinking once the drug enters the pipeline, and associated losses are represented by two penalty terms: γ_i^D (loss of market share) and γ_i^L (loss of patent life). A slightly revised version of the MSSP formulation of Colvin and Maravelias (2008) and the corresponding nomenclature is given in the Appendix.

The GKDA starts by generating the eligible items set E for the root node KSP. The items $k \in E$ are created by enumerating all possible decisions under the logic of the sequencing constraints for the clinical trials defined by the MSSP model. In KSPs, the decision variable $x_{i,j,t,s}$ is identified by x_k, where k corresponds to a specific drug-clinical trial pair.

The objective function of the original MSSP is given in Eq. (4.31), which reveals that the revenue associated with the drug i can only be realized once the drug completes its last clinical trial ($PIII$) successfully, i.e., the revenue is *only associated with certain decision variables*, i.e., $X_{i,PIII,t,s}$, which satisfies CONDITION 1. Thus, the item values are generated using MPGA, *viz* Eq. (4.32). They are generated considering the potential revenue associated with drug i, which is calculated by taking the maximum revenue associated with drug i and deducting the losses due to loss of active patent life and clinical trial costs.

$$
\begin{aligned}
ENPV = f(X_{i,j,t,s}) \\
&= \sum_s p_s \left\{ \sum_i \sum_t \left[rev_i{}^{max} X_{i,PIII,t,s} \right.\right. \\
&- \gamma_i{}^D \sum_{j=PII,PIII} \left(-X_{i,j,1,s} + \sum_{t'>\tau_{i,j-1}}^{t' \le t} X_{i,j-1,t'-\tau_{i,j-1},s} - \sum_{t'}^{t' \le t} X_{i,j,t',s} \right) \\
&\left. - \gamma_i{}^L (t + \tau_{i,PIII}) X_{i,PIII,t,s} \right] \\
&+ \sum_i \sum_{j \neq PI} rev_{i,j}^{open} f_{i,j} \left(-X_{i,j,1,s} + \sum_{t'>\tau_{i,j-1}}^{t' \le |T|} X_{i,j-1,t'-\tau_{i,j-1},s} \right. \\
&\left. - \sum_{t'}^{t' \le |T|} X_{i,j,t',s} \right) + \sum_i rev_{i,PI}^{open} f_{i,PI} \left(1 - \sum_{t'}^{t' \le |T|} X_{i,PI,t',s} \right) \\
&\left. + \sum_i \sum_{j \in \{PI,PII\}} \sum_{t>|T|-\tau_{i,j}} rev_{i,j,t}^{run} f_{i,j+1} X_{i,j,t,s} - \sum_{i,j,t} cd_t C_{i,j} X_{i,j,t,s} \right\}
\end{aligned}
\tag{4.31}
$$

$$
V_{k(i,j),t} = \mathrm{E}\left[rev_i{}^{max} - \gamma_i{}^L \left(t + \sum_{j'>j} \tau_{i,j'} \right) - cd_t C_{i,j} \right] \tag{4.32}
$$

Item weight $W_{k,r}$ is equal to the resource requirement of the associated drug-clinical trial pair for resource type r. The $|R|$-dimensional capacity vector $\{W_r^{max}\}$ contains the available resources, ρ_r^{max}. The general form of KSPs for the clinical trial planning problem is given in Eqns. (4.33)-(4.35), and they are $|R|$-dimensional 0-1 knapsack problems. For this problem, the GKDA is identical to the KDA. Eqn. is the scenario specific constraints of the original MSSP other than sequencing constraints, which are used for generating eligible items along with NACs.

Once the root node KSP is solved, the number of child KSPs generated and their generation times are based on the outcomes of the root node KSP solution. For example, suppose the solution of the root node KSP selects two items associated with first clinical trials (PI) of drugs A and B,

i.e., (A, PI) and (B, PI), a total of $2^2 = 4$ child KSPs are generated at $t = \max(\tau_{A,PI}, \tau_{B,PI})$. These four KSPs represent the four possible outcomes of the trial (A, PI) and (B, PI): (1) both drugs pass the clinical trial, (2) trail (A, PI) fail while trial (B, PI) pass, (3) trail (A, PI) pass while trial (B, PI) fail, and (4) both drugs fail clinical trial PI. The algorithm then updates the eligible item sets, objective functions and weight constraints for all four KSPs. Because scenario specific constraints of clinical trial planning MSSP model ($g_{i,t,s}(.)$ and $h_{i,t,s}(.)$) do not contain any uncertain parameters, the values of the recourse-action decision variables determined based on the optimum solution of the KSPs are guaranteed to be feasible. The algorithm continues until either the eligible item sets are empty or the end of the planning horizon is reached. A detailed discussion of KDA and its application to this problem can be found in Christian and Cremaschi (2015, 2017).

$$max: \sum_{k \in E} V_k x_k \tag{4.33}$$

$$\sum_{k} W_{k,r} x_k \leq W_r^{max} \qquad \forall r \in R \tag{4.34}$$

$$x_k \in \{0, 1\} \qquad \forall k \in E \tag{4.35}$$

4.3.3 Process network synthesis under uncertain process yields

The schematic of a simple process network is given in Figure 4.2.2. Currently, the final product A is produced from chemical B via Process 3, which has a fixed capacity of 3×10^6 tons/year and a fixed yield of 0.7. Currently, chemical B is purchased. Two new processes, Process 1 and Process 2, are available to install for producing chemical B from raw materials C and D, respectively. Both of these processes, neither of them, or either of them may be installed to produce all or some of the required amount of chemical B. Nodes N1 and N2 represent stream mixing points. If demand

for product A cannot be satisfied by production at a given time period, the product may be purchased to satisfy demand.

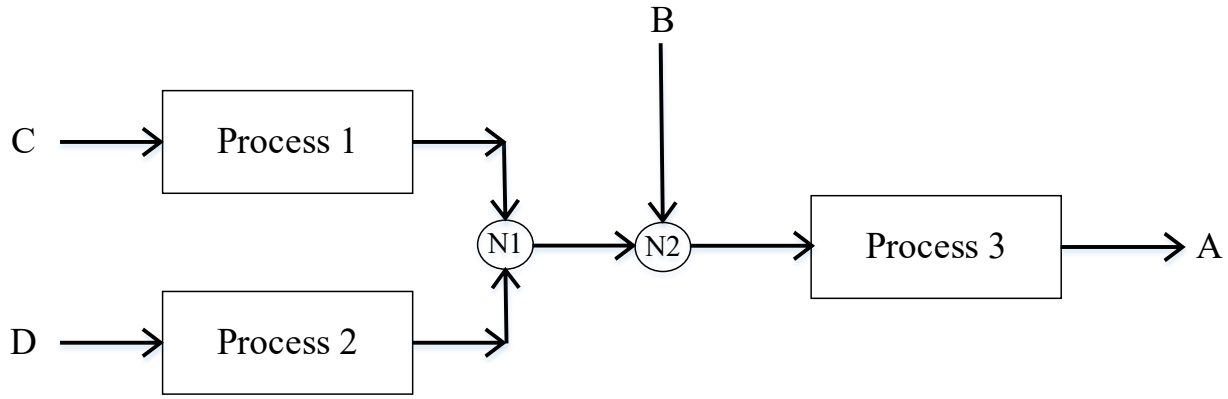

Figure 4.2.2: Schematic of a simple process network

The goal of the process synthesis problem is to determine optimal capacity expansion investments and process operation decisions that will maximize the total expected profit from the sales of product A. The decisions are which (if any) of the processes ($i \in \{Process\ 1, Process\ 2, Process\ 3\}$) should be installed or expanded ($y_{i,t}^{exp,s} \in \{0,1\}$), which processes should be operated ($y_{i,t}^{oper,s} \in \{0,1\}$, the capacity levels of the expansions ($w_{i,t}^{QE,s} \in \mathbb{R}^{\geq 0}$), the flowrates throughout the network ($w_{k,t}^{rate,s} \in \mathbb{R}^{\geq 0}$), purchases ($x_t^{purch,s} \in \mathbb{R}^{\geq 0}$), sales ($x_t^{sales,s} \in \mathbb{R}^{\geq 0}$), and level of inventory ($w_t^{inv,s} \in \mathbb{R}^{\geq 0}$) of product A for satisfying the demand (d_t) at each time period ($t \in T$) under each scenario ($s \in S$). The endogenous uncertain parameters of the problem are the yields (tons of product per ton of raw material) of Process 1 and Process 2, which are represented by (θ_i^s) and are only resolved once the process is installed. We assume that the demand for the planning horizon is known, and that there no exogenous uncertain parameters. The original nomenclature and MSSP formulation and the corresponding nomenclature developed

by Goel and Grossmann (2006) for the process synthesis problem is given in Appendix. Next, we explain how GKDA solves this process synthesis problem.

The algorithm starts by identifying the eligible items for constructing the root-node KSP. The eligible item set E includes all current-stage decision variables and recourse actions, which are expansion decisions (y_i^{exp}), operation decisions (y_i^{oper}), capacity levels of expansions (w_i^{QE}), flowrates throughout the network (w_k^{rate}), purchases (x^{purch}), sales (x^{sales}), and level of inventory (w^{inv}) for product A. In KSP, the decision variables y_i^{exp} and y_i^{oper} are binaries, while the variables w_i^{QE}, w_k^{rate}, x^{purch}, x^{sales}, and w^{inv} are continuous. Note that KSP constructed only considers the current time period, hence, the index for time period is not needed in the model. The NACs in GKDA are enforced by the KSP generation scheme, and hence, the scenario index is also not necessary for the decision variables of KSPs. The item costs (which can be seen as the negative of item values) are fixed expansion cost of process i (εFE_i), fixed operating cost of process i (εFO_i), variable expansion cost of process i (VE_i), variable operating cost for flowrate k (VO_k), purchasing price of product A (α), and cost of maintaining inventory for product A for one time period (γ). The sales price of product A (β) is the true item value for the item associated with sales. The coefficient ε is a discount factor between zero and one, and it is used to better approximate the fixed expansion and fixed operating costs of a process for a single time period. The general form of the KSPs constructed by GKDA for the process synthesis problem is given in (4.36)-(4.53).

$$max: -\sum_i \left(\varepsilon FE_i y_i^{exp} + \varepsilon FO_i y_i^{oper} + VE_i w_i^{QE}\right) - \sum_k VO_k w_k^{rate} - \left(\alpha x^{purch} - \beta x^{sales} + \gamma w^{inv}\right) \tag{4.36}$$

$$w_3^{rate} = E[\theta_1] w_1^{rate} \tag{4.37}$$

$$w_4^{rate} = E[\theta_2] w_2^{rate} \tag{4.38}$$

$$w_8^{rate,} = \theta_3 w_7^{rate} \tag{4.39}$$

$$w_5^{rate} = w_3^{rate} + w_4^{rate} \tag{4.40}$$

$$w_7^{rate} = w_5^{rate,s} + w_6^{rate} \tag{4.41}$$

$$w^{inv} = w_{pervious}^{inv} + (w_8^{rate} + x^{purch} - x^{sales}) \tag{4.42}$$

$$x^{sales} = d \tag{4.43}$$

$$w_i^{cap} = w_i^{cap,previos} + w_i^{QE} \tag{4.44}$$

$$w_3^{rate} \leq w_1^{cap} \tag{4.45}$$

$$w_4^{rate} \leq w_2^{cap} \tag{4.46}$$

$$w_8^{rate} \leq w_3^{cap} \tag{4.47}$$

$$L_i^{QE} y_i^{exp} \leq w_i^{QE} \leq U_i^{QE} y_i^{exp} \tag{4.48}$$

$$L_1^{inflow} y_1^{oper} \leq w_3^{rate} \leq U_1^{inflow} y_1^{oper} \tag{4.49}$$

$$L_2^{inflow} y_2^{oper} \leq w_4^{rate} \leq U_2^{inflow} y_2^{oper} \tag{4.50}$$

$$L_3^{inflow} y_3^{oper} \leq w_8^{rate} \leq U_3^{inflow} y_3^{oper} \tag{4.51}$$

$$y_i^{exp} + \mathrm{y}^{\mathrm{exp}\ \_previous} \geq y_i^{oper} \tag{4.52}$$

$$y_i^{oper} \geq y_i^{exp} \tag{4.53}$$

(4.37) – (4.44) are the equality constraints of the KSPs (4.16), and they ensure mass balance throughout the network, calculate inventory level, and enforce demand satisfaction. (4.45) – (4.53)

are the inequality constraints of the KSPs (4.15), and they constrain flowrates throughout the network either to the capacity levels of the processes or to the upper and lower limits defined by the problem, and ensure that only installed capacities for processes are operated. The KSPs use the expected values of the uncertain parameters, $E[\theta_1]$ and $E[\theta_2]$, for the yields of Process 1 and Process 2 before they are realized. When the solution of a KSP recommends installing and operating Process 1 or Process 2, the yield of the installed process is realized. Since the constraints $g_{i,t,s}(.)$ and $h_{i,t,s}(.)$ of the process synthesis MSSP include uncertain parameters, the values of the recourse-action decision variables are calculated using the stage decision variable values determined by the optimum solution of the KSP for each unique realized yield value as outlined in Figure 4.2.1.

4.3.4 Artificial lift infrastructure planning problem

The ALIP MSSP model contains two binary decision variables, $w_{i,t,s}$ and $z_{i,t,s}$. The variable $w_{i,t,s}$ is equal to one if ALM $i \in I$ is installed at time $t \in T$ under scenario $s \in S$, and zero otherwise. The variable $z_{i,t,s}$ is equal to one if ALM i is uninstalled at time t under scenario s. The ALIP MSSP includes nine candidate ALMs ($|I| = 9$), which are Plunger Lift (PL), Foam Lift (FL), Well Head Compression (WHC), Velocity Strings (VS), Gas Lift (GL), Electrical Submersible Pump (ESP), Sucker Rod Pump (SRP), Progressing Cavity Pump (PCP) and Jet Pump (JP). The model assumes that the well is loaded at the first time period, and the first ALM is installed at the second time period. The original MSSP model for the ALIP was introduced in Zeng and Cremaschi (2017a) and it can be found in Appendix A along with its nomenclature.

The first step of GKDA is to construct the root node KSP. The items $i \in E$ for the root node KSP are created by enumerating all allowable decisions at the second time period of the planning horizon under the logic of the sequencing constraints of the ALIP MSSP model. Therefore, the

eligible item set of the root node KSP contains all candidate ALMs. The decision variable of KSP, y_i, represents ALM i that is installed and currently operating. The item value is calculated using the expected net present value of the potential revenue and costs of installing and operating the associated ALM i for the rest of the planning horizon. As this item includes a capital investment decision, it satisfies CONDITION 2. Therefore, in item value calculations, the capital cost $CC_{i,t}$ at time t for ALM i is converted to annual capital charge using annual capital charge ratio ε. Figure 4.2.3 summarizes how the item values are calculated for KSPs generated at time period t based on the original MSSP objective function.

MSSP

Max: Net Present Value

$$ENPV = \sum_s P_s \left(\sum_r \left(GI_{r,s} - TI_{r,s} \cdot x_{r,s} \cdot FT\right) \left(\frac{1}{(1+MARR)^r}\right) - CC_s(1-FT) \right)$$

Gross Income

$$GI_{r,s} = \left(Rev_{r,s}(1-RT)(1-LT) - \sum_{i,t \geq r, p \leq r} \left(Cm_i y_{i,t,p,s}\right) \right) WI \quad \forall r, \mathrm{s}$$

Revenue

$$Rev_{r,s} = P_g Qg_{r,s} + P_o Qo_{r,s} + P_{ng} Qng_{r,s} \quad \forall r, s$$

Taxable Income

$$TI_{r,s} = GI_{r,s} - Dep_{r,s} \quad \forall r, \mathrm{s}$$

$$TI_{r,s} \leq x_{r,s} M \quad \forall r \in T$$

$$TI_{r,s} > \left(x_{r,s} - 1\right) M \quad \forall r \in T$$

Depreciation

$$Dep_{r,s} = \sum_{i,t \geq r, p \geq r-n+1}^{p \leq r} \left(\frac{Ce_i}{n} y_{i,t,p,s}\right) \quad \forall t, p, r, i, s$$

Capital Cost

$$CC_s = \sum_{i,t,p} \left(Co_i y_{i,t,p,s} \frac{1}{(1+MARR)^p}\right) \quad \forall t, p, \mathrm{s}$$

Decline Curve Function

$$Qg_{r,s} = \sum_{i,t,p} \left\{ y_{i,t,p,s} Qg_{p-1,s} Qrc_{i,s} \left(1 + bD(r-p+1)\right)^{-\frac{1}{b}} \right\}$$

$$Qo_{r,s} = \sum_{i,t,p} \left\{ y_{i,t,p,s} Qo_{p-1,s} Qrc_{i,s} \left(1 + bD(r-p+1)\right)^{-\frac{1}{b}} \right\}$$

$$Qng_{r,s} = \sum_{i,t,p} \left\{ y_{i,t,p,s} Qng_{p-1,s} Qrc_{i,s} \left(1 + bD(r-p+1)\right)^{-\frac{1}{b}} \right\}$$

KSP:

Max: Item Values in Knapsack

$$V_{i,t} = \sum_{r \geq t}^{T} \left(GI_{i,r,t} - TI_{i,r,t} \cdot FT\right) \left(\frac{1}{(1+MARR)^r}\right) - \varepsilon CC_{i,t}(1-FT)$$

Gross Income

$$GI_{i,r,t} = \left(Rev_{i,r,t}(1-RT)(1-LT) - Cm_i \right) WI \quad \forall r, \mathrm{s}$$

Revenue

$$Rev_{i,r,t} = P_g Qg_{i,r,t} + P_o Qo_{i,r,t} + P_{ng} Qng_{i,r,t} \quad \forall r, s$$

Taxable Income and Depreciation

$$TI_{i,r,t} = GI_{i,r,t} - \frac{Ce_i}{n} \quad \forall \mathrm{i}, r$$

Capital Cost

$$CC_{i,t} = Co_i \frac{1}{(1+MARR)^t} \quad \forall t, p, \mathrm{s}$$

Decline Curve Function

$$Qg_{i,r,t} = Qg_{p-1} E[Qrc_i] \left(1 + bD(r-t+1)\right)^{-\frac{1}{b}}$$

$$Qo_{i,r,t} = Qo_{p-1} E[Qrc_i] \left(1 + bD(r-t+1)\right)^{-\frac{1}{b}}$$

$$Qng_{i,r,t} = Qng_{p-1} E[Qrc_i] \left(1 + bD(r-t+1)\right)^{-\frac{1}{b}}$$

Figure 4.2.3: Objective function of MSSP and item values in KSPs for artificial lift infrastructure planning problem (See Appendix for nomenclature)

The weight of each item is equal to one, and at most one ALM can be operational at any time period, which defines the capacity constraints of the KSPs (4.55). The general form of the KSPs generated at time period t used for solving the ALIP using GKDA is given in Eqns. (4.54)-(4.56). The KSPs for this problem are 0-1 knapsack problems.

$$max: \sum_{i \in E} V_i x_i \tag{4.54}$$

$$\sum_{i \in E} x_i \leq 1 \tag{4.55}$$

$$x_i \in \{0, 1\} \qquad \forall i \in E \tag{4.56}$$

The GKDA solves the root node KSP, which results in the realization of production rate uncertainty for the selected ALM. Next, the GKDA determines N_{t+1}, which is equal to the number of realizations of the production rate uncertainty for the selected ALM, and generates N_{t+1} child KSPs. Because the scenario specific constraints of ALIP MSSP model do not contain any uncertain parameters, the values of the recourse-action decision variables determined based on the optimum solutions of the KSPs are guaranteed to be feasible.

4.4 Results and Discussion

We apply GKDA to generate feasible solutions for different instances of the four planning problems. To assess the GKDA solution quality, deterministic equivalents of the MSSPs given in Appendices are solved for the instances when possible. All models and algorithms were implemented in Pyomo and solved using CPLEX 12.6.3 on a standard node of Auburn University Hopper Cluster. Data for all instances are available upon request from the corresponding author.

4.4.1 New technology investment planning problem

We solve four instances of the NTIP problem with different network topologies (Figure 4.2.4) (Christian and Cremaschi, 2018). The first instance (Figure 4.2.4a) compares a new technology at a Laboratory stage, T1, and a mature technology, T2, at Commercialization stage. The second one (Figure 4.2.4b) compares two Laboratory stage technologies (T1 and T2). Instance 3 (Figure 4.2.4c) considers a retrofitting problem, where two of the technologies (T2 and T4) are at Laboratory stage, and two are mature (T1 and T3).

The last NTIP problem instance considers a segment of the chemical process industry, where the desired product is ethylene (Figure 4.2.4d). Ethylene can be produced from naphtha using mature cracking technology and from biomass using two routes, Fermentation-Catalytic Dehydration or Gasification-Catalytic Conversion-Catalytic Dehydration. We assume that Fermentation and Catalytic Dehydration are mature, and that Gasification and Catalytic Conversion are at Laboratory stage. The planning horizon for all instances is three years discretized into one-year time periods.

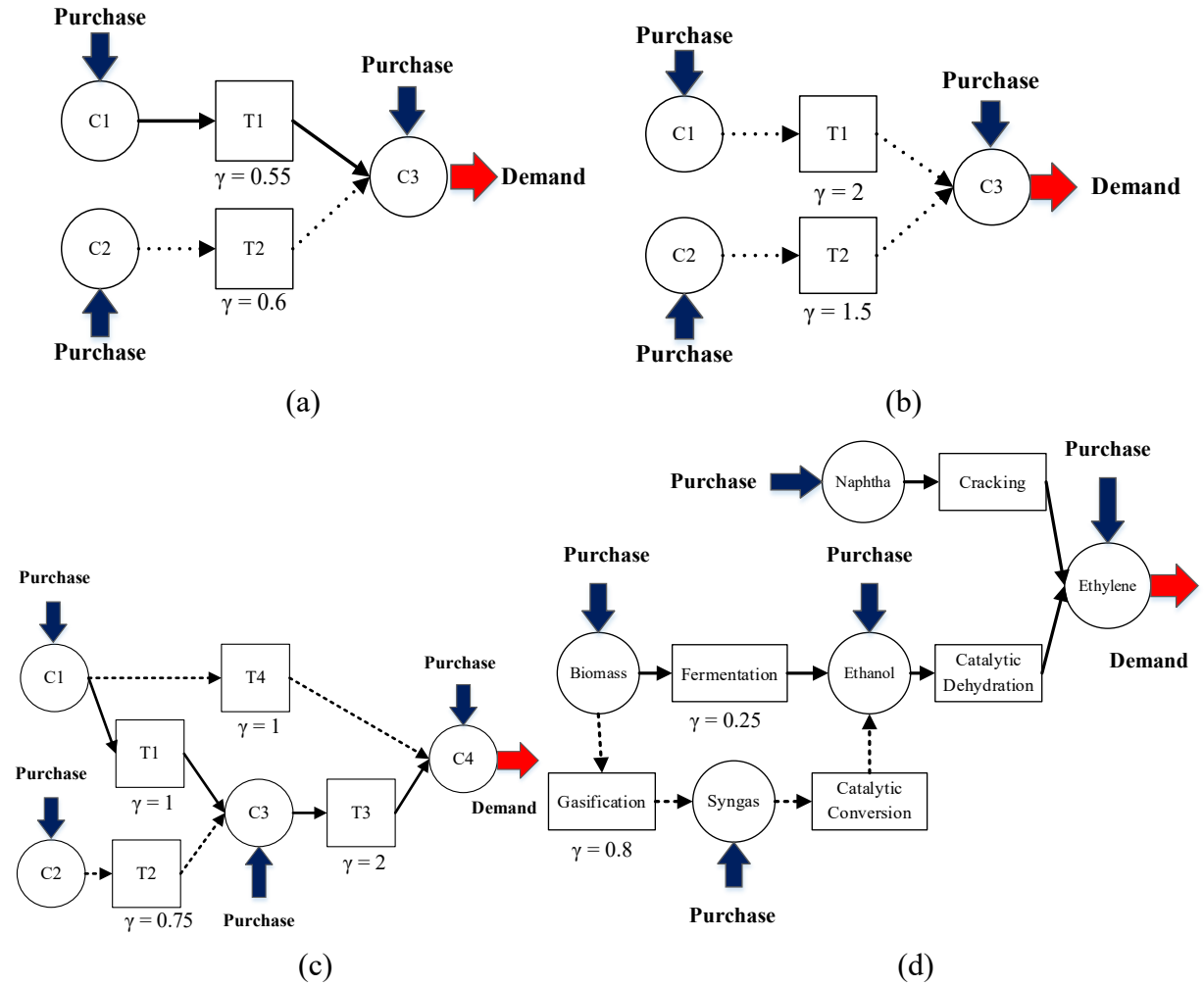

Figure 4.2.4: Network topologies of the NTIP problem instances (Christian and Cremaschi, 2018)

The deterministic equivalent of the NTIP problem MSSP yields a large-scale non-convex MINLP model. For example, the MINLP model of the first instance has 1728 binary and 9409 continuous variables, and 46,561 constraints, and the MINLP model of the last instance has 1.983 million variables, of which are 89,856 binaries, and 3.905 million constraints. These problems cannot be solved to global optimality using commercial global solvers. A relaxed MILP model was constructed using piecewise linear relaxation (Misener et al., 2011) and exact linearization (Oral and Kettani, 1992), and solved to obtain dual bounds. The GKDA provides the primal bounds. The results are summarized in Table 4.1.

Table 4.1: Primal (GKDA) and dual (relaxed MILP) bounds, solution times, and relative gaps for NTIP problem instances

	Primal bound - GKDA		Dual bound - MILP relaxation		Gap (%)
	ENPV ($)	Solution Time (CPUs)	ENPV ($)	Solution Time (CPUs)	
Instance 1	152	40.0	147	4	3.4
Instance 2	101	1.1	91.3	1074	9
Instance 3	6.88	0.6	6.88	138	0
Instance 4	168	1.7	161	2313	4.1

For problem instance 1, the GKDA solved 217 MINLP KSPs to optimality generating a feasible solution with an objective value of $151.8 billion. Each KSP required an average of 0.01 CPUs to solve and resulted in a total algorithm time of 40 CPUs. The dual bound is $146.4 billion, yielding a relative gap of 3.4%. The GKDA solution recommends installing 6 megatons (Mt) per technology at the first time period despite the possibility of T2 failing to be developed. The GKDA determines investment decisions based on the expected outcomes. For this problem instance, the expected success is high enough to justify a full development of T2. At time period two, the solution obtained by the GKDA recommends maximum research and capacity expansion investments in T2 if the development of T2 is successful. If the development of T2 is unsuccessful, the algorithm suggests installing 6 Mt of T1 capacity. At time period three, for scenarios where the development of T2 is successful, the GKDA solution selects to install capacity for T2. The level of capacity expansion depends on the realized yield value but is approximately 2 Mt. For scenarios where T2 is not successful, the GKDA solution installs an additional capacity of 6 Mt of T1.

The GKDA yielded a feasible solution in 1.1 CPUs for the second instance (Table 4.1). The algorithm solved four MINLP KSPs, and the relative gap is 9%. The GKDA solution does not recommend any investments in either technology. In the MILP model, the capacity expansion cost is under-estimated thus the value of expanding the undeveloped technologies is higher. The probability of success is not high enough to offset the potential loss in the KSPs.

The third instance had a primal objective function value of $6.884 billion, which is equal to the relaxed MILP solution, and hence the GKDA solution is optimal. The optimal investment strategy is to use the existing capacities to produce the chemicals. The GKDA found the optimal solution in 0.6 CPUs by solving eight MINLP KSPs. The difference in the solution times among the three instances demonstrate that the solution time of the GKDA depends on the number of KSPs solved, which is a function of the uncertainty realizations that occur during the planning horizon.

When applied to the fourth instance, the GKDA solution has an objective value of $168 billion. The dual bound is $161 billion, which yields a percent gap of 4.1%. The GKDA solved 19 MINLP KSPs with a total runtime of 1.7 CPUs. The solutions are very similar. The MILP solution recommends producing ethylene using the cracking technology, and purchasing ethylene when necessary to satisfy demand. No capacity expansions are recommended. The GKDA solution also does not recommend installing any new capacity. However, it recommends using existing fermentation and dehydration technologies to produce ethylene from biomass.

4.4.2 Clinical trial planning problem

In this section, we apply GKDA to find feasible solutions to instances of clinical trial planning problem with three-, four-, five-, six-, seven- and ten-products, and three clinical trials with different planning horizon lengths. We also solved the deterministic equivalents of the MSSPs for

these instances to 0.1% optimality gap when possible. Number of trials, time periods, and scenarios for the instances are summarized in Table 4.2 along with the optimum ENPVs and the MSSP solution times for the instances that were solved to optimality. Table 4.2 illustrates the exponential growth in number of scenarios with the increases in the number of drugs. With this growth, the MSSP models quickly become computationally intractable due to their space and time complexities. As expected, the MSSP solution times grow exponentially with the number of scenarios. For example, CPLEX 12.6.3 required more than eight CPU hours to solve the six-product instance. The MSSP models of seven-, and ten-product problems cannot be generated in RAM, and hence cannot be solved using CPLEX 12.6.3 due to their space complexities.

Table 4.2: Main characteristics, ENPVs and MSSP solution times of the clinical trial planning problem instances

Case Name	Number of Trials	Number of Time Periods	Number of Scenarios	ENPV ($M)	MSSP Solution Time (CPUs)
3-prod	3	12	64	1189	4
4-prod	3	6	256	1696	7
5-prod	3	6	1,024	2082	185
6-prod	3	6	4,096	2450	29,357
7-prod	3	8	16,384	--	--
10-prod	3	10	1,048,576	--	--
15-prod	3	10	1,073,741,824	--	--

The ENPVs of the GKDA solutions, the percent gaps from the optimum solutions (when available), and the GKDA solution times are compiled in Table 4.3. The percent gaps ranged from 0.9% to 1.8%. The GKDA obtained these feasible solutions very quickly, in 1.3 to 3.4 CPUs. For

seven-, and ten-product instances, whose deterministic equivalents cannot be generated and solved, the GKDA yielded feasible solutions in 20.0 and 1579.3 CPUs.

Table 4.3: The objective function values of the GKDA solutions, the GKDA solution times, and the corresponding relative gaps from the optimum solutions (where available) for the clinical trial planning problem instances

Case Name	ENPV of the GKDA Solution	Relative Gap	Solution Time (CPUs)
3-prod	1178	0.9%	1.3
4-prod	1677	1.1%	1.5
5-prod	2052	1.5%	4.7
6-prod	2407	1.8%	3.4
7-prod	2874	NA	20.0
10-prod	4078	NA	1579.3

4.4.3 Process network synthesis problem with uncertain process yields

The problem instance solved is derived from Goel and Grossmann (2006), and its schematic is given in Figure 4.2.2. The planning horizon is ten years discretized into one-year time periods. The realizations for the yield of Process 1 are 0.69 or 0.81 with equal probabilities. The uncertain parameter for yield of Process 2 has two equally likely realizations as well, 0.65 and 0.85. Process 3 has an initial capacity of 3×10^6 tons/year and a fixed yield of 0.7.

The deterministic equivalent of the MSSP contains 400 binary variables, 685 continuous variables, and 5753 constraints. This is a relatively small MILP and CPLEX 12.6.3 solves it in 1 CPUs. The optimum ENPV is \$11.459×106. The solution recommends installing Process 2 with a capacity of 10×10^6 tons/year at the first time period and expanding the capacity of Process 3 to 8×10^6 tons/year at the second time period.

The ENPV of the GKDA solution is \$9.159×106, which yields a percent gap of 20.0%. The capital charge rate used is calculated by $\frac{ir(1+ir)^t}{(1+ir)^t-1}$, where $ir = 0.1$. The GKDA solution recommends installing 2.85×10^6 tons/year of Process 2 capacity in the second time period. The capacity of Process 2 is further expanded to 5.71×10^6 tons/year in the third time period, and to 10×10^6 tons/year in fourth time period. In the solution, the capacity of Process 3 is expanded twice, to 4×10^6 tons/year in the third time period and to 8×10^6 tons/year in the fourth time period. The GKDA solution never recommends holding inventory because their item values in KSPs are negative, and hence the current item generation rules of the GKDA fails to correctly assess the potential benefit of holding inventory for future time periods. The differences in inventory levels along with the capacity expansion decisions explain the relatively large percent gap between the GKDA and optimum solutions for this problem.

4.4.4 Artificial lift infrastructure planning problem

We also constructed a hypothetical instance of an ALIP based on a real-world field application: Woodford shale horizontal well. The production before the loading indicated a gas rate of 800 thousand standard cubic feet per day (Mscf/D), a liquid rate of 520 barrels per day (BPD). The well is currently operated using gas lift with a well-site compressor24. The hypothetical instance assumes that the planning was carried out at the first loading prior to the installation of gas lift. The uncertain production rate parameter for each ALM is assumed to have two realizations, a High outcome, which is assumed to be 20% above its nominal value, and a Low outcome 20% below with equal probabilities. For the hypothetical instance, operating envelope of SRP is modified by setting the minimum allowable flow rate to 300 BPD. The problem is solved for planning horizon lengths of 12, 16, 24 and 36 months (time periods). The maximum CPU time is limited to 20 hours.

The MSSP models are solved to 0.1% optimality gap except for 24- and 36-month planning horizons, which cannot be solved within the allowed time limitations.

Table 4.4 summarizes the optimum ENPVs and the GKDA solution ENPVs along with the solution times. The decision trees are presented in Figure 4.2.5. The decision trees of the optimum and the GKDA solutions are identical for 12-month (Figure 4.2.5a) and 16-month (Figure 4.2.5b) planning horizons. Both solutions recommend installing SRP at the second time period to unload the well initially. If the realized production rate using SRP is the High value, then the solution recommends operating SRP until the end of the planning horizon. On the other hand, if the realized production rate is the Low value for SRP, then it is replaced by ESP at $t = 7$. The solution recommends this switch for the Low value because the operational envelop of SRP is violated before the end of the planning horizon is reached for these scenarios. It should be noted that the optimality of the GKDA solutions for this problem cannot be guaranteed.

Table 4.4: The optimum ENPVs, MSSP solution times, ENPVs of the GKDA solutions and GKDA solution times for hypothetical ALIP instances with different planning horizons

	MSSP		GKDA	
Planning Horizon (months)	ENPV ($)	Solution Time (CPUs)	ENPV ($)	Solution Time (CPUs)
12	2,783,894	21	2,783,894	0.2
16	3,416,506	4659	3,416,506	0.4
24	NA	72000	4,260,763	0.7
36	NA	72000	5,252,280	1.3

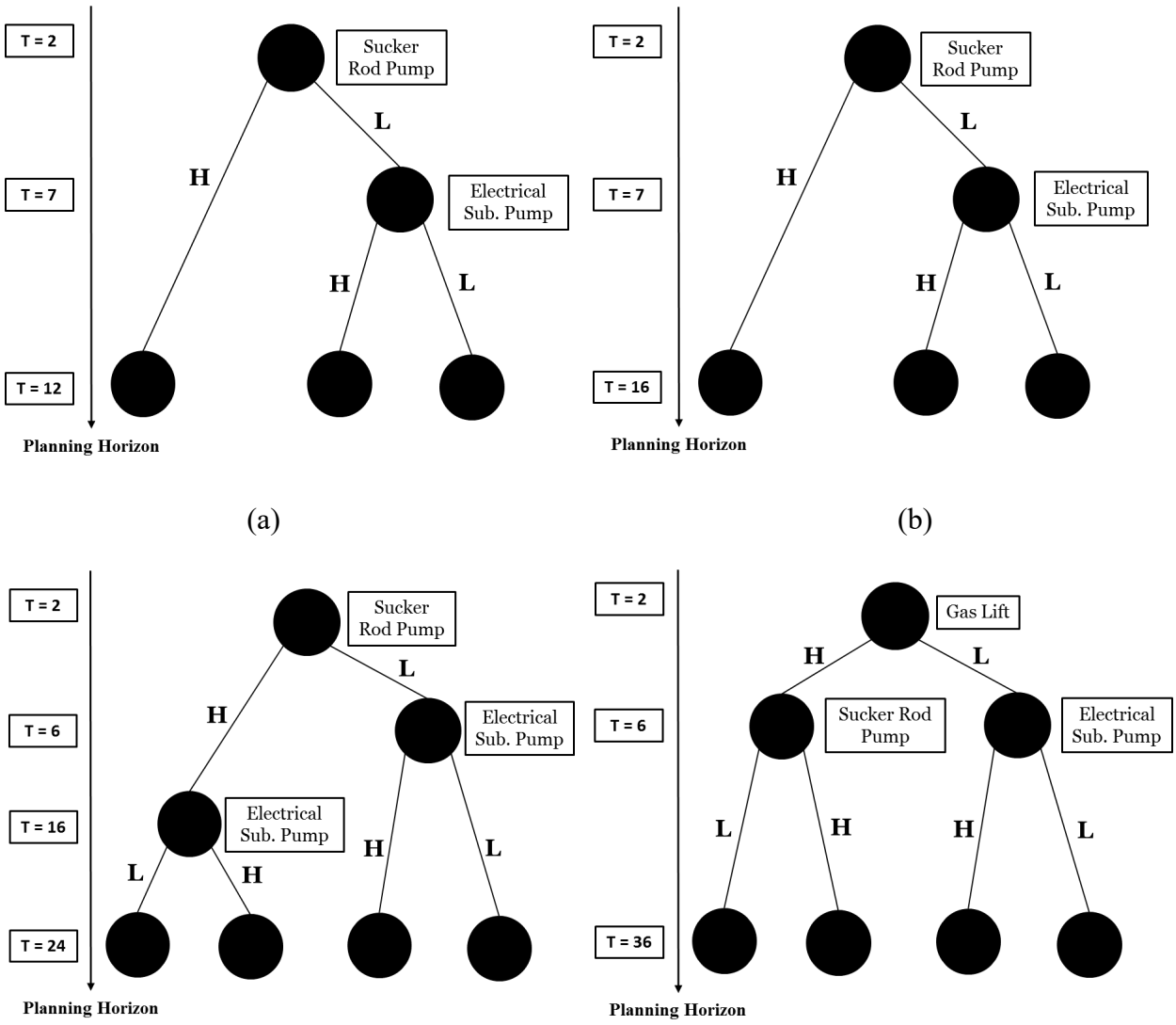

Figure 4.2.5: Decision trees of the optimum (a-b) and the GKDA solutions (a-d) for the ALIP instances

The decision trees for the 24-month and 36-month planning horizons are given in Figure 4.2.5c and Figure 4.2.5d. For longer planning horizons, even for scenarios with High production rate realization of SRP, its operational envelope is violated before the end of the planning horizon. For these problems, the solution recommends installing ESP for scenarios with High SRP production

rate realization. For scenarios with Low SRP production rate realization, the ALM selected is Gas Lift.

Comparison of the optimum ENPVs to the ENPVs of the GKDA solutions in Table 4.4 reveals that the GKDA yielded the optimum solutions for the 12- and 16-month planning horizons. The MSSP of the ALIP does not have any resource constraints and allows at most one ALM to be installed and operational at any given time period of the planning horizon. These constraints are easily enforced by the GKDA. Furthermore, the item values in KSPs are able to capture the impact of each ALM's operation fairly accurately at any given time period. It is important to note that the GKDA obtains these solutions with solution times up to six orders of magnitude faster than solving the deterministic equivalent of the ALIP MSSP.

4.5 Conclusion

This section introduces a new Generalized Knapsack-problem based Decomposition Algorithm, GKDA, to quickly obtain feasible solutions and primal bounds for large-scale multistage stochastic programming (MSSP) models under endogenous and/or exogenous uncertainties. The GKDA decomposes time periods and scenarios of an MSSP model into a series of continuous or 0-1 Knapsack Sub-Problems (KSPs). The algorithm, then, generates and solves these KSPs based on the realizations of uncertainty. Therefore, the GKDA addresses both time and space complexities associated with solving large-scale MSSPs under endogenous and/or exogenous uncertainties. The GKDA extends our previously developed heuristic algorithm, Knapsack-problem based Decomposition Algorithm (KDA), to obtain feasible solution for MSSPs under both endogenous and exogenous uncertainties with both continuous and discrete state variables and recourse actions that may also impact scenario specific constraint satisfaction.

The GKDA is applied to solve four different planning problems under uncertainty: new technology investment planning, clinical trial planning, process network synthesis, and artificial lift infrastructure planning. For all problems considered, GKDA efficiently generated feasible solutions and primal bounds. The relative gap of the GKDA solutions from the optimum solutions ranged from 0.0% to 20.0% for all problem instances solved in this paper, and the GKDA generated these solutions with up to six orders of magnitude faster than solving the deterministic equivalents of the MSSPs.

The GKDA is a greedy algorithm, and hence, fails to capture the potential impact of current time period decisions on future time periods. In next chapter, another AEEV approach has been developed to generate tighter primal bounds and address the greedy nature of GKDA.

CHAPTER 5

Absolute Expected Value Solution (AEEV)

In this chapter, to address the greedy nature of the GKDA, a general primal-bounding framework has been developed based on extending the concepts of expected value solution and value of stochastic solution from multistage stochastic programs under exogenous uncertainties. To measure the expected loss when using a deterministic model (i.e., ignoring uncertainty) instead of its stochastic counterpart, Birge and Louveaux (1997) introduced the value of the stochastic solution (VSS) concept for two-stage stochastic programming models with complete recourse. The VSS is obtained by taking the difference between the solution of the stochastic model (SRP) and the expected value of the solution of the deterministic problem (EEV): $VSS = SRP - EEV$. A large VSS indicates that it is crucial to obtain the solution of the stochastic program, while a small VSS means that it may be acceptable to obtain a deterministic solution using expected values of random variables and to avoid the computational time associated with solving the stochastic model. The EEV is calculated by implementing the first-stage decisions obtained as the solution of the deterministic problem in stochastic model (RP, also referred to as the recursion problem). The deterministic model, EV, is constructed by replacing all random variables with their expected values in the RP. The EEV yields a feasible solution for the original two-stage stochastic programming model, where decision makers are to assume expected value results for the first-stage decisions and determine the optimal recourse plan according to these results. Because the EEV is a feasible solution, it is naturally a primal bound for the two-stage stochastic program.

Escudero et al. (2007) first extended the concept of VSS (Birge and Louveaux, 1997) to multistage stochastic linear programs under exogenous uncertainties. They first solve the deterministic problem (EV) of the original MSSP. Then, the here-and-now decisions of first t

stages from the optimal solution of the EV are fixed in the original MSSP model. This MSSP model is solved to generate EEV_t, which is the expected result in stage t of using the expected value solution and is a valid primal bound for the RP. For minimization problems, we would have $EEV_t \geq RP$. The EEV, defined for two-stage models, is equivalent to EEV_T for MSSP models with T stages, where all stages' here-and-now decision variables are fixed to their values from the solution of EV. However, the EEV_T solution may be infeasible for MSSP problems with partial recourse. A few studies utilize the deterministic problem to generate various primal bounds for MSSP models under exogenous uncertainties, and they are reviewed in Section 2.2.

To the best of our knowledge, no studies define expected value problem, discuss its properties, and utilize its solution for generating a primal bound for MSSP models under endogenous uncertainties. Unlike MSSP models under exogenous uncertainties, where the scenario structure is known in advance, observation of uncertainties and scenario structure in MSSP with endogenous uncertainties are dependent on decisions made in previous stages. In this paper, we propose a general primal-bounding framework based on the expected value solution for large-scale MSSP models with complete recourse under endogenous uncertainties. The MSSP models we consider may have continuous and/or discrete state variables. The proposed framework follows the nature of decision-making when planning in multiple decision stages under uncertainty, which shown in Figure 5.2.1:

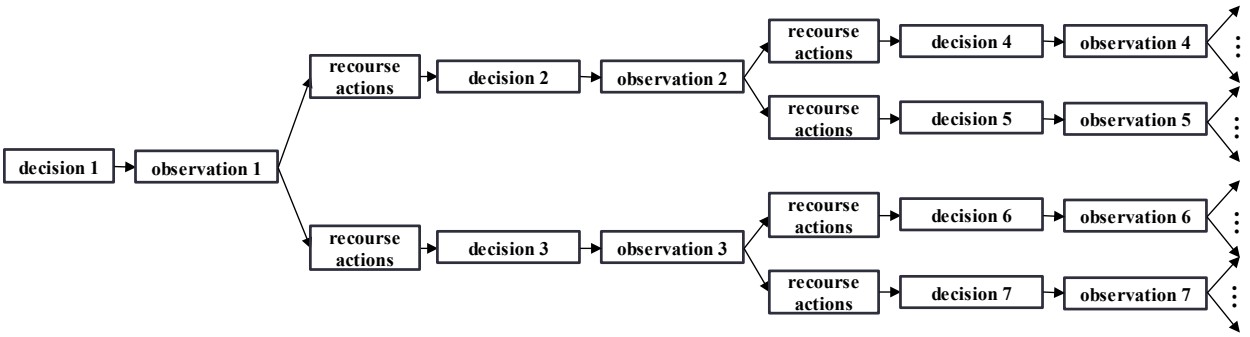

Figure 5.2.1: Steps of the proposed primal-bounding framework for MSSP models under endogenous uncertainties

where here-and-now decisions are made based on current available information and future unrevealed information is assumed based on expected results. After the here-and-now decisions are made, uncertainties are observed based on those decisions, and recourse actions are taken. In the planning horizon, decision makers repeat such decision processes until the end of planning horizon. The proposed framework, which is developed based on this process, yields a feasible solution for the original MSSP. We call this implementable solution absolute expected value solution (AEEV). The framework has a scenario-free structure and solves one deterministic problem at initialization. It only generates and solves deterministic problems based on the realized outcomes, where the non-anticipativity is implicitly enforced. The objective function value of the AEEV is a valid primal bound for the original large-scale MSSP under endogenous uncertainties if the MSSP has complete recourse.

The proposed algorithm is tested on three planning problems under uncertainties from PSE literature: pharmaceutical R&D pipeline planning problem (Colvin and Maravalias 2008), process network synthesis problem (Tarhan and Grossmann 2008), and artificial lift infrastructure planning problem (Zeng and Cremaschi 2017a). The results show that the proposed framework performs well both in terms of its computational time and its solution quality for these large-scale MSSP models. We also believe the framework and the AEEV can be utilized:

- To evaluate heuristic approaches proposed for generating feasible solutions for MSSPs in terms of both their computational times and solution qualities. The proposed framework follows the traditional pattern of decision making under uncertainty. The difference between the objective functions of the RP solution and the AEEV truly reflects the benefits of solving MSSP models with uncertainties rather than solving their deterministic counterparts. Hence, a new heuristic approach should not be considered if both the quality of the solution obtained and its computational performance are worse than those of the proposed framework.
- To be the initial primal bound for MSSP models with complete recourse under endogenous uncertainties. The proposed framework only solves relevant deterministic problems based on the observations of uncertainties. It addresses both time and space complexities of the original MSSP. Thus, obtaining the AEEV is much easier than solving the large-scale MSSP model and the objective function value at the AEEV could be regarded as the initial primal bound for these problems.

5.1 Primal bounds in MSSP models

The expected value solutions (EEV_t and EEV_T in Escudero et al. (2007)) provide valid primal bounds for MSSP problems with complete recourse under exogenous uncertainties. Several sources in literature utilize the concepts similar to the expected value solution to generate primal bounds for the MSSP problems under exogenous uncertainties. In addition to EEV_t and EEV_T, Escudero et al. (2007) also proposed dynamic solution of the average scenario ($EDEV_t$) approach for multistage stochastic linear programs with exogenous uncertainties. The approach obtains $EDEV_t$ by solving deterministic models in a rolling horizon, where individual deterministic models are constructed for each node of the scenario tree. In these models, the state decision variables are

fixed to their values at the expected value solution for the prior t stages. The authors calculated the EEV_T and the $EDEV_T$ for an instance of the financial planning problem (Birge and Louveaux, 1997), where return on each investment is an exogenous uncertain parameter and recourse actions are only taken at the last stage. The results revealed that the values of EEV_T and $EDEV_T$ were the same, and had a 60% relative gap from the optimal solution.

Maggioni and Wallace (2010) proposed the expected skeleton solution value ($ESSV$) and the expected input value (EIV) for two-stage stochastic programs. The $ESSV$ is obtained by fixing the values of all first stage decision variables which were zero in the expected value solution to zero, and, then solving the resulting stochastic program. To obtain EIV, the original stochastic program is modified by adding lower bounds to the state variables using the expected value solution. Both the EIV and the $ESSV$ can be used to evaluate the usefulness of the expected value solution. Maggioni et al. (2014) extended these concepts to MSSP with exogenous uncertainties and proposed the multistage $ESSV$ ($MESSV$) and the multistage EIV ($MEIV$). For multistage problems, Maggioni et al. (2014) utilize the expected value solution for the first t stages and generate the corresponding $MESSV_t$ and $MEIV_t$. In the same reference, the authors also proposed the multistage expected value of the reference scenario solution ($MEVRS$), and the optimal bound of all reference scenario solutions ($MEPEV$). To calculate $MEVRS$, the uncertain parameter values in the MSSP are fixed to their respective values in a certain reference scenario. The solution of this deterministic problem is implemented for state variables in the original MSSP and the recourse actions are determined. The $MEPEV$ is the best bound obtained by comparing all $MEVRS$ values obtained by using each scenario as a reference scenario. All above bounds are generated by solving the original MSSP where the state decision variables are fixed to their corresponding values obtained as the solution of the corresponding deterministic models. The authors stated that the

solutions obtained for calculating $MEVRS$, $MESSV$ and $MEIV$ were often infeasible because they all require fixing many variables using the solution of a deterministic model.

Similar to the dynamic solution of the average scenario approach ($EDEV_t$ in Escudero et al., 2007), Maggioni et al. (2014) also proposed rolling horizon approaches utilizing the fixed scenario tree in MSSP with exogenous uncertainties. Based on a fixed scenario tree, these approaches can take into account the realized values of uncertain parameters for different nodes of the stages and include uncertainty information that is more precise compared to expected value solution approach. The rolling horizon approach solves a series of deterministic models for each node in a scenario tree. In each deterministic model, the decision variables of the previous stages are fixed and the expected values of unrealized uncertain parameters are used. Maggioni et al. (2014) applied this approach to a single-sink transportation problem with demand and production capacity as exogenous uncertain parameters. The rolling horizon approach yielded a feasible solution with a 2.3% relative gap within a second.

5.2 A Systematic Framework to Generate Absolute Expected Value Solution (AEEV)

Our systematic framework utilizes a scenario-free approach to generate the absolute expected value solution (AEEV). It constructs and solves a series of deterministic optimization problems, which we will refer to as deterministic expected value sub-problems ($DEVSPs$) and recourse deterministic expected value sub-problems ($DEVSPs^{recourse}$). The here-and-now decisions and corresponding recourse actions are determined using the solutions of these sub-problems, which utilize the current available information and the expected value for future events.

The framework starts by constructing one $DEVSP$ at the initial time period. The values of all endogenous uncertain parameters are set to their expected values for this problem. The framework solves this $DEVSP$, stores and implements the here-and-now decisions recommended by the

solution for the first time period, and determines the uncertain parameter realizations associated with the implemented here-and-now decisions. To determine the recourse actions, the framework constructs a set of $DEVSPs^{recourse}$. The number of $DEVSPs^{recourse}$ generated depends on whether there have been any uncertain parameter realizations or not. If there have been no realizations, no $DEVSP^{recourse}$ is generated. If there have been uncertainty realizations, then the number of $DEVSPs^{recourse}$ constructed is equal to the number outcomes of the realized uncertain parameters. Each $DEVSP^{recourse}$ corresponds a unique outcome. In each $DEVSP^{recourse}$, the here-and-now decisions are fixed to the solution of $DEVSP$ and the values of the realized uncertain parameters are replaced by one of their outcomes for the first-time period. The recourse actions for each outcome are then determined by solving these $DEVSPs^{recourse}$.

The framework moves to the next time period, and continues to generate and solve $DEVSPs$ and $DEVSPs^{recourse}$. At the next time period, the number of $DEVSPs$ generated also depends on whether there have been any uncertain parameter realizations or not. If there were none, one $DEVSP$ is constructed in which the here-and-now and recourse action decision variables of previous time period(s) are fixed. If there were uncertainty realization, a unique $DEVSP$ is constructed for each unique outcome. In each $DEVSP$, all decisions made at previous time periods are fixed, and the values of the remaining uncertain parameters are taken to be at their expected values.

The framework repeats the process of constructing and solving $DEVSPs$ and $DEVSPs^{recourse}$ until the end of the planning horizon. At the end of the planning horizon, it yields a feasible solution for the MSSP problem for all time periods. The progression of the framework along the planning horizon follows the traditional decision-making process under uncertainties, where a decision-maker uses the expected outcome for the future events and incorporates available information at

the current stage. Then, the decision maker starts to observe the uncertain events, and makes new decisions and takes recourse actions accordingly.

The formulations of $DEVSP$ and $DEVSP^{recourse}$, which are essential for the implementation of the framework, are introduced in the next section. Section 5.1.2 outlines in detail the algorithm used to generate AEEV using the framework.

5.2.1 The formulations of deterministic expected value sub-problems ($DEVSPs$) and recourse deterministic expected value sub-problems ($DEVSP^{recourse}$)

The model for the deterministic expected value sub-problem that is used in the framework is given in Eqns. (5.1) – (5.6).

$$min \sum_{i} \sum_{t'} G_{i,t'}^{est}\left(V_{i,t'}, \theta_i^n \in U_{n^{pre}}, \xi_{t'}^n \in U_{n^{pre}}, b_{i,t'}, y_{i,t'}, \gamma_{t'}\right) \tag{5.1}$$

$$h\left(b_{i,t'}, y_{i,t'}, \gamma_{t'}, \theta_i^n \in U_{n^{pre}}, \xi_{t'}^n \in U_{n^{pre}}\right) = 0 \quad \forall i \in \boldsymbol{I}, t' \in \boldsymbol{T} \tag{5.2}$$

$$g(b_{i,t'}, y_{i,t'}, \gamma_{t'}, \theta_i^n \in U_{n^{pre}}, \xi_{t'}^n \in U_{n^{pre}}) \leq 0 \quad \forall i \in \boldsymbol{I}, t' \in \boldsymbol{T} \tag{5.3}$$

$$b_{i,t'} \in F_{n^{pre}}\big|_{t'<t}, b_{i,t'} \in \{0,1\}\big|_{t' \geq t} \quad \forall t' \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \tag{5.4}$$

$$y_{i,t'} \in F_{n^{pre}}\big|_{t'<t}, y_{i,t'} \in \mathbb{R}\big|_{t' \geq t} \quad \forall t' \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \tag{5.5}$$

$$\gamma_{t'} \in F_{n^{pre}}\big|_{t'<t}, \gamma_{t'} \in \mathbb{R}\big|_{t' \geq t} \quad \forall t' \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \tag{5.6}$$

Each $DEVSP_{n,t}$ is defined for sub-problem n at time t. In $DEVSP_{n,t}$, n is the index for the sub-problem, n^{pre} represents its parent sub-problem, and t corresponds to the current time period. The scenario indices of the original MSSP model along with the NACs are removed to construct $DEVSP_{n,t}$. Because the NACs and scenario dependencies of the decision variables are removed, the indicator variable ($Z_t^{s,s'}$) and its logic expression are also dropped from the model. The values

of the uncertain parameters are set to their expected or realized values depending on the stage and time period the corresponding $DEVSP_{n,t}$ is generated. As shown in Eqns. (5.4)-(5.6), in $DEVSP_{n,t}$, all decision variables of the previous stages ($t' < t$) are fixed using $F_{n^{pre}}$, which stores the decisions taken at the previous sub-problems n^{pre}. The solution of $DEVSP_{n,t}$ determines the here-and-now decision variable ($b_{i,t}, y_{i,t}$) values for $t' \geq t$. The values of uncertain parameters in $DEVSP_{n,t}$ are updated with $U_{n^{pre}}$, which stores the uncertain parameter values. In $U_{n^{pre}}$, if uncertain parameters are already observed, the values are set to their realized values, otherwise, they are set to their expected values.

A set of recourse deterministic expected value sub-problems ($DEVSP^{recourse}_{n^{child},t}$) are generated based on the realized outcomes of uncertain parameters as shown in Eqns. (5.7)-(5.12). The formulation of $DEVSP^{recourse}_{n^{child},t}$ are similar to $DEVSP_{n,t}$.

$$min \sum_{i} \sum_{t'} G^{est}_{i,t'}\left(V_{i,t'}, \theta^{n}_{i} \in U_{n^{child}}, \xi^{n}_{t'} \in U_{n^{child}}, b_{i,t'}, y_{i,t'}, \gamma_{t'}\right) \tag{5.7}$$

$$h\left(b_{i,t'}, y_{i,t'}, \gamma_{t'}, \theta^{n}_{i} \in U_{n^{child}}, \xi^{n}_{t'} \in U_{n^{child}}\right) = 0 \quad \forall i \in \boldsymbol{I}, t' \in \boldsymbol{T} \tag{5.8}$$

$$g(b_{i,t'}, y_{i,t'}, \gamma_{t'}, \theta^{n}_{i} \in U_{n^{child}}, \xi^{n}_{t'} \in U_{n^{child}}) \leq 0 \quad \forall i \in \boldsymbol{I}, t' \in \boldsymbol{T} \tag{5.9}$$

$$b_{i,t'} \in F_n\big|_{t' \leq t}, b_{i,t'} \in \{0,1\}\big|_{t' > t} \quad \forall t' \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \tag{5.10}$$

$$y_{i,t'} \in F_n\big|_{t' < t}, y_{i,t'} \in \mathbb{R}\big|_{t' > t} \quad \forall t' \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \tag{5.11}$$

$$\gamma_{t'} \in F_n\big|_{t' < t}, \gamma_{t'} \in \mathbb{R}\big|_{t' \geq t} \quad \forall t' \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \tag{5.12}$$

At time period t, after solving the $DEVSP_{n,t}$, new child sub-problems, $DEVSP^{recourse}_{n^{child},t}$, are generated if there are uncertainty realizations. In each child sub-problem, the values of realized uncertain parameters in $U_{n^{child}}$ are updated to their corresponding outcomes, the here-and-now

decisions ($b_{i,t'}, y_{i,t'}$) are fixed for time periods $t' \leq t$, and the recourse actions are fixed for time periods $t' < t$. The framework obtains recourse actions for time period t by solving these problems, $DEVSP^{recourse}_{n^{child},t}$.

5.2.2 An algorithmic implementation of the systematic framework

The steps implementing the proposed systematic framework is given in Figure 5.2.2. The algorithm starts by initializing n, N_t, U_n, and F_n. The variable N_t is a set that monitors and stores the number and indices of sub-problems at time t. At initialization, there is only one sub-problem, so N_1 is initialized to $\{0\}$. The variable $n \in N_t$ represents the specific sub-problems in N_t and is equal to 0 at $t = 1$. The set U_n for subproblem n contains the values of the uncertain parameters, and hence, U_0 is initialized to the expected values of all uncertain parameters. In the algorithm (Figure 5.2.2), the variable F_n represents the fixed decision variables, which include the values of all here-and-now decision variables and recourse actions for sub-problem n up to time t. The variable F_n is a dictionary, where the key represents the indices of decision variables and the value of the key represents the value of the corresponding decision variable. At the beginning ($t = 1$), this variable is empty as there are no fixed decision variables. At $t = 1$ (root node), one deterministic expected value sub-problem, $DEVSP_{0,1}$, is constructed. As no uncertainty is realized yet, all uncertain parameter values are set equal to their expected values in $DEVSP_{0,1}$. The solution of $DEVSP_{0,1}$ determines the values of here-and-now decisions at $t = 1$, and these values are stored in variable F_n, in this case, F_0.

For $t = 1$ to T do
    If $t == 1$ then
        Set n: $=$ '0'
        Initialize sub-problem monitor $N_t = \{'0'\}$
        Initialize uncertain parameter value set U_n
        Initialize fixed decision dictionary F_n
        Generate and solve sub-problem $DEVSP_{n,t}(U_n, F_n)$
        Obtain here-and-now decisions for t: $F_n \leftarrow$ solution of $DEVSP_{n,t}(U_n, F_n)$
    Else
        For n in N_{t-1} do
            Use solution F_n to determine the realized outcomes
            Determine the number of new sub-problems and add their indices to N_t
            Update uncertain parameter value set U_n with realized outcomes
        End
        If there is no realizations
            $N_t = N_{t-1}$
        For n in N_t do
            If n not in N_{t-1} then
                Retrieve parent sub-problem indices n^{pre}
            Else
                $n^{pre} \leftarrow n$
            If recourse actions are needed then
                Generate and solve sub-problem $DEVSP_{n,t-1}^{recourse}(U_n, F_{n^{pre}})$
                Obtain recourse decisions for $t-1$: $F_n \leftarrow DEVSP_{n,t-1}^{recourse}(U_n, F_{n^{pre}})$
            Generate and solve sub-problem $DEVSP_{n,t}(U_{n^{pre}}, F_{n^{pre}})$
            Obtain here-and-now decisions for t: $F_n \leftarrow$ solution of $DEVSP_{n,t}(U_n, F_{n^{pre}})$
        End
End

Figure 5.2.2: Steps of the proposed primal-bounding framework for MSSP models under endogenous uncertainties

Next, at $t = 2$, the algorithm first determines whether there have been any uncertainty realizations, and then determines how many sub-problems should be generated based on the realized outcomes of uncertain parameters, and updates N_2 (Figure 5.2.2). Similar to MSSP, where here-and-now decisions are used to determine the values of the Boolean variable, the generation of sub-problems in the framework is based on the here-and-now decisions of the previous stages.

Each sub-problem has an updated uncertain parameter value set U_n based on realized values of uncertain parameters. If an uncertain parameter is observed, the value of that uncertain parameter in U_n takes its realized value for sub-problem n. For all other uncertain parameters, U_n contains their expected values.

After the here-and-now decisions and recourse actions are both obtained for the first time period, the algorithm continues by generating and solving $DEVSP_{n,t}(U_n, F_{n^{pre}})$ at $t = 2$. The overall process is repeated until the end of planning horizon. Finally, the framework generates the AEEV and evaluates the objective function at the AEEV ($OAEEV$). At termination, the calculated AEEV is a valid primal bound for the original MSSP objective function value if the MSSP has complete recourse. The framework calculates the recourse variables by solving recourse deterministic expected value sub-problems ($DEVSP^{recourse}_{n^{child},t}$) based on the realized outcomes of uncertain parameters in every step. However, for problems without complete recourse, the solution generated by the framework may be infeasible, though the probability of obtaining infeasible solution is limited. Below Proposition shows that the bounds obtained by AEEV is guaranteed to be better than EEV_T.

Definition 1 Let us define the *absolute value of the stochastic solution in a maximization problem*, say *AVSS*, as

$$AVSS = OV - OAEEV.$$

Proposition 1 *For any multistage stochastic program with complete recourse,*

$$0 \leq AVSS \leq VSS_T$$

Proof The *AEEV* is a feasible solution for the original MSSP, then $OAEEV \leq OV \Longrightarrow (OV - OAEEV) \geq (OV - OV) \geq 0$. Thus $0 \leq AVSS$. $VSS_T = OV - EEV_T$, where EEV_T takes expected value of all random variables and is equal to solution of $DEVSP_{0,1}$. At the beginning, $DEVSP_{0,1}$

uses the expected values for all random variables. After obtaining solutions at $t = 1$ in $DEVSP_{0,1}$, the framework generates the new sub-problem n and re-optimizes $DEVSP_{n,2}$ and the expected objective function value in the second stage will surely be larger than $DEVSP_{0,1}$. Thus, $EEV_T \leq OAEEV \Longrightarrow (OV - EEV_T) \geq (OV - OAEEV) \Longrightarrow VSS_T \geq AVSS$.

5.3 Application of AEEV to example planning problems

We generate a valid primal bound by applying the framework to three different problems: pharmaceutical R&D pipeline planning problem (Colvin and Maravalias 2008), process network synthesis problem (Tarhan and Grossmann 2008), and artificial lift infrastructure planning problem (Zeng and Cremaschi 2017a). The following sections outline the corresponding deterministic expected value sub-problem ($DEVSP$) formulations and the details of algorithm's implementation for these problems.

5.3.1 Clinical trial planning problem

Following the approach explained in previous section, the deterministic models of clinical trial planning problem, $DEVSP_{n,t}$, are generated as in Eqns. (5.13) – (5.17):

$$
\begin{aligned}
max \sum_i P_i \sum_t \Bigg[& rev_i{}^{max} X_{i,PIII,t} \\
& - \gamma_i{}^{D} \sum_{j=PII,PIII} \left(-X_{i,j,1} + \sum_{t' > \tau_{i,j-1}}^{t' \le t} X_{i,j-1,t'-\tau_{i,j-1}} - \sum_{t'}^{t' \le t} X_{i,j,t'} \right) \\
& - \gamma_i{}^{L} \left(t + \tau_{i,PIII}\right) X_{i,PIII,t} \Bigg] \\
& + \sum_i P_i \sum_{j \neq PI} rev_{i,j}^{open} f_{i,j} \left(-X_{i,j,1,s} + \sum_{t' > \tau_{i,j-1}}^{t' \le |T|} X_{i,j-1,t'-\tau_{i,j-1}} \right. \\
& \left. - \sum_{t'}^{t' \le |T|} X_{i,j,t'} \right) + \sum_i P_i rev_{i,PI}^{open} f_{i,PI} \left(1 - \sum_{t'}^{t' \le |T|} X_{i,PI,t'} \right) \\
& + \sum_i \sum_{j \in \{PI,PII\}} \sum_{t > |T| - \tau_{i,j}} rev_{i,j,t}^{run} f_{i,j+1} X_{i,j,t} - \sum_{i,j,t} P_i cd_t C_{i,j} X_{i,j,t}
\end{aligned}
\tag{5.13}
$$

$$
\sum_t X_{i,j,t} \le 1 \quad \forall i, j \tag{5.14}
$$

$$
\sum_{t'}^{t} X_{i,j,t'} \le \sum_{t'}^{t-\tau_{i,j-1}} X_{i,j-1,t'} \quad \forall i, j > 1, t' \tag{5.15}
$$

$$
\sum_i \sum_j \sum_{t' > t-\tau_{i,j}}^{t' \le t} \rho_{i,j,r} X_{i,j,t'} \le \rho_r{}^{max} \quad \forall r, t \tag{5.16}
$$

$$
X_{i,j,t} \in F_n \quad \forall i, j, t \tag{5.17}
$$

In the deterministic models ($DEVSP_{n,t}$), the parameter P_i represents the probability of drug i successfully completing all three clinical trials, and, therefore, the objective function considers the expected profit for each drug. In the clinical-trial planning MSSP model, there are no uncertain

parameters in the scenario specific constraints. The realized uncertain parameters are the outcomes of completed clinical trials. For failure outcomes, the values of the decision variables associated with subsequent trials are set to zero and stored in set F_n in child $DEVSPs$ reflecting the impact of the realized outcomes. For success outcomes, no such action is required. It should be noted that this planning problem has no recourse actions. Thus, as discussed in previous sections, generating and solving $DEVSPs$ are sufficient to obtain the AEEV.

5.3.2 Process network synthesis

The schematic of a simple process network has been given in Figure 4.2.2. The deterministic sub-problems ($DEVSP_{n,t}^{synthesis}$) obtained from the MSSP formulation are given in Eqns. (5.18)-(5.37).

$$max -\sum_{t}\sum_{i}\left(FE_i y_{i,t}^{exp} + FO_i y_{i,t}^{oper} + VE_i w_{i,t}^{QE}\right) - \sum_{t}\sum_{k} VO_k w_{k,t}^{rate} - \sum_{t}\left(\alpha x_t^{purch} - \beta x_t^{sales} + \gamma w_t^{inv}\right) \tag{5.18}$$

$$w_{3,\mathrm{t}}^{rate} = \theta_1 w_{1,\mathrm{t}}^{rate} \quad t \in T \tag{5.19}$$

$$w_{4,\mathrm{t}}^{rate} = \theta_2 w_{2,\mathrm{t}}^{rate} \quad t \in T \tag{5.20}$$

$$w_{8,\mathrm{t}}^{rate,} = \theta_3 w_{7,\mathrm{t}}^{rate} \quad t \in T \tag{5.21}$$

$$w_{5,\mathrm{t}}^{rate} = w_{3,\mathrm{t}}^{rate} + w_{4,\mathrm{t}}^{rate} \quad t \in T \tag{5.22}$$

$$w_{7,\mathrm{t}}^{rate} = w_{5,\mathrm{t}}^{rate,s} + w_{6,\mathrm{t}}^{rate} \quad t \in T \tag{5.23}$$

$$w_t^{inv} = w_{t-1}^{inv} + (w_{8,\mathrm{t}}^{rate} + x_t^{purch} - x_t^{sales}) \quad t \in T \tag{5.24}$$

$$x_t^{sales} = d_t \quad t \in T \tag{5.25}$$

$$w_{i,t}^{cap} = w_{i,t-1}^{cap} + w_{i,t}^{QE} \quad t \in T \tag{5.26}$$

$$w_{3,\mathrm{t}}^{rate} \leq w_{1,\mathrm{t}}^{cap} \quad t \in T \tag{5.27}$$

$$w_{4,\mathrm{t}}^{rate} \leq w_{2,\mathrm{t}}^{cap} \quad t \in T \tag{5.28}$$

$$w_{8,\mathrm{t}}^{rate} \leq w_{3,\mathrm{t}}^{cap} \quad t \in T \tag{5.29}$$

$$L_i^{QE} y_{i,t}^{exp} \leq w_{i,t}^{QE} \leq U_i^{QE} y_{i,t}^{exp} \quad t \in T \tag{5.30}$$

$$L_1^{inflow} y_{1,\mathrm{t}}^{oper} \leq w_{3,\mathrm{t}}^{rate} \leq U_1^{inflow} y_{1,\mathrm{t}}^{oper} \quad t \in T \tag{5.31}$$

$$L_2^{inflow} y_{2,\mathrm{t}}^{oper} \leq w_{4,\mathrm{t}}^{rate} \leq U_2^{inflow} y_{2,\mathrm{t}}^{oper} \quad t \in T \tag{5.32}$$

$$L_3^{inflow} y_{3,\mathrm{t}}^{oper} \leq w_{8,\mathrm{t}}^{rate} \leq U_3^{inflow} y_{3,\mathrm{t}}^{oper} \quad t \in T \tag{5.33}$$

$$\sum_{tt}^{t} y_{i,tt}^{exp} \geq y_{i,t}^{oper} \quad t \in T, i \in [1,2] \tag{5.34}$$

$$y_{i,t}^{oper} \geq y_{i,t}^{exp} \quad t \in T, i \in [1,2] \tag{5.35}$$

$$\theta_1, \theta_2 \in U_n \tag{5.36}$$

$$y_{i,t'}^{oper}, y_{i,t'}^{exp}, w_{\mathrm{k},t'}^{rate}\Big|_{t'<t} \in F_n \quad t' \in T, i \in I, k \in [1,2,6] \tag{5.37}$$

The $DEVSP_{n,t}^{recourse}$ models for the process synthesis problem has the same formulations as $DEVSP_{n,t}^{synthesis}$ where the values of the here-and-now decision variables are fixed to their corresponding values. By solving the $DEVSP_{n,t}^{synthesis}$, the here-and-now decisions at time t ($y_{i,t}^{exp}, y_{i,t}^{oper}, w_{k,t}^{rate,s} \in \mathbb{R}^{\geq 0}, k = \{1,2,6\}, w_{i,t}^{QE,s} \in \mathbb{R}^{\geq 0}$) are obtained. After fixing the here-and-

now decisions and solving the same deterministic model, the values of the recourse action decision variables at time t ($x_t^{purch}, x_t^{sales}, w_t^{inv}, w_{k,t}^{rate} \in \mathbb{R}^{\geq 0}, k = \{3,4,5,7,8\}$) are obtained.

The algorithm starts by taking the expected value of the uncertain yields, 0.7, for both Process 1 and 2 for constructing the root-node $DEVSP_{0,1}$ ($n = 0, t = 1$). From the solution of $DEVSP_{0,1}$, the here-and-now decisions ($y_{i,1}^{exp}, y_{i,1}^{oper}, w_{k,1}^{rate,s} \in \mathbb{R}^{\geq 0}, k = \{1,2,6\}, w_{i,1}^{QE,s} \in \mathbb{R}^{\geq 0}$) for $t = 1$ are determined. Based on the values of the here-and-now decision variables, specifically operating decisions $y_{i,t}^{oper}$, the algorithm next checks if any uncertain parameters are realized. If there are realizations, sub-problem monitor for next time period N_2 is updated according to the number of realizations, and the values of uncertain parameters (yields) are updated in $U_{n \in N_2}$. If there are no realizations, the sub-problem monitor is set as $N_2 = N_1$ for the next time period. The corresponding sub-problem(s) are generated for realized uncertain parameters, and the solution(s) of these problem(s) yield the recourse actions for the corresponding outcome(s). At the next time period, $t = 2$, all the variables of the previous time periods are fixed and realized uncertain parameter values are passed to the sub-problems. The algorithm terminates at the end of planning horizon.

5.3.3 Artificial lift infrastructure planning problem (ALIP)

The $DEVSP$ model for ALIP is given in Eqns. (5.38) - (5.55). The decision variable in $DEVSP_{n,t}^{ALIP}$ is $w_{i,p}$, which is obtained by removing the scenario index of the original binary decision variable $w_{i,p,s}$.

$$max \sum_r \left\{ (GI_r - TI_r \cdot x_r \cdot FT) \left(\frac{1}{(1 + MARR)^r} \right) \right\} - CC(1 - FT) \quad \forall r \in T \tag{5.38}$$

$$GI_r = \left(Rev_r(1-RT)(1-LT) - \sum_{i,t\geq r,p\leq r} \left(Cm_i y_{i,t,p}\right) \right) WI \quad \forall r \in T \tag{5.39}$$

$$Rev_r = P_g Qg_r + P_o Qo_r + P_{ng} Qng_r \quad \forall r \in T \tag{5.40}$$

$$TI_r = GI_r - Dep_r \quad \forall r \in T \tag{5.41}$$

$$TI_r \leq x_r M \quad \forall r \in T \tag{5.42}$$

$$TI_r > (x_r - 1)M \quad \forall r \in T \tag{5.43}$$

$$Dep_r = \sum_{i,t\geq r,p\geq r-n+1}^{p\leq r} \left(\frac{Ce_i}{n} y_{i,t,p}\right) \quad \forall t,p,r \in T, i \in I \tag{5.44}$$

$$CC = \sum_{i,t,p} \left(Co_i y_{i,t,p} \frac{1}{(1+MARR)^p} \right) \quad \forall t,p \in T, i \in I \tag{5.45}$$

$$Qg_r = \sum_{i,t,p} \left\{ y_{i,t,p} Qg_{p-1} Qrc_i \left(1 + bD(r-p+1)\right)^{-\frac{1}{b}} \right\} \tag{5.46}$$

$$Qo_r = \sum_{i,t,p} \left\{ y_{i,t,p} Qo_{p-1} Qrc_i \left(1 + bD(r-p+1)\right)^{-\frac{1}{b}} \right\} \tag{5.47}$$

$$Qng_r = \sum_{i,t,p} \left\{ y_{i,t,p} Qng_{p-1} Qrc_i \left(1 + bD(r-p+1)\right)^{-\frac{1}{b}} \right\} \tag{5.48}$$

$$f_{i,t,p,r}\left(LFR_r, y_{i,t,p}\right) \leq 0 \quad \forall i,t,p,r \tag{5.49}$$

$$g_{i,t,p,s}\left(y_{i,t,p}\right) \leq 0 \quad \forall i,t,p \tag{5.50}$$

$$y_{i,\mathrm{t,p}} \geq z_{i,\mathrm{t}} + w_{i,\mathrm{p}} - 1 \quad \forall i,t,p \tag{5.51}$$

$$z_{i,\mathrm{t}} \geq y_{i,\mathrm{t,p}} \quad \forall i,t,p \tag{5.52}$$

$$w_{i,\mathrm{p}} \geq y_{i,\mathrm{t,p}} \qquad \forall i, t, p \tag{5.53}$$

$$Qg_{p-1}, Qo_{p-1}, Qng_{-1} \in U_n \tag{5.54}$$

$$w_{i,\mathrm{p}}\big|_{p<t} \in F_n \quad p \in T, i \in I \tag{5.55}$$

The algorithm starts by taking the expected value of the uncertain gas flow rate change ratios for ALMs for constructing $DEVSP_{0,1}$ ($n = 0, t = 1$), solution of which yields the here-and-now decisions, $w_{i,\mathrm{p}}$. Based on the values of the here-and-now decision variables, the algorithm updates the uncertain parameter set, generates and solves sub-problems to determine recourse actions, and increments the time. The algorithm terminates at the end of planning horizon.

5.4 Results and Discussion

We solved the MSSP models of all three problems to optimality and applied the proposed framework to obtain the AEEVs and the corresponding objective function values. All models and algorithms were implemented in Pyomo 5.5.0 and solved using CPLEX 12.6.3 on a standard node of Auburn University Hopper Cluster. Data for all instances are available upon request from the corresponding author.

5.4.1 Clinical trial planning problem

We applied the primal-bounding framework to obtain AEEVs and the corresponding objective function values (i.e., primal bounds) for instances of clinical-trial planning problem with three-, four-, five-, six- and seven-drugs, and three clinical trials with different planning horizon lengths. We also solved the deterministic equivalents of the MSSP models for these instances to 0.1% optimality gap when possible. Table 5.1 summarizes the number of trials, time periods, and

scenarios for these instances along with the optimum ENPVs and the solution times for the instances that were solved to optimality using the MSSP formulation.

Table 5.1: The number of trials, time periods and scenarios for clinical-trial planning instances and the optimal solutions of these instances (optimality gap = 0.1%) where available.

Case Name	Number of Trials	Number of Time Periods	Number of Scenarios	ENPV ($M)	MSSP Solution Time (CPUs)
3-drug	3	12	64	1193	12
4-drug	3	6	256	1700	20
5-drug	3	6	1,024	2087	1776
6-drug	3	6	4,096	2460	104,541
7-drug	3	8	16,384	--	--

For the clinical-trial planning problem instances, the ENPVs of the AEEVs, the percent gaps from the optimum solutions (when available), and the algorithm times to generate the AEEVs and the corresponding objective function values are compiled in Table 5.2. The framework yields very tight primal bounds, within 0.75% of the true solution, for all instances of the clinical-trial planning problem where we were able to solve the original MSSP. For the 4-drug case, the AEEV is optimal. As can be seen from Table 5.2, the primal-bounding framework obtains the AEEVs relatively quickly with solution times up to four orders of magnitude faster than the solution times of the original MSSP models. The MSSP model for the 6-drug case was solved to optimality in about 20 CPU-hours, while our framework took about one CPU-minutes to generate the AEEV with an objective function value that is within 0.65% of the true solution. The MSSP model of the 7-drug case cannot be generated in a standard node of the Auburn University Hopper Cluster with 128GB of memory using CPLEX 12.6.3. In contrast, the framework generated the AEEV and calculated its objective function value in about 17 CPU-minutes. By applying the relaxed knapsack-problem

based decomposition (RKDA) (Zeng and Cremaschi, 2018c), we obtained a dual bound for the 7-drug case, which is $3008 million. The relative gap between the objective function value of AEEV (primal bound) and the dual bound (RKDA) is 3.82%. It is worth noting that the solution times for the framework can be further reduced by solving the deterministic sub-problems in parallel after the root node.

Table 5.2: The ENPVs, solution times and relative gaps of AEEV for the clinical-trial planning instances. Percent gaps are calculated using the optimum ENVPs of these instances (optimality gap = 0.1%) where available

	AEEV		
Case Name	ENPV (million $)	Percent Gap from the Optimum ENVP	Solution time (CPUs)
3-prod	1184	0.75%	7
4-prod	1700	0.00%	6
5-prod	2078	0.43%	10
6-prod	2444	0.65%	62
7-prod	2893	NA	987

We compare the AEEV and the optimum-solution decision trees in detail for the 3-drug case. This case has three candidate drugs (D1, D2 and D3) in the R&D pipeline, and a 36-month planning horizon divided into 12 three-month time periods. Each drug has to pass all three trials (PI, PII and PIII) before its revenue can be realized. Figure 5.2.3 and 5.2.4 depict the optimum decision tree and the decision tree of the AEEV, respectively. The optimal solution has an ENVP of $1,193 million and recommends starting first clinical trial of drug D1, i.e., (D1, PI), at the first time period, $t = 1$, and (D3, PI) at the second time period, $t = 2$. The outcome of (D1, PI) is realized at $t = 3$. If drug D1 successfully completes the first clinical trial, the solution recommends

starting the second trial for drug D1 (D1, PII) at $t = 3$. The ENPV of AEEV is \$1,184 million and the AEEV has a very similar decision tree (Figure 5.2.3) when compared to the optimum one given in Figure 5.2.4. For the first four time periods, the decision trees are identical for both solutions. Both MSSP and AEEV recommend starting Drug 1 Phase I (D1, PI) clinical trial at the first time period, and then starting Drug 3 Phase I (D3, PI) at the second time period. The differences lie in decisions after the failure outcomes are observed. For example, at time period 5 ($t = 5$), in the case where the outcomes of both (D1, PI) and (D3, PI) are failures, the AEEV recommends starting (D2, PI) while the optimal solution recommends not starting any other clinical trials at that time period. For the remainder of the planning horizon, the AEEV decision tree is identical to the optimum one for the scenarios where the outcomes of clinical trials are successes, but may differ when the outcomes are failures.

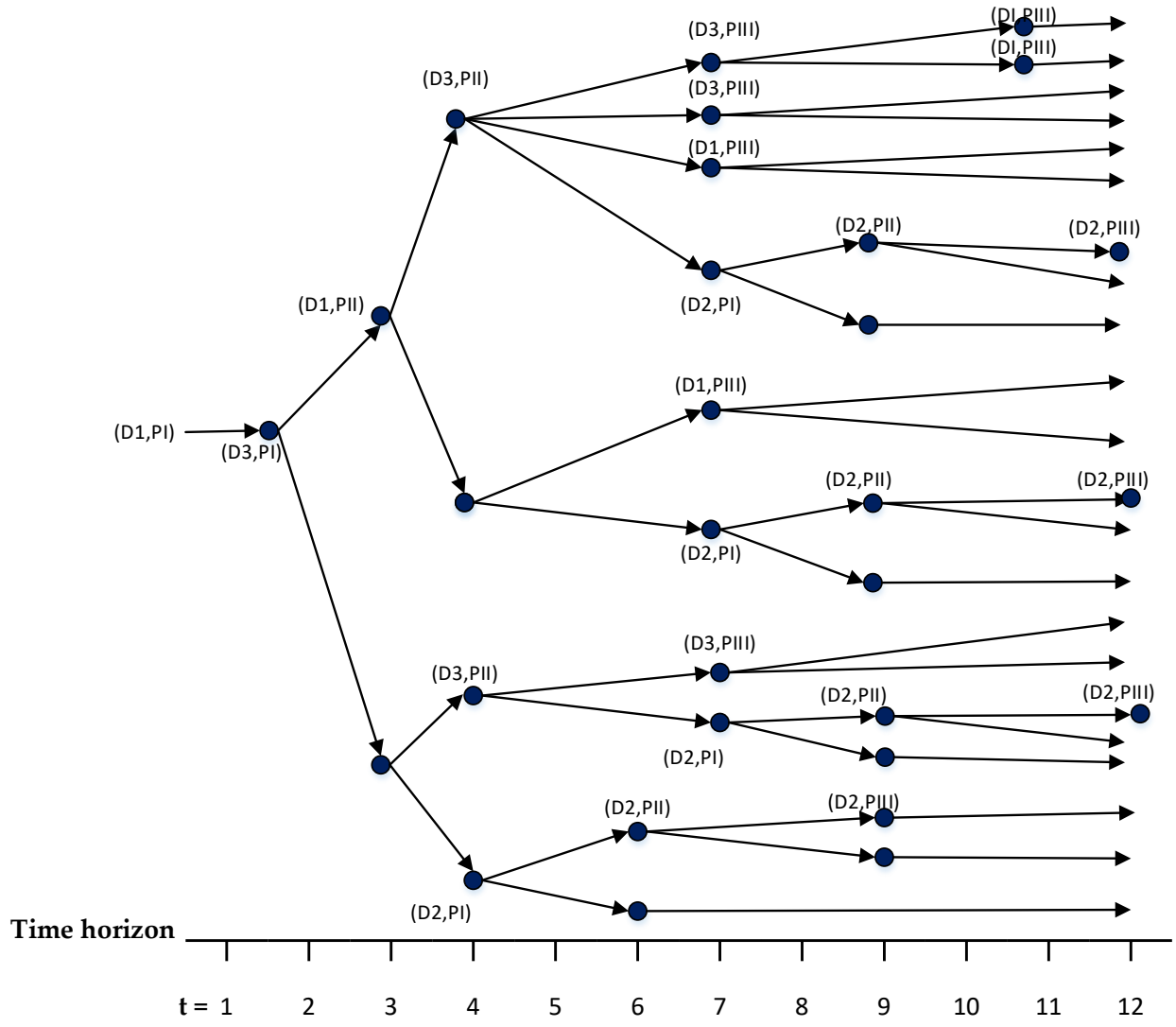

Figure 5.2.3: The optimum decision tree for the clinical trial planning problem of the 3-Drug instance

We also compared the decision trees for 4-, 5-, and 6-drug instances. The comparisons revealed that the first time-period decisions of the AEEV were identical to the first-time period decisions of the optimum solutions for all cases. Furthermore, for sub-problems that were generated for the success outcomes of the started clinical trials, the AEEV and optimum decision trees were identical. The solutions only differed for some of the failure outcomes.

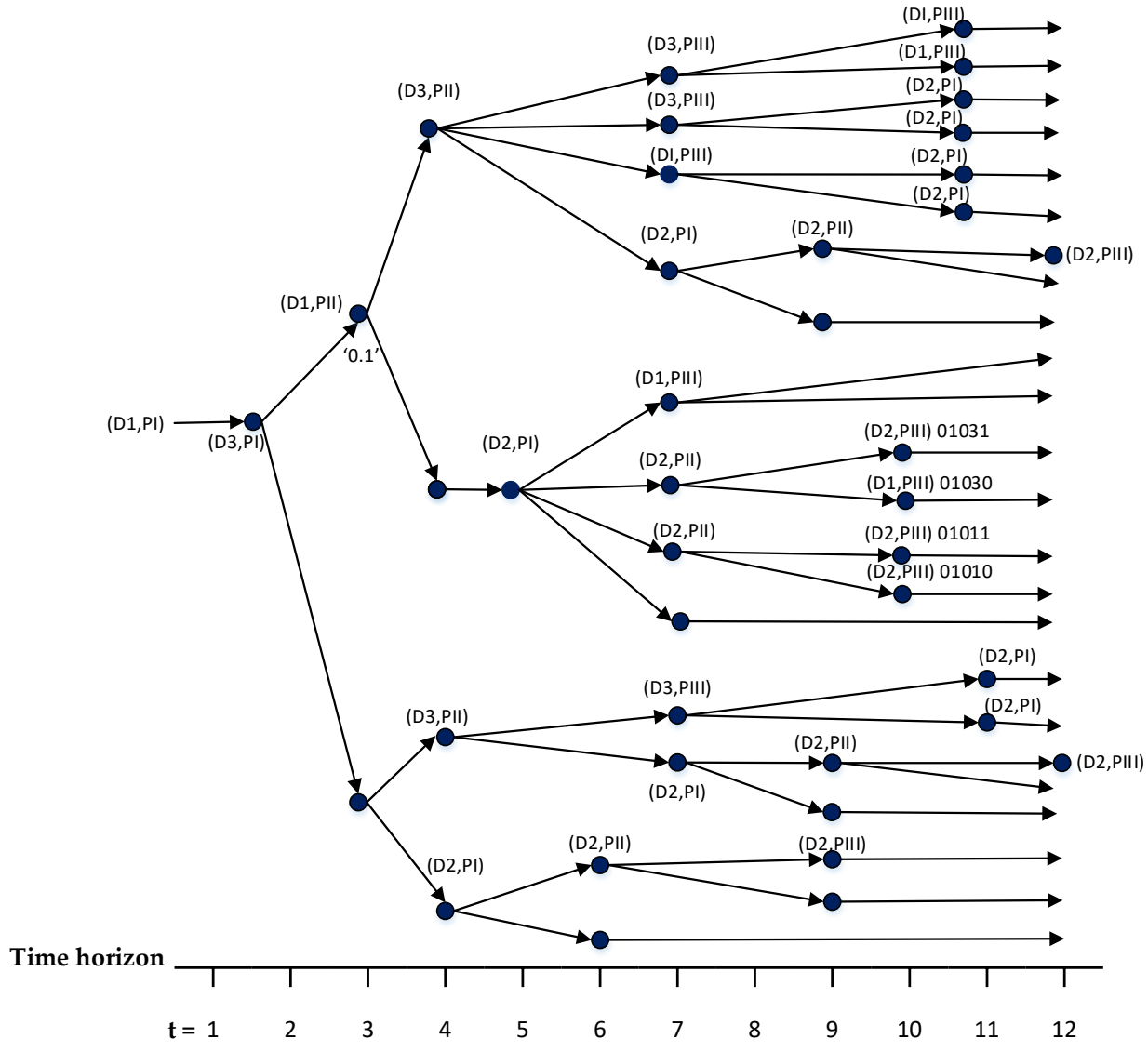

Figure 5.2.4: The AEEV decision tree for the clinical trial planning problem of 3-Drug instance

5.4.2 Process network synthesis

The problem instance solved is derived from Goel and Grossmann (2006) and has been described in Chapter 4. The planning horizon is ten years discretized into one-year time periods. The realizations for the yield of Process 1 are 0.69 and 0.81, and of Process 2 are 0.65 and 0.85 both with equal probabilities. Process 3 has an initial capacity of 3×10^6 tons/year and a fixed yield of 0.7.

The deterministic equivalent of the MSSP contains 400 binary variables, 685 continuous variables, and 5753 constraints, which is a relatively small MILP, and CPLEX 12.6.3 solves it in one CPU-second. The optimum ENPV is \$11.459×106. The optimum solution recommends installing and operating Process 2 at $t = 1$ with a capacity of 10×10^6 tons/year and expanding the capacity of Process 3 to 8×10^6 tons/year at $t = 3$ for scenarios 1 and 3 and at $t = 4$ for scenarios 2 and 4. In scenarios 2 and 4, the yield of Process 2, which is now observed, is 0.81, i.e., the high outcome. The higher yield from Process 2 results in extra inventory for the product. Therefore, the optimal solution for these scenarios recommend capacity expansion at a later time period compared to the recommendations for scenarios 1 and 3.

The ENPV of AEEV is \$11.451×106, which is within 0.06% of the optimum solution. The framework obtains this solution within 3 CPU seconds. The expansion and operation decisions of the AEEV are identical to the optimum solution. The small difference in the ENPV is due to slight differences in inventory levels and inflow rate decisions.

5.4.3 Artificial lift infrastructure planning problem

The problem solved has been introduced in chapter 4 and with four planning horizons, 12, 16, 20 and 24 months. We discretize the planning horizon into one-month long time periods. The maximum time for obtaining a solution is limited to 20 CPU-hours. Table 5.3 summarizes the optimum ENPVs and the ENPVs of the AEEV for all planning horizons along with the solution times. The decision trees of the AEEV, which are identical for the optimum and the AEEV for 12 and 16-month planning horizons, are presented in Figure 5.2.5.

Table 5.3: The optimum ENPVs, MSSP solution times, ENPVs of the AEEV solutions and AEEV solution times for hypothetical ALIP instances with different planning horizons

	MSSP			AEEV	
Planning Horizon (months)	ENPV ($)	Gap	Solution Time (CPUs)	ENPV ($)	Solution Time (CPUs)
12	2,783,894	0.0%	21	2,783,894	79
16	3,416,506	0.0%	4659	3,416,506	217
20	3,916,996	11.9%	72000	3,943,560	945
24	4,240,937	31.1%	72000	4,359,789	2824

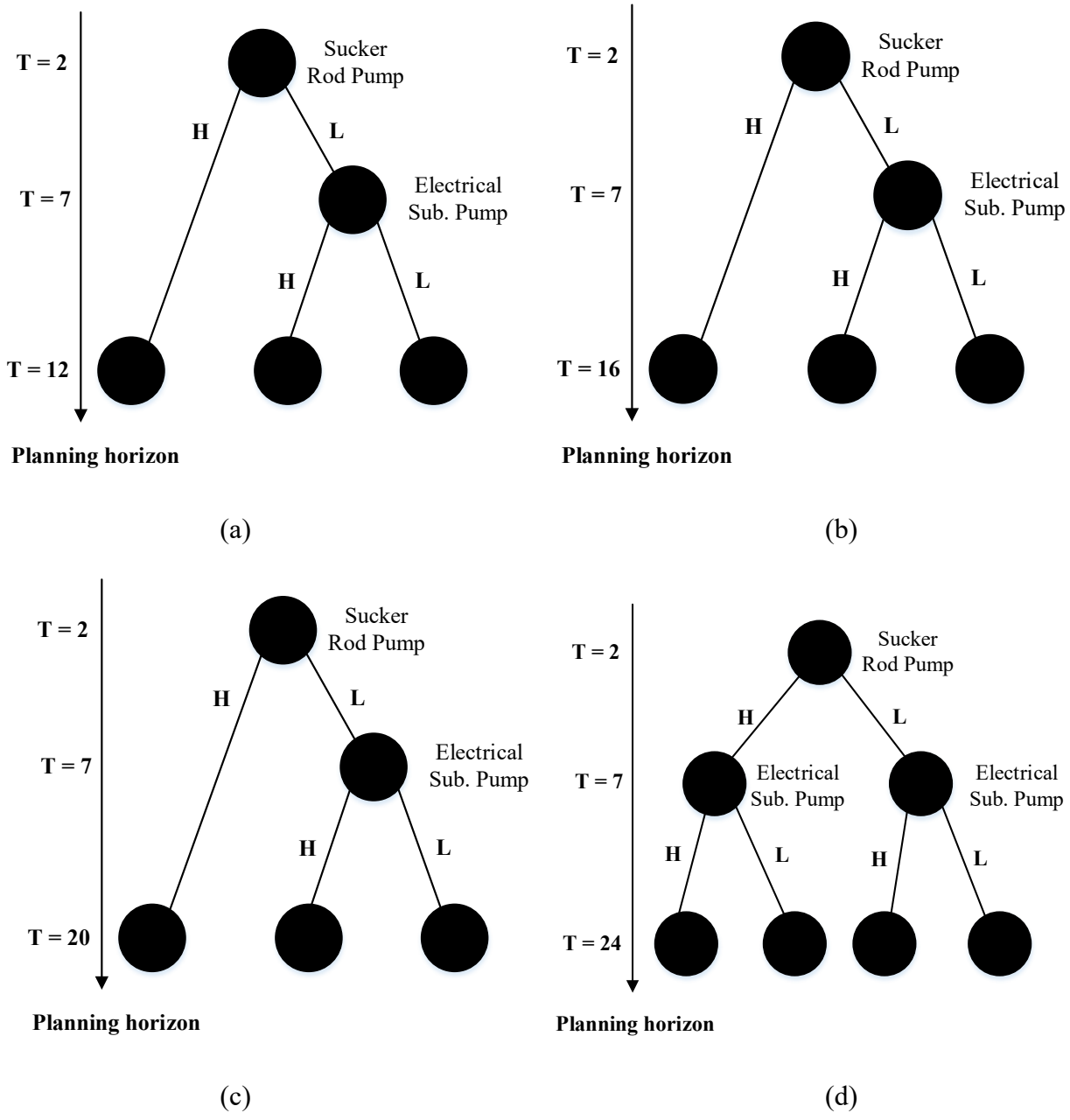

Figure 5.2.5: Artificial lift infrastructure plan suggested by the AEEV for planning horizons of (a)12, (b) 16, (c) 20, and (d) 24 months. The plans suggested by the AEEV are optimal for 12- and 16-month planning horizons

For 12- and 16-month planning horizon cases, the primal-bounding framework yields the optimal solutions as can be seen from Table 5.3. The optimum solution recommends installing a SRP at $t = 2$ on the well for the 12- and 16-month planning horizons. If the production rate of

SRP is realized to be the High outcome, then the solution recommends operating SRP until the end of the planning horizon. On the other hand, if the realized production rate of SRP is Low, then it is replaced by ESP at $t = 7$. It should be noted that for a 12-month planning horizon case, the solution time for the AEEV is larger than the solution time for the original MSSP because the framework has to repeatedly solve the deterministic model at every time period. However, there is a significant difference in solution times for the 16-month planning horizon as the MSSP and its solution time grows exponentially, whereas the sub-problem sizes and the solution times for the framework do not grow as fast. Therefore, in 20- and 24-month planning horizon cases, the MSSP can not be solved to optimality within 20 CPU hours. The AEEV obtain feasible solutions and primal bounds for both cases within one hour.

5.5 Selection of GKDA and AEEV

In Chapter 4, we introduce a new Generalized Knapsack-problem-based Decomposition Algorithm, GKDA, to quickly obtain feasible solutions and to decompose large-scale multistage stochastic programming (MSSP) models into a series of knapsack subproblems (KSP). The GKDA decomposes time periods and scenarios of an MSSP model and approximates the contribution of the decision variables associated with that item to the objective function when KSPs are generated.

AEEV generates and solves a series of deterministic problems based on the traditional decision-making process, which utilizes the expected values for unrealized information, exploits all known information, and fixes here-and-now decisions for the current time only.

GKDA takes advantage of removing both time and scenario indices and addresses both time and space complexities associated with solving large-scale MSSPs under endogenous uncertainties. However, due to its greedy nature, it fails to capture the impact of the decisions, which have immediate negative short-term consequences at the current time period but have long-

term benefits on future time periods. For example, inventory and expansion decisions have negative impacts on the net present value within a range of short time period. In the GKDA applications of process network synthesis problem, the solutions never recommend holding inventory due to the negative item values of inventory decisions. The item values generated in KSPs fail to correctly assess the potential benefit of inventory levels and capacity expansions. Therefore, the GKDA solution has a 20% gap from the optimum.

Unlike the GKDA, which removes both time and scenario indices, AEEV only removes scenario indices and generates deterministic subproblems. AEEV determines the here-and-now decisions for the current time period and also considers their impact on the whole planning horizon. However, if the deterministic problems require significant computational effort to solve, we recommend GKDA for full decomposition.

5.6 Conclusion

In this chapter, we proposed a general primal-bounding framework and an algorithm to implement this framework for large-scale multistage stochastic programming (MSSP) problems that involve endogenous uncertain parameters. For MSSP models under exogenous uncertainties, we can take advantage of fixed scenario trees and develop bounding approaches based on solving expected value problems. However, to the best of our knowledge, this paper is the first one that explores and extends these concepts for MSSP models with endogenous uncertain parameters. By generating and solving a series of deterministic expected value models, the framework obtains the absolute expected value solution (AEEV), a feasible and implementable solution for MSSP problems with complete recourse. The objective function evaluated at the AEEV is shown to provide tight primal bounds for MSSP models under endogenous uncertainties based on the results of the case studies. It should be noted that the AEEV is different from the expected value solution

of deterministic model EEV, where all random variables are fixed at their expected values and the original MSSP is solved by fixing all here-and-now decisions to the solution of the deterministic model. The AEEV is generated based on the traditional process of decision-making facing uncertainties, which utilizes the expected values for unrealized information, exploits all known information, and fixes here-and-now decisions for the current time only.

We apply the framework to generate the AEEVs of three planning problems and compare the objective function values of the AEEVs with the optimum solutions of the original MSSP models. All three planning problems contain endogenous uncertainties, where the scenario tree structure is dependent on decisions. The primal bounds generated by the AEEV for all three problems are within 1% of the optimum. The framework reduces solution times up to four orders of magnitude compared to solving the original MSSP models for large problems.

The results suggest that the proposed framework has significant potential to be a standard to evaluate the benefit of solving a stochastic model and as an initial primal bound for solving large-scale multistage stochastic programs under endogenous uncertainties. We recommend regarding the AEEV and its objective function value as an essential primal bound/cut for solving all large-scale MSSP models with complete recourse under endogenous uncertainties, not only because AEEV is generated by a straightforward decision-making process but also due to its quality and small computational time.

CHAPTER 6

A Modified Lagrangian Relaxation Algorithm (mLR) and A Relaxed Knapsack-Problem Based Decomposition (RKDA)

This chapter presents two dual bounding approaches to efficiently generate tight dual bounds for large-scale MSSPs. The first approach is the modified Lagrangian relaxation algorithm (mLR), which relaxes the conditional NACs, and by exploiting the structure of the MSSP, a tight dual problem is formulated, which reduces the total number of Lagrange multipliers. The second approach employs a relaxed knapsack-problem based decomposition algorithm (RKDA) to solve a relaxed version of the MSSP to optimality. The relaxed MSSP is obtained by removing its resource constraints. As the NACs contribute to the main computational complexities in MSSP, like most heuristics, mLR regards and removes the NACs as complicated constraints. The RKDA relaxes the resource constraints instead and keeps the NACs in its framework.

6.1 A modified Lagrangian relaxation approach (mLR)

This section presents a new Lagrangian relaxation formulation for bounding MSSPs under endogenous uncertainties. Similar to the existing Lagrangian relaxation schemes (Gupta and Grossmann, 2014), the new Lagrangian relaxation regards NACs as complicating constraints, however, utilizes the special structure of the NACs to formulate a tighter dual problem, which also eliminates half of the Lagrange multipliers. To obtain tight dual bounds using Lagrangian relaxation, the multipliers should be updated appropriately. The most widely used technique to update the multipliers is the sub-gradient optimization method (Ardalan et al., 2016). An algorithm to be used in the sub-gradient optimization method to update multipliers is introduced. The algorithm dynamically updates the multipliers based on the ratio of the value of the dualized constraints in the objective function to the original objective function value. We applied the new

Lagrangian relaxation to bound instances of ALIP problem (Zeng and Cremaschi, 2017a) under uncertain production rate and the CTP problem (Colvin and Maravelias, 2008) under uncertain clinical trial outcomes.

6.1.1 Illustrative MSSP formulation

A simple deterministic equivalent formulation of an MSSP with complete recourse under endogenous uncertainties is derived from Zeng et al. (2018). The uncertain source $i \in \boldsymbol{I}$ associate with its uncertain parameter (θ_i) has finite outcomes represented by finite set $\{\theta_i^1, \theta_i^2, \dots, \theta_i^{R_i}\}$. The scenario set ($\boldsymbol{S}$) is constructed using the Cartesian product: $\boldsymbol{S} := \{\times_{i \in \boldsymbol{I}} \theta_i\}$. The planning horizon is discretized, $t \in \boldsymbol{T} = \{1, 2, 3, .., T\}$. The uncertain parameters are endogenous, and we define an endogenous scenario pair set $(s, s') \in S_E$ for scenarios s and s' that are indistinguishable in terms of the realization of the endogenous uncertain parameter i: $S_E := \{(i, s, s') : i \in I, \theta_i^s = \theta_i^{s'}\}$. The MSSP is defined by Eqns. (6.1) - (6.6):

$$OV = max \sum_s p_s \sum_i \sum_t G_{i,t,s}\left(\theta_i^s, b_{i,t}^s, \gamma_t^s\right) \tag{6.1}$$

$$h\left(b_{i,t}^s, \gamma_t^s, \theta_i^s\right) = 0 \quad \forall i \in \boldsymbol{I}, t \in \boldsymbol{T}, s \in \boldsymbol{S} \tag{6.2}$$

$$g(b_{i,t}^s, \gamma_t^s, \theta_i^s) \leq 0 \quad \forall i \in \boldsymbol{I}, t \in \boldsymbol{T}, s \in \boldsymbol{S} \tag{6.3}$$

$$b_{i,1}^s = b_{i,1}^{s'} \qquad \forall i \in \boldsymbol{I}, \forall (s, s') \in \boldsymbol{S} \tag{6.4}$$

$$\begin{bmatrix} Z_t^{s,s'} \\ b_{i,t}^s = b_{i,t}^{s'} \end{bmatrix} \vee \left[\neg Z_t^{s,s'}\right] \qquad \forall (i, s, s') \in \boldsymbol{S_E}, \forall t \in \boldsymbol{T}, t > 1 \tag{6.5}$$

$$b_{i,1}^s, Z_t^{s,s'} \in \{0,1\}, \gamma_t^s \in \mathbb{R} \qquad \forall (i, s, s') \in \boldsymbol{S_E}, \forall t \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \tag{6.6}$$

Binary variables $b_{i,t}^{s}$ represent state decisions made at the beginning of each time period, and variables γ_t^s represent recourse-action decisions that follow the state decisions. The state decisions $b_{i,t}^{s}$ are enforced to be identical until the differentiating event occurs either by initial or conditional NACs. The recourse variables γ_t^s are determined by scenario specific constraints. The optimum objective function value is defined by variable OV.

The objective function (Eqn. (6.1)) maximizes the expected value of the total profit/gain associated with state variables $b_{i,t}^{s}$ and recourse-action decision variables γ_t^s, where p_s is the probability of scenario s and the function $G_{i,t,s}\left(V_{i,t}, \theta_i^s, b_{i,t}^{s}, y_{i,t}^{s}, \gamma_t^s\right)$ calculates the total profit/gain for the specific scenario s. Equations (6.2) and (6.3) are scenario specific inequality and equality constraints that may or may not include the uncertain parameters (θ_i^s). Functions $G_{i,t,s}(\cdot)$, $h(\cdot)$, and $g(\cdot)$ in Eqns. (1), (2), and (3) can either be linear or nonlinear. Equation (6.4) is initial NACs, and Eqn. (6.5) is conditional NACs that are enforced depending on the values of the Boolean variable, $Z_t^{s,s'}$. This variable is equal to one if scenarios s and s' are indistinguishable at time period t, and zero otherwise.

6.1.2 Original Lagrangian relaxation problem (LR)

Lagrangian relaxation removes the complicating constraints from the original problem and adds them to the objective function with penalty terms, which are called Lagrangian multipliers. The Lagrangian multipliers represent penalties for solutions not satisfying the particular constraints (Conejo et al., 2006). In Lagrangian relaxation of a MSSP, the initial and conditional NACs are regarded as complicated constraints. The conditional NACs enforcing $b_{i,t}^{s} = b_{i,t}^{s'}$ in Eqn. (6.5) can be reformulated as two inequalities, Eqns. (6.7) and (6.8):

$$b_{i,t}^{s} - b_{i,t}^{s'} \leq 1 - Z_t^{s,s'} \quad \forall i \in \boldsymbol{I}, t \in \boldsymbol{T}, s, s' \in \boldsymbol{S} \tag{6.7}$$

$$b_{i,t}^{s} - b_{i,t}^{s'} \geq Z_t^{s,s'} - 1 \quad \forall i \in \boldsymbol{I}, t \in \boldsymbol{T}, s, s' \in \boldsymbol{S} \tag{6.8}$$

To construct Lagrangian relaxation of a MSSP, both initial (Eqn. 4) and reformulated conditional NACs (Eqns. (6.7) and (6.8)) are dualized in the objective function. The original Lagrangian relaxation of MSSP (LR) is given in Eqns. (6.9) – (6.12):

$$\begin{aligned} OV = max \sum_s p_s \sum_i \sum_t G_{i,t,s}\left(\theta_i^s, b_{i,t}^s, \gamma_t^s\right) \\ + \sum_{i,t,s,s'} \Big(\lambda 0_{i,s,s'} \left(b_{i,1}^s - b_{i,1}^{s'}\right) \\ + \lambda 1_{i,t,s,s'} \left(1 - Z_t^{s,s'} - b_{i,t}^s + b_{i,t}^{s'}\right) \\ + \lambda 2_{i,t,s,s'} \left(1 - Z_t^{s,s'} + b_{i,t}^s - b_{i,t}^{s'}\right) \Big) \end{aligned} \tag{6.9}$$

$$h\left(b_{i,t}^s, \gamma_t^s, \theta_i^s\right) = 0 \quad \forall i \in \boldsymbol{I}, t \in \boldsymbol{T}, s \in \boldsymbol{S} \tag{6.10}$$

$$g(b_{i,t}^s, \gamma_t^s, \theta_i^s) \leq 0 \quad \forall i \in \boldsymbol{I}, t \in \boldsymbol{T}, s \in \boldsymbol{S} \tag{6.11}$$

$$\begin{aligned} b_{i,1}^s, Z_t^{s,s'} \in \{0,1\}, \gamma_t^s \in \mathbb{R}, \lambda 0_{i,s,s'} \in \mathbb{R}, \lambda 1_{i,t,s,s'}, \lambda 2_{i,t,s,s'} \geq 0 \mathbb{R} \quad \forall (i,s,s') \\ \in \boldsymbol{S_E}, \forall t \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \end{aligned} \tag{6.12}$$

The multipliers $\lambda 0_{i,s,s'}$ for initial NACs can either be positive or negative. The multipliers of conditional NACs $\lambda 1_{i,t,s,s'}$ and $\lambda 2_{i,t,s,s'}$ can only take nonnegative values, which enforce the dualized constraints to be great than or equal to zero. The solution of the relaxed problem LR yields a valid dual bound for the original MSSP.

6.1.3 A modified Lagrangian relaxation formulation (mLR)

In the original relaxation, Lagrangian multipliers penalize the violation of the dualized constraints (Eq. (4.9)). All possible values of the conditional NACs (Eqns. (4.7) and (4.8) in Eq. (4.9)) are given in Table 6.1.

Table 6.1: All possible values of dualized constraints (Eqns. (6.7) and (6.8) in Eq. (6.9))

$Z_t^{s,s'}$	$b_{i,t}^{s}$	$b_{i,t}^{s'}$	$1 - Z_t^{s,s'} - b_{i,t}^{s} + b_{i,t}^{s'}$	$1 - Z_t^{s,s'} + b_{i,t}^{s} - b_{i,t}^{s'}$
0	0	0	1	1
0	0	1	2	0
0	1	0	0	2
0	1	1	1	1
1	0	0	0	0
1	0	1	1	-1
1	1	0	-1	1
1	1	1	0	0

As can be seen from Table 6.1, the conditional NACs are always satisfied when indicator variables ($Z_t^{s,s'}$) are equal to zero and can only be violated for indicator variable values equal to one. To tighten the dual bound (Eqn. (9)), we propose a modified relaxation that is shown in Eqns. (6.13) and (6.14), where the constraints ($b_{i,t}^{s} - b_{i,t}^{s'}$) are dualized only when indicator variables ($Z_t^{s,s'}$) values are equal to one.

$$OV = max \sum_{s} p_s \sum_{i} \sum_{t} G_{i,t,s}\left(\theta_i^s, b_{i,t}^s, \gamma_t^s\right) + \sum_{i,t,s,s'} \left(\lambda 0_{i,s,s'}\left(b_{i,1}^{s} - b_{i,1}^{s'}\right) + u_{i,t,s,s'}\right) \tag{6.13}$$

$$\begin{bmatrix} Z_t^{s,s'} \\ u_{i,t,s,s'} = \lambda_{i,t,s,s'}(b_{i,t}^{s} - b_{i,t}^{s'}) \end{bmatrix} \vee \begin{bmatrix} \neg Z_t^{s,s'} \\ u_{i,t,s,s'} = 0 \end{bmatrix} \quad \forall (i,s,s') \in \boldsymbol{S_E}, \forall t \in \boldsymbol{T}, t > 1 \tag{6.14}$$

The disjunction in Eq. (6.14) ensures that a conditional NAC is only dualized and added to the objective function (Eq. (6.13)) with a Lagrangian multiplier $\lambda_{i,t,s,s'} \in \mathbb{R}$ if its associated indicator variable ($Z_t^{s,s'}$) is equal to one. We utilize the disjunctive structure to indicate the selection of dualized constraints in the optimization model. If the indicating variable is equal to zero, the dualized constraint variable $u_{i,t,s,s'}$ is set to zero. In this modified relaxation, the dualized constraint is equivalent to an equality constraint for indistinguishable scenarios (i.e., for $Z_t^{s,s'} = 1$) with assigned Lagrangian multipliers. The modified Lagrangian relaxation formulation has half of the Lagrangian multipliers compared to the original one ($\lambda_{i,t,s,s'}$ vs. $\lambda 1_{i,t,s,s'}, \lambda 2_{i,t,s,s'}$). Eqns. (6.13) and (6.14) can be reformulated as Eq. (6.15).

$$OV = max \sum_s p_s \sum_i \sum_t G_{i,t,s}\left(\theta_i^s, b_{i,t}^s, \gamma_t^s\right) + \sum_{i,t,s,s'} \left(\lambda 0_{i,s,s'}\left(b_{i,1}^s - b_{i,1}^{s'}\right) + \lambda_{i,t,s,s'} Z_t^{s,s'}\left(b_{i,t}^s - b_{i,t}^{s'}\right)\right) \tag{6.15}$$

The modified Lagrangian relaxation formulation (mLR) is given by Eqns. (6.15), (6.10), (6.11), (6.12). Both LR and mLR have the same KKT conditions, and they are both dual bounds for the original MSSP.

6.1.4 Two Lagrangian multiplier updating schemes for sub-gradient optimization method

To obtain tight dual bounds using Lagrangian relaxation, the multipliers should be updated appropriately. The most widely used technique to update the multipliers is the sub-gradient optimization method (Gupta and Grossmann, 2014). Given an initial $\lambda 0_{i,s,s'}$, $\lambda_{i,t,s,s'}$($\forall (i,s,s') \in \boldsymbol{S_E}, \forall t \in \boldsymbol{T}$) (multipliers for a total of $|\boldsymbol{S_E}||T|$ dualized constraints), the multipliers at the k^{th} iteration, $\lambda 0_{i,s,s'}^k, \lambda_{i,t,s,s'}^k$, are calculated using Equation (6.16)-(6.17) (Fisher, 1981),

$$\lambda 0_{i,s,s'}^{k+1} = max\left\{0,\ \lambda 0_{i,s,s'}^{k} - \gamma^k (b_{i,1}^{s} - b_{i,1}^{s'})\right\} \quad \forall (i,s,s') \in \boldsymbol{S_E} \tag{6.16}$$

$$\lambda_{i,t,s,s'}^{k+1} = max\left\{0,\ \lambda_{i,t,s,s'}^{k+1} - \gamma^k (b_{i,t}^{s} - b_{i,t}^{s'})\right\} \quad \forall (i,s,s') \in \boldsymbol{S_E}, \forall t \in \boldsymbol{T}, t > 1 \tag{6.17}$$

where $b_{i,1}^{s}, b_{i,1}^{s'}, b_{i,t}^{s}, b_{i,t}^{s'}$ is the solution of Lagrangian relaxation at iteration k. $(b_{i,1}^{s} - b_{i,1}^{s'})$ and $(b_{i,t}^{s} - b_{i,t}^{s'})$ is the dualized initial and conditional NACs in the objective function. The variable γ^k is the positive step size at iteration k, which is calculated by Equation (17):

$$\gamma^k = scale \frac{Z(\lambda_n^k) - Z^*}{violated\ constraints} \tag{6.18}$$

where $Z(\lambda_n^k)$ is the dual bound from Lagrangian relaxation at the k^{th} iteration and Z^* is the best known primal bound generated from GKDA and AEEV. In Equation (6,17), the parameter $scale$ is a scalar satisfying $0 \leq scale \leq 2$. Often the value of $scale$ is set to 2 at first iteration. If the dual bound fails to converge after a number of iterations, the value of $scale$ is decreased.

Here, we introduce an algorithm for dynamically updating the values of γ^k and $scale$ in the sub-gradient optimization method. This algorithm is summarized in Figure 6.2.1. The iteration counter (k) and multipliers are initialized as zero, and the scale is set to two at initialization. The first dual bound is obtained by solving the Lagrangian relaxation. The algorithm then determines the number of dualized constraints that were violated. The primal bound in Equation (6.17) is obtained by the generalized knapsack decomposition algorithm GKDA, $GKDA^*$ in Figure 6.2.1 (Zeng et al., 2018). The difference between dual and primal bounds are divided by the total number of violated dualized constraints. The step size variable γ^k is updated using the difference of dual and primal bounds and multiplying this difference with $scale$ and dividing it by the number of violated dualized constraints. Starting with the second iteration, the algorithm dynamically updates the value of $scale$ based on the ratio of dualized constraints to the original objective function to reduce the zigzagging behavior of the Lagrangian dual solutions as iterations progress. The

stopping criteria are either reaching a predetermined maximum number of iterations or satisfying all dualized constraints.

$k, \lambda_n^k\ (n \epsilon N) \coloneqq 0, scale^k \coloneqq 2$
FOR $k = 1\ to\ K_{max}$
    $Z(\lambda_n^k) \coloneqq$ dual bound from Lagrangian relaxation with multipliers $\lambda_n^k\ (n \epsilon N)$
    $x^k \coloneqq$ solution of Lagrangian relaxation
    $N^v \coloneqq$ number of violated constraints of dualized constraints
    $GKDA^* \coloneqq$ primal bound from generalized knapsack decomposition algorithm
    IF $k \geq 2$ THEN
        IF $\frac{dualized\ constaints\ value}{Original\ objective\ function\ value} \geq 10\%$ THEN
            $scale^k = \frac{scale^{k-1}}{1.618}$
    $\gamma^k \coloneqq scale^k \frac{Z(\lambda_n^k) - GKDA^*}{N^v}$
    FOR $n = 1\ to\ N$
        Update λ_n^{k+1} with dualized constraints and γ^k
    IF no violated constraints THEN BREAK

Figure 6.2.1: Modified sub-gradient optimization method for updating Lagrangian multipliers

6.1.5 Case Study

We applied the original Lagrangian relaxation (RL) and modified Lagrangian relaxation (mLR) formulations to generate dual bounds for ALIP problem with planning horizons of 12 months and CTP problem (3 product 3 clinical trials with a planning horizon of 12 periods). The complete parameters of the instances can be found in Zeng and Cremaschi (2017b). The models and algorithms were implemented in Pyomo and solved using CPLEX 12.6.3 on a standard node of Auburn University Hopper Cluster. The maximum number of iterations is set to 100 for the sub-gradient optimization method. The ALIP and clinical trial planning problem solution results are presented in Table 6.2, respectively.

Table 6.2: The ENPV, relative gap, and terminated iterations of LR with a standard sub-gradient method (LR standard), LR with a modified sub-gradient method (LR modified), and mLR with a modified sub-gradient method (mLR modified) for artificial lift infrastructure planning problem and clinical trial planning problem

	LR (standard)			LR (modified)			mLR (modified)		
Instance	ENPV	Gap	Iterations	ENPV	Gap	Iterations	ENPV	Gap	Iterations
6-month (ALIP)	1597031	0.0%	7	1597031	0.0%	7	1597031	0.0%	7
12-month (ALIP)	3219895	15.7%	100	2915898	4.7%	100	2783908	0.0%	79
2-prod (CTP)	1107	0.3%	100	1105	0.1%	100	1104	0.0%	28
3-prod (CTP)	1276	10.4%	100	1262	9.2%	100	1233	6.7%	100

Table 6.2 reveals that the modified sub-gradient updating scheme improves bounds qualities for tested problems. The mLR obtains 0% percentage relative gaps for three tested instances. For the 3-prod case in clinical trial planning problem, both LR and mLR can not converge within 100 iterations, while LR and mLR obtain 10.38% and 6.66% relative gap, respectively.

6.1.6 Conclusion

This section developed a new Lagrangian relaxation reformulation and modified sub-gradient updating scheme for multistage stochastic programs under endogenous uncertainties. The modified Lagrangian relaxation (mLR) utilizes a disjunctive structure to indicate which conditional NACs should be dualized to objective function in the model. The mLR reduces the number of dualized constraints and Lagrangian multipliers and obtains tight dual bound. We compare the original (LR) and modified (mLR) Lagrangian relaxation with two planning problems. The mLR is able to obtain optimality gaps for three instances. The results indicate that

mLR generates tighter relative gaps in fewer iterations than the original Lagrangian relaxation formulation.

6.2 A Relaxed Knapsack-Problem Based Decomposition (RKDA)

As shown in the previous section, for cases with a significant number of scenarios, relaxing NACs, which are the linkage between scenario pairs in MSSP, will naturally result in large relative gaps. In this section, the Relaxed Knapsack-problem based Decomposition Algorithm (RKDA) builds upon has been developed for solving the MSSP model of resource-constrained multiple projects scheduling with stochastic task success.

6.2.1 Introduction

Resource-constrained project scheduling problem (RCPSP) is a well-known and studied problem with practical applications in many fields, including the chemical process industry, manufacturing, logistics, and project management (Hartmann and Briskorn 2010). The problem is defined by a project that should complete a set of non-preemptive activities with different durations and resource requirements. Some activities are linked by precedence relations, and there are limited renewable and/or non-renewable resources available for completing the activities. The traditional objective of RCPSP is to minimize makespan, however cost and robust schedule based objectives, e.g., minimization of total project costs or net present value, are some of the popular extensions (Hartmann and Briskorn 2010). Many approaches have been investigated to model and solve RCPSP, which is an NP-hard problem in the strong sense (Bartusch, Möhring et al. 1988). These approaches include discrete and continuous time mixed integer linear programming (e.g., (Christofides, Alvarez-Valdes et al. 1987, Zapata, Hodge et al. 2008, Naber and Kolisch 2014)), constraint programming (e.g.,(Liess and Michelon 2008)), satisfiability module theory (e.g.,

(Ansotegui et al., 2011)), and stochastic optimization approaches such as particle swarm optimization (e.g.,(Koulinas et al. 2014)) among others. An excellent review of RCPSP and its variants can be found in Hartmann and Briskorn (2010).

Resource-constrained multi-project scheduling problem (RCMPSP) is an extension of the conventional RCPSP and considers the interaction and utilization of shared resources among multiple projects (Zapata et al., 2008). Many applications from the process industry can be classified as RCMPSP such as pharmaceutical and agrochemicals new product development tests planning (Jain and Grossmann, 1999; Maravelias and Grossmann, 2004), batch scheduling (Schwindt and Trautmann, 2000), and new technology investment planning problem (Christian and Cremaschi, 2018). Therefore, the process systems engineering community made significant contributions to the modeling and solution approaches for RCPSP and RCMPSP. The classical RCPSP and RCMPSP is deterministic, where all parameter values of the problems are assumed to be known with certainty. However, for real applications, the parameters of these problems are subject to considerable uncertainty. For example, resource availabilities may change, task durations may fluctuate due to equipment failure, or unexpected weather events (shipment delays), project(s) may not be able to successfully complete one or more tasks and hence be dropped, etc. Stochastic task success for a project, which is only realized after the task is scheduled for execution, and hence, is decision dependent, is an endogenous uncertain parameter.

6.2.2 A general MSSP model for RCMPSP with stochastic task success

In this section, a deterministic equivalent formulation of an MSSP for the RCMPSP with stochastic task success is introduced (Eqns. (6.19) - (6.26)). The candidate projects are defined as $i \in \boldsymbol{I}$ and each project needs to complete tasks $j \in \boldsymbol{J}$. The uncertain task outcome(s) associated with candidate project i is defined by a finite set $\theta_i = \{\theta_i^1, \theta_i^2, \dots, \theta_i^{R_i}\}$. The scenario set and the

endogenous scenario pair set have the same definition in Chapter 4. The MSSP formulation can then be defined as:

$$OV = max \sum_{s} p_s \sum_{i} \sum_{j} \sum_{t} G_{i,j,t,s}\left(\theta_i^s, b_{i,j,t}^s\right) \tag{6.19}$$

$$h\left(b_{i,j,t}^s\right) = 0 \quad \forall i \in \boldsymbol{I}, j \in \boldsymbol{J}, t \in \boldsymbol{T}, s \in \boldsymbol{S} \tag{6.20}$$

$$g(b_{i,j,t}^s) \leq 0 \quad \forall i \in \boldsymbol{I}, j \in \boldsymbol{J}, t \in \boldsymbol{T}, s \in \boldsymbol{S} \tag{6.21}$$

$$\sum_{i} \sum_{j} \sum_{t' > t-\tau_{i,j}}^{t' \leq t} \rho_{i,j,r} b_{i,j,t'}^s \leq {\rho_r}^{max} \qquad \forall r \in \boldsymbol{R}, j \in \boldsymbol{J}, t \in \boldsymbol{T}, s \in \boldsymbol{S} \tag{6.22}$$

$$b_{i,j,1}^s = b_{i,j,1}^{s'} \qquad \forall i \in \boldsymbol{I}, j \in \boldsymbol{J}, \forall (s, s') \in \boldsymbol{S} \tag{6.23}$$

$$\begin{bmatrix} Z_t^{s,s'} \\ b_{i,j,t}^s = b_{i,j,t}^{s'} \end{bmatrix} \vee \left[\neg Z_t^{s,s'}\right] \qquad \forall (i, s, s') \in \boldsymbol{S_E}, j \in \boldsymbol{J}, \forall t \in \boldsymbol{T}, t > 1 \tag{6.24}$$

$$Z_t^{s,s'} \Leftrightarrow H\left(b_{i,j,1}^s, b_{i,j,2}^s, \dots, b_{i,j,t}^s\right) \quad \forall (i, s, s') \in \boldsymbol{S_E}, j \in \boldsymbol{J}, \forall t \in \boldsymbol{T}, t > 1 \tag{6.25}$$

$$b_{i,j,t}^s, Z_t^{s,s'} \in \{0,1\} \qquad \forall (i, s, s') \in \boldsymbol{S_E}, s \in \boldsymbol{S}, j \in \boldsymbol{J}, \forall t \in \boldsymbol{T}, \forall i \in \boldsymbol{I} \tag{6.26}$$

The MSSP model has binary state variables, $b_{i,j,t}^s$, which is equal to one if the task j of projects i is selected to be started at time t in scenario s. The problem has no recourse actions and selection decisions are made sequentially along the planning horizon. Task j of project i requires $\rho_{i,j,r}$ amount of resource type $r \in \boldsymbol{R}$ for execution. The resource constraint, Eq. (6.22), ensures that the total consumed resources for task execution does not exceed the limited budget available for resources type r, ${\rho_r}^{max}$, at any time period $t \in \boldsymbol{T}$. In the model, the uncertain parameters are only incorporated in the objective function (Eq. (6.19)), which maximizes the expected value of total profit/gain associated with state variable $b_{i,j,t}^s$, where p_s is the probability of scenario s and the

function $G_{i,j,t,s}\left(\theta_i^s, b_{i,j,t}^s\right)$ calculates the total profit/gain for specific scenario s. Equations (6.20) and (6.21) are sequencing inequality and equality constraints that do not include any uncertain parameters. Functions $G_{i,j,t,s}(\cdot)$, $h(\cdot)$, and $g(\cdot)$ can either be linear or nonlinear. Equations (6.23) and (6.24) are initial and conditional NACs. The conditional NACs are enforced depending on the values of the Boolean variable, $Z_t^{s,s'}$, which is equal to one if scenarios s and s' are indistinguishable at time period t, and zero otherwise. The values of $Z_t^{s,s'}$ are determined based on the decisions made up to the previous time period via the expression $H(.)$ in Eq. (6.25).

6.2.3 Scheme and formulation of Relaxed Knapsack-Problem Based Decomposition (RKDA)

The RKDA generates a valid dual bound for the MSSP of the RCMPSP with stochastic task success by solving a relaxed version of the original MSSP to optimality. The original MSSP is relaxed by removing the resource constraints. We refer to this relaxed MSSP formulation as RMSSP. The optimal solution of RMSSP provides a dual bound for the original MSSP.

KSP generation in GKDA has been introduced in Chapter 3. The weight constraints of the KSPs in the GKDA are equivalent to the resource constraints of the MSSP. Therefore, to solve the RMSSP, the weight constraints of the KSPs are removed yielding relaxed KSPs given in Eqns. (6.27) – (6.28).

$$max: \sum_{i \in E} V_i x_i \tag{6.27}$$

$$x_i \in \{0,1\} \qquad \forall i \in E \tag{6.28}$$

The original GKDA is modified to generate and solve the relaxed KSPs every time period at which the outcome of an uncertain parameter is realized. We refer to this modified GKDA as a

relaxed GKDA (RKDA). The solution obtained by RKDA to the RMSSP satisfies the sequencing constraints (Eqns. (6.20) – (6.21)) and the NACs (Eqns. 6.23 – 6.24) of the RMSSP , and recommends starting all eligible tasks of all revenue generating projects at the earliest times allowable. Therefore, the RKDA yields the optimal solution of the RMSSP.

6.2.4 Case Study

We applied GKDA and RKDA to generate initial primal and dual bounds for the clinical trial planning problem with two-, three-, four-, five-, six-, seven- and ten-products and three clinical trials with different planning horizons. We also solved the original MSSPs of these problems to 0.1% optimality gap when possible. The models and algorithms were implemented in Pyomo and solved using CPLEX 12.6.3 on a standard node of Auburn University Hopper Cluster. Problem data are available upon request.

6.2.4.1 Primal and dual bounds obtained by KDA and RKDA

Number of trials, time periods, and scenarios for the problems are summarized in Table 6.3 along with the optimum ENPVs and the MSSP solution times for the problems that were solved to optimality. As expected, the MSSP solution times grow exponentially with the number scenarios. For example, CPLEX 12.6.3 required more than 8 hours to solve the six-product problem. The MSSP enumerates all scenarios, and quickly becomes computationally intractable due to its space and time complexity. Therefore, the MSSP formulations of seven- and ten-product problems cannot be generated in RAM and hence cannot be solved using CPLEX 12.6.3 due to their space complexities.

Table 6.3: Main characteristics and the MSSP solutions of the problems

Case Name	Number of Trials	Number of Time Periods	Number of Scenarios	ENPV ($M)	MSSP Solution Time (CPUs)
2-prod	2	5	9	1104	0.1
3-prod	3	12	64	1189	4
4-prod	3	6	256	1696	7
5-prod	3	6	1,024	2082	185
6-prod	3	6	4,096	2450	29,357
7-prod	3	8	16,384	--	--
10-prod	3	10	1,048,576	--	--

As can be seen from Table 6.4, the results reveal that the initial relative gaps between KDA and RKDA solutions for two-, three-, four-, five-, six-, seven-, and ten-product instances are 2.10%, 3.65%, 2.62%, 3.65%, 4.53%, 4.66%, and 3.90%, respectively. Knapsack-problem based decomposition algorithm (KDA) and RKDA yields these gaps in 0.1, 2, 3, 7, 18, 69, and 2377 CPU seconds for two-, three-, four-, five-, six-, seven-, and ten-product instances, respectively. The total solution times for obtaining good feasible solutions and tight primal and dual bounds using KDA and RKDA are under one hour for all instances considered. The Pearson correlation coefficient between the total solution time and the number of scenarios is 1.00, revealing the strong linear relationship between them. Therefore, unlike the MSSP solution time which grows exponentially with the number of scenarios, the total solution time for KDA and RKDA grows linearly with the number of scenarios. Additionally, we can obtain feasible solutions using KDA and assess the quality of these feasible solutions using RKDA for seven and ten-product instances, which yield significantly large MSSP models with more than one million scenarios.

Table 6.4: Primal and dual bounds, relative gap, and solution times for the KDA and RKDA

Case Name	KDA Primal Bound ($M)	RKDA Dual Bound ($M)	Relative Gap	Solution Time (CPUs)
2-prod	1097	1120	2.10%	0.1
3-prod	1178	1221	3.65%	2
4-prod	1677	1721	2.62%	3
5-prod	2052	2127	3.65%	7
6-prod	2407	2516	4.53%	18
7-prod	2874	3008	4.66%	69
10-prod	4078	4237	3.90%	2377

Unlike the deterministic equivalent of MSSP formulations, the GKDA and RKDA do not suffer from space complexity, and hence, can be used to generate a feasible solution and to obtain its solution quality for problems that are currently computationally intractable. The future work will focus on the framework development of relaxed version of knapsack-problem based decomposition algorithm, which will be able to apply to different MSSPs.

6.2.4.2 Available resource variation case study for RKDA

The resource availability would have a significant impact on solution quality. Thus, we investigate our sensitivity analysis by changing the resource availability for studied cases and we choose base cases for 3, 4, 5, and 6-drug case. It should be noted that original cases already have a very limited resources. To study the impact of overall resource availability, we construct four problems with varying degrees of overall resource availabilities: (1) fully constrained; (2) 40 percent unconstrained; (3) 70 percent unconstrained; and (4) unconstrained. The base case problem is assumed to be fully constrained. The fully unconstrained resource case is equivalent to removing all resource constraints, which will allow completing clinical trials of all potential drugs simultaneously yielding a relaxed problem. The 40 percent-unconstrained case is generated by

increasing the available resource of the fully constrained case by 40 percent. In the 70 percent unconstrained case, this increase is 70%. The results are summarized in Table 6.5.

Table 6.5: Resource constraint perturbation results for RKDA

Cases	MSSP	RKDA	Relative Gap
3-Drug	1189	1221	2.69%
3-Drug-R40%	1211	1221	0.83%
3-Drug-R70%	1220	1221	0.08%
3-Drug-R100%	1221	1221	0.00%
4-Drug	1697	1721	1.41%
4-Drug-R40%	1719	1721	0.12%
4-Drug-R70%	1721	1721	0.00%
4-Drug-R100%	1721	1721	0.00%
5-Drug	2083	2128	2.16%
5-Drug-R40%	2123	2128	0.24%
5-Drug-R70%	2128	2128	0.00%
5-Drug-R100%	2128	2128	0.00%
6-Drug	2450	2517	2.73%
6-Drug-R40%	2510	2517	0.28%
6-Drug-R70%	2517	2517	0.00%
6-Drug-R100%	2517	2517	0.00%

As shown in Table 6.5, the solution of Case with R100% is equivalent to RMSSP and results further approve that RKDA obtain the true solution of RMSSP. With relaxing the resource limits, the relative gaps obtained from RKDA decrease to 0%. The full constrained case present the largest relative gap. The Table 6.5 indicates that relaxing the resource constraints for fully constrained cases may still yield tight dual bound.

6.3 Conclusion

This chapter contributes to a modified Lagrangian relaxation approach and RKDA to obtain dual bounds for resource-constrained multiple project scheduling problems (RCMPSP) with stochastic task success endogenous uncertainty. Unlike most decomposition approaches by

addressing the complicated constraints NACs, RKDA keeps NACs and only relax resource constraints. Section 6.2 presents GKDA and RKDA to quickly obtain a feasible solution, a primal, and a dual bound for the stochastic RCMPSP under endogenous uncertainty. The GKDA and RKDA sequentially generate and solve a series of knapsack problems to find a feasible solution for the MSSP model of the RCMPSP with stochastic task success and to find the optimum solution of a relaxed version of the MSSP model of the RCMPSP with stochastic task success. Instead of enumerating all the scenarios like the MSSP model, the GKDA and RKDA generate knapsack problems based on the selected projects' outcomes and their associated tasks. Therefore, unlike the deterministic equivalent of MSSP formulations, the GKDA and RKDA do not suffer from space complexity. Hence, it can be used to generate a feasible solution and obtain its solution quality for stochastic RCMPSP under endogenous uncertainty that is currently computationally intractable.

This chapter also compared the strength of the dual bounds obtained by RKDA to dual bounds obtained by modified Lagrangian relaxation. The strategies are obtained by determining which NACs to dualize to the objective function. We also propose a new multiplier updating algorithm for stabilizing and strengthening the dual bounds obtained by the Lagrangian relaxation strategies. The comparisons revealed that the initial gaps obtained by GKDA and RKDA were tighter than the ones obtained by the Lagrangian relaxation strategies considered even after 100 iterations.

The large relative gaps with Lagrangian relaxation solutions are due to relaxed NACs, and results indicate that the relative gaps will be larger, with an increasing number of scenarios. The comparison of RKDA and Lagrangian relaxation reveals that relaxing resource constraints other than NACs would like to generate much tighter dual bounds, especially for a large number of scenarios. Furthermore, GKDA and RKDA do not suffer from space complexity, and hence, the

proposed approach is an initial step towards solving large-scale MSSP for real-world sized problems.

CHAPTER 7

A FRAMEWORK FOR SOLVING MULTISTAGE STOCHASTIC PROGRAMS WITH ENDOGENOUS UNCERTAINTY

We have developed GKDA and AEEV as primal bounds and RKDA and modified Lagrangian relaxation to generate dual bounds in previous chapters. This chapter explores the decomposition framework combining primal and dual bounding approaches for solving multistage stochastic programs with endogenous uncertain parameters. The proposed framework decomposes the problem to individual scenario subproblems and is applied to solve several instances of the pharmaceutical R&D pipeline management problem.

7.1 Lagrangian generalized knapsack-problem based decomposition framework (LGKDA)

In this section, we propose a new decomposition framework for solving multistage stochastic programs with endogenous uncertainties. The framework builds upon our generalized knapsack-problem based decomposition approach (GKDA) (chapter 4), which decomposes the original MSSPs into a series of knapsack subproblems (KSP) and solves these problems at appropriate decision points of the planning horizon.

We proposed four different strategies for applying Lagrangian decomposition to the original MSSP model to generate dual bounds as outlined below. All strategies remove initial and conditional NACs of the original MSSP model and differ in terms of which NAC groups are dualized. The strategies are:

(1) LD1: only dualize the **initial** NACs to the objective function,

(2) LD2: only dualize the **conditional** NACs to the objective function,

(3) LD3: dualize all the **initial, and** the **conditional** NACs to the objective function, and

(4) LD4: gradually dualize **violated** NACs to the objective function.

In the LD4 strategy, at the first iteration, we assume that no NACs are violated. The objective function of LD4 is equivalent to LD1, where only initial NACs are dualized. In solution from Lagrangian dual problem, the violated conditional NACs will be checked and dualized to the objective function in the next iteration. The violated conditional NACs will be sequentially checked and dualized to the objective function, and LD4 may be equivalent to LD3 after certain iterations.

The framework decomposes the Lagrangian problem to individual scenario problems, which are solved efficiently. The solution of the Lagrangian problem yields a dual bound. The decomposition framework-Lagrangian GKDA (LGKDA) employs GKDA with Lagrangian decomposition. The framework is illustrated in Figure 7.2.1.

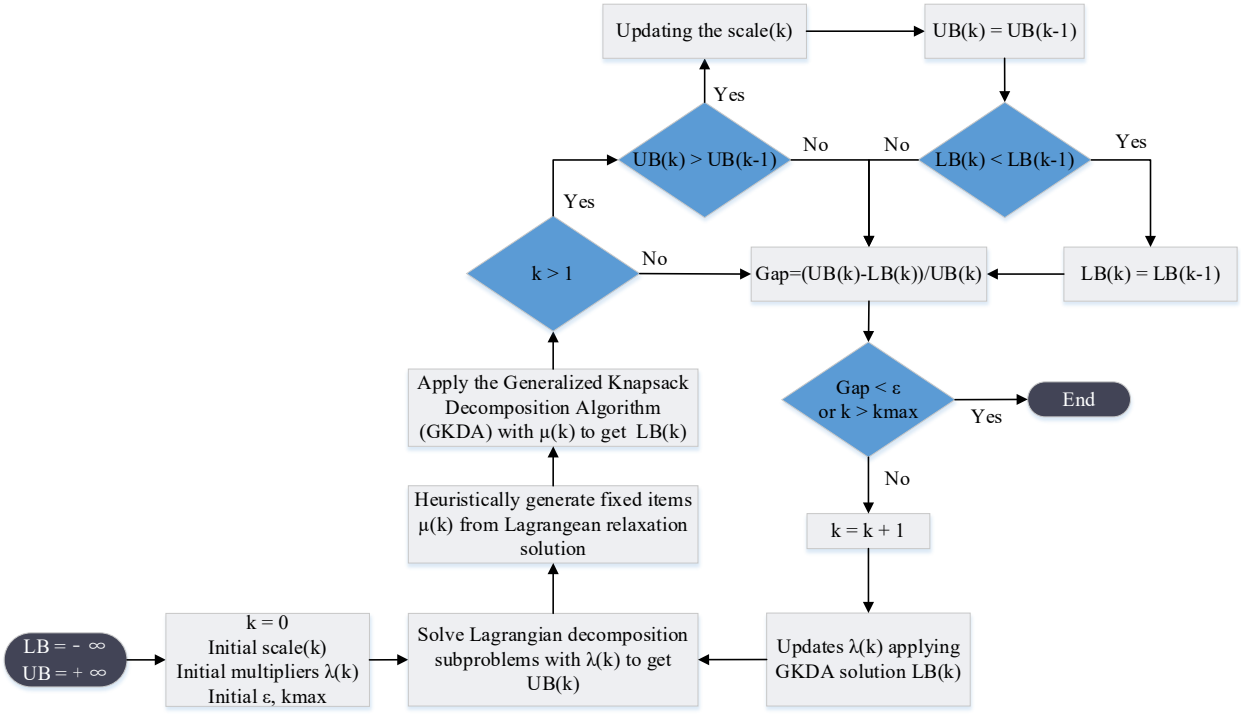

Figure 7.2.1: Lagrangian decomposition and Generalized Knapsack-Problem based Decomposition Framework

The algorithm starts by initializing the first set of Lagrange multipliers λ, scale value, maximum iterations, and threshold values. The Lagrangian decomposition problems are solved with fixed multipliers and generate an upper bound (UB, assuming minimization). Next, the GKDA is directed and solved with fixed items μ generated using the solution of the Lagrangian problem. Then, the Lagrangian multipliers are updated utilizing the tight primal bounds obtained from GKDA. The algorithm terminates when the optimality gap is below a pre-specified value or when the maximum number of iterations is reached.

We also developed a dynamic multiplier updating algorithm to be used in the sub-gradient optimization method to update the multipliers at every iteration in the dual problem. The algorithm determines the number of dualized constraints that were violated. The difference between dual and primal bounds are divided by the total number of nonzero dualized constraints. The step size variable γ^k is updated using the difference of dual and primal bounds and multiplying this difference with $scale$ and dividing it by the number of violated dualized constraints. To reduce the zigzagging behavior of the Lagrangian dual solutions as iterations progress, the algorithm dynamically updates the value of $scale$ based on the relative gap of Lagrangian dual solutions between two iterations.

7.2 Case study

We applied the LGKDA to solve the instances of pharmaceutical clinical trial planning problem with two-, three-, four- and five-products, and three clinical trials with different planning horizon lengths. The MSSP formulation of the clinical trial planning problem was introduced in previous chapters. The number of trials, time periods, scenarios, and optimum ENPV for the instances of the clinical trial planning problem are summarized in Table 7.1.

Table 7.1: Number of trials, time periods, scenarios, and ENPV of the clinical trial planning problem instance

Instance	Number of Trials	Number of Time Periods	Number of Scenarios	ENPV ($M)
2-prod	2	5	9	1104
3-prod	3	12	64	1189
4-prod	3	6	256	1696
5-prod	3	6	1,024	2082

The algorithms were implemented in Pyomo (Hart et al., 2012) and solved using CPLEX 12.6.3 on a standard node of Auburn University Hopper Cluster (each node has 128 GB of memory and 2 "Broadwell" CPUs (14 cores/CPU)).

7.2.1 Comparison of conventional multipliers updating scheme and proposed dynamic updating scheme

In this section, we take a 3-Drug case as an example to compare the conventional and our proposed dynamic updating scheme. The maximum iteration is set to one hundred. For the traditional updating scheme, the scale is set to constant 2. The dual bound obtained using a dynamic updating scheme, and the conventional updating scheme is presented in Figure 7.2.2. As shown in Figure 7.2.2, the dynamic updating scheme reduces the zigzagging behavior of the conventional way. In addition, Lagrangian decomposition obtained a tighter dual bound with the proposed dynamic updating scheme.

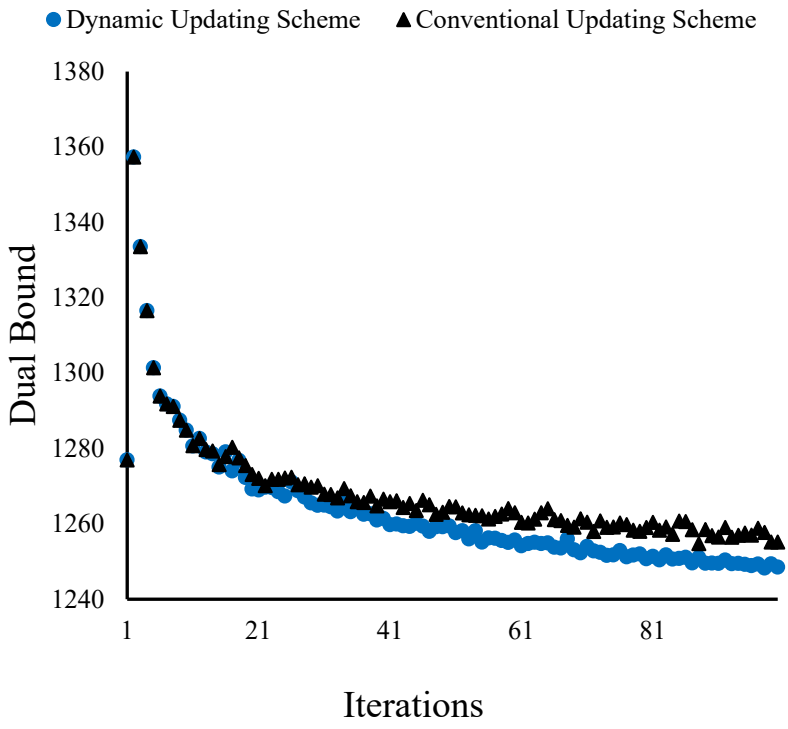

Figure 7.2.2: Dual bound obtained by Dynamic updating scheme and conventional updating scheme

7.2.2 Solutions by LGKDA with four strategies

In this section, we applied LGKDA with four proposed strategies named LD1, LD2, LD3, and LD4. The relative gaps obtained by primal and dual bounds within 100 iterations are shown in Figure 7.2.3.

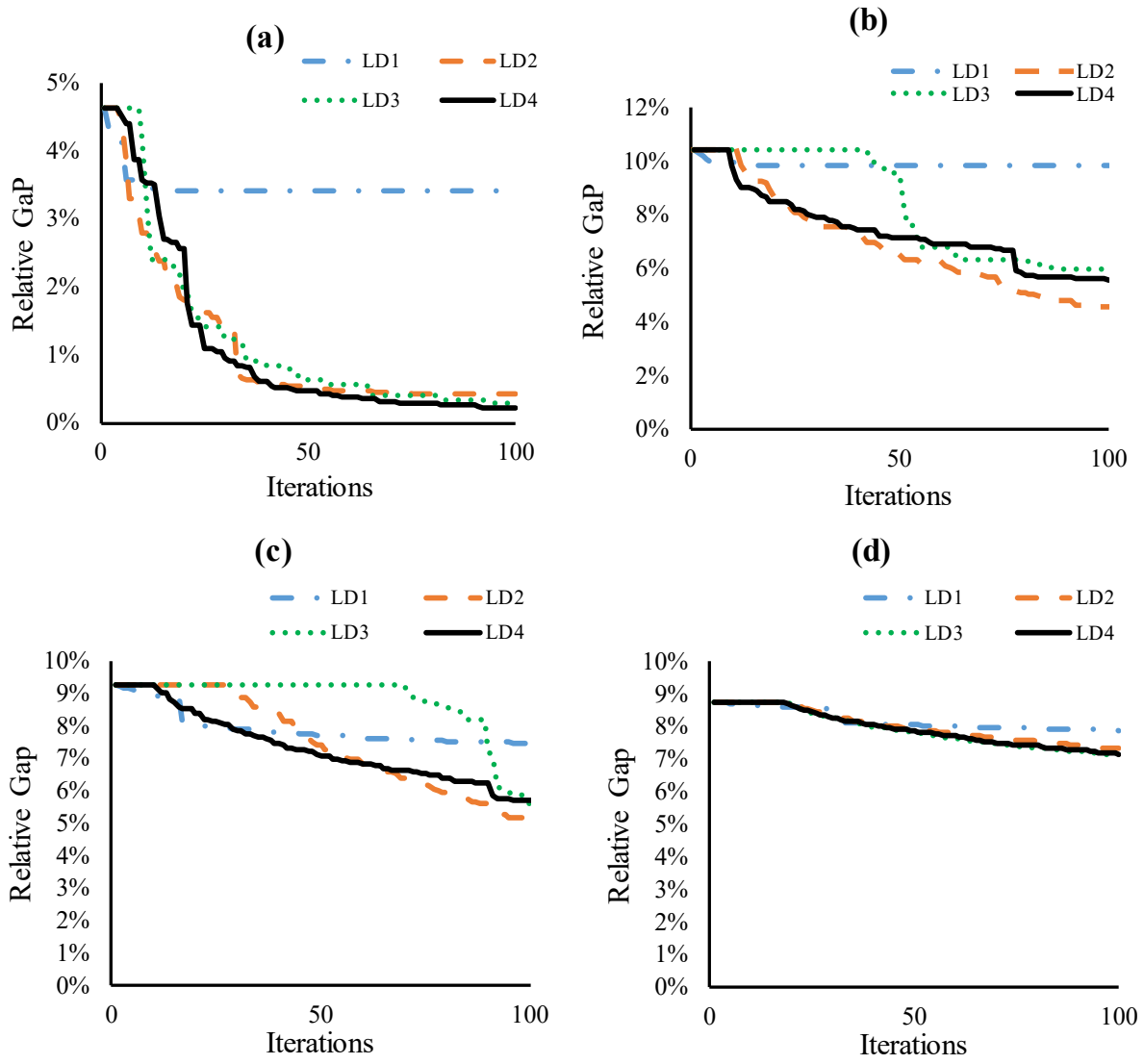

Figure 7.2.3: Evolution of relative percent gaps obtained by four decomposition strategies (LD1, LD2, LD3, and LD4) for (a) 2-product, (b) 3-product, (c) 4-product, and (d) 5-product instances of the clinical trial planning problem

As can be seen from Figure 7.2.3, LD2 has the best performance in general yielding tighter relative gaps when compared to LD1, LD3, and LD4, while LD1 consistently results in the largest gap after 100 iterations. In early iterations, LD3 obtains a larger gap than the one obtained by LD2. With gradually adding violated NACs, LD4 performs better than LD3, however it was observed that LD4 very quickly dualized all NACs in the first few iterations and became equivalent to LD3.

Figure 7.2.4 shows the 100 iterations between GKDA and Lagrangian decomposition for a 3-Drug case with the LD4 strategy. As shown in Figure 7.2.4, lower bounds from GKDA are improved by fixed items from solutions of Lagrangian decomposition. The optimum ENPV of the 3-Drug case is 1189. The ENPV of lower bounds is improved from 1156 to 1182, which reaches a 0.59% optimality gap. However, upper bounds generated from Lagrangian decomposition has large initial value and converges slowly with iterations. The final relative gap is 5.21% at 100 iterations.

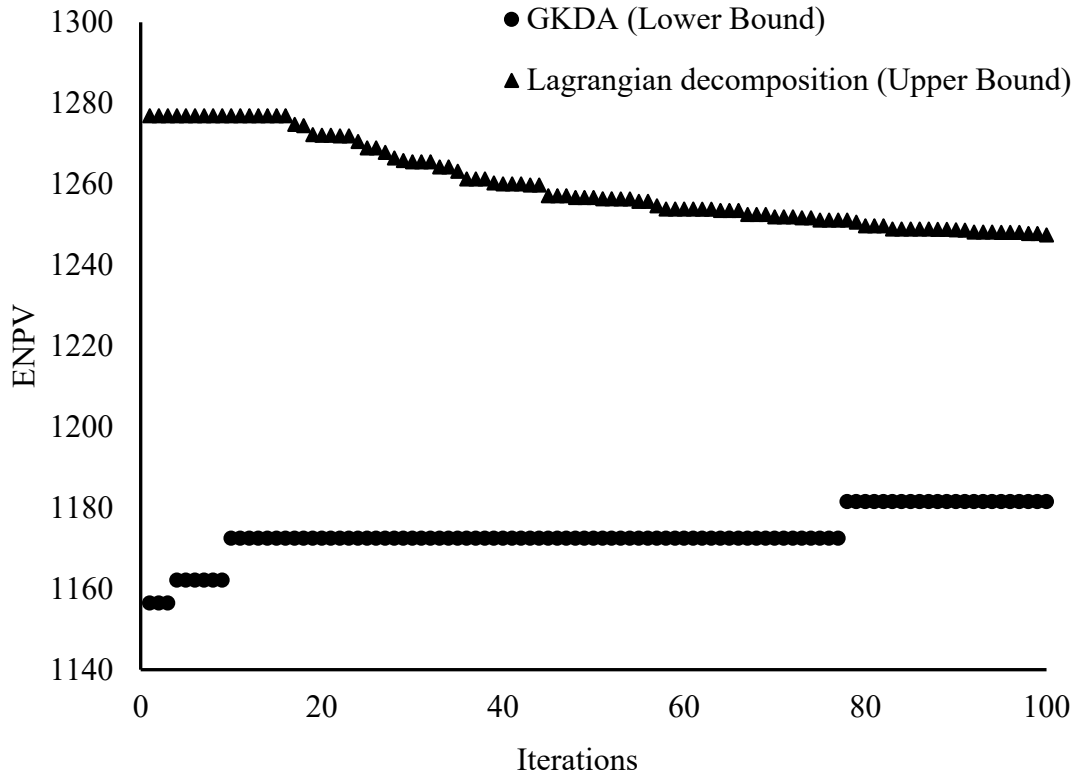

Figure 7.2.4: GKDA and Lagrangian decomposition (LD4) for Drug-3 within 100 iterations

7.3 Conclusion and discussion

The proposed frameworks utilize solutions of GKDA to generate Lagrangian multipliers and use Lagrangian decomposition solutions to direct and improve the primal bounds obtained from

GKDA. The proposed framework was able to reach reasonable relative gaps for small cases within 100 iterations. The relative gaps are larger and cannot be converged for larger instances with increasing scenarios. For example, the best relative percent gap is 0.021% for the 2-product instance (9 scenarios), whereas it is 7.15% for the five-product one (1024 scenarios) at 100 iterations.

The large relative gaps obtained by the proposed decomposition framework are main contributed by dual bounds. The Lagrangian problem is developed by relaxing all NACs and can be decomposed into individual scenario problems. However, with an increasing number of scenarios, solutions from all scenario subproblems are converged slowly and have large relative gaps. Besides, a large number of scenarios increase the computational effort to update the Lagrangian multipliers for individual scenario problems.

In the previous chapter, we have developed a new Lagrangian relaxation reformulation for multistage stochastic programs under endogenous uncertainties. The modified Lagrangian relaxation (mLR) reduces the total number of dualized constraints and multipliers and obtains tight dual bound by utilizing a disjunctive structure to indicate which conditional NACs should be dualized to the objective function. We recommend developing the decomposable formulation of modified Lagrangian relaxation (mLR) to parallelize and use frameworks proposed in this chapter to direct the primal bounds.

Chapter 6 has also shown relaxing resource constraints other than NACs would like to generate much tighter dual bounds, especially for a large number of scenarios. The RKDA proposed in chapter 6, implicitly incorporate NACs and relax the resource constraints and obtain dual bounds for resource-constrained multiple project scheduling problem (RCMPSP) with stochastic task success. Since RKDA generates tighter initial dual bounds than approaches relaxing the NACs and

does not suffer the computational effort to update Lagrangian multipliers, we recommend extending the framework of RKDA for general MSSP under endogenous uncertainties as to future work.

CHAPTER 8 SURROGATE-BASED OPTIMIZATION USING RANDOM FORESTS

The random forest algorithm has been used successfully as a general method for classification and regression since its introduction in 2001 (Breiman 2001). Random forests are decision tree-based surrogate models that can be used to map the relationship between input and output data when the actual one between the two is unknown or computationally expensive to evaluate for surrogate-based optimization. The random forest's decision tree structure allows for it to be formulated as a mixed-integer linear program (MILP), which can be readily solved using existing commercial solvers. Random forest models consist of a large number of individual uncorrelated decision trees as an ensemble, and generate their predictions using the expected value of individual decision tree predictions.

The corresponding MILPs are generally large and require significant computational effort to solve to optimality, especially for random forest models with a large number of decision trees. The MILP of random forest is similar to two-stage stochastic programs, which contains the complicating variables enforcing the same input value for all individual uncorrelated decision trees. Both two-stage stochastic programs and MILP of random forests enforce first-state decisions of individual scenarios or input values for trees with the equalities.

In this section, we propose frameworks with different dual and primal bounding approaches that are inspired by Lagrangian decomposition with cutting plane method (LD), Progressive Hedging (PH), and Alternating Direction Method of Multipliers (ADMM) as well as clustering approaches to solve the MILP of random forest models.

8.1 Random Forest Structure and MILP Formulation

Each RF model is composed of several trees with a series of decision nodes. The decision node consists of a parent, or test, node, with a left child node and a right child node. If the indicated input value is less than or equal to a threshold value for the given test node, then the left child node is selected, otherwise the right child node is selected. Decisions are made at each branch in a tree until a final node, or leaf node is reached. An example of a simple decision tree structure is illustrated in Figure 8.1. The value that is output for a tree for given inputs is the value of the final leaf node reached, and the output value for the entire RF model is the average value of the outputs for every decision tree in the forest.

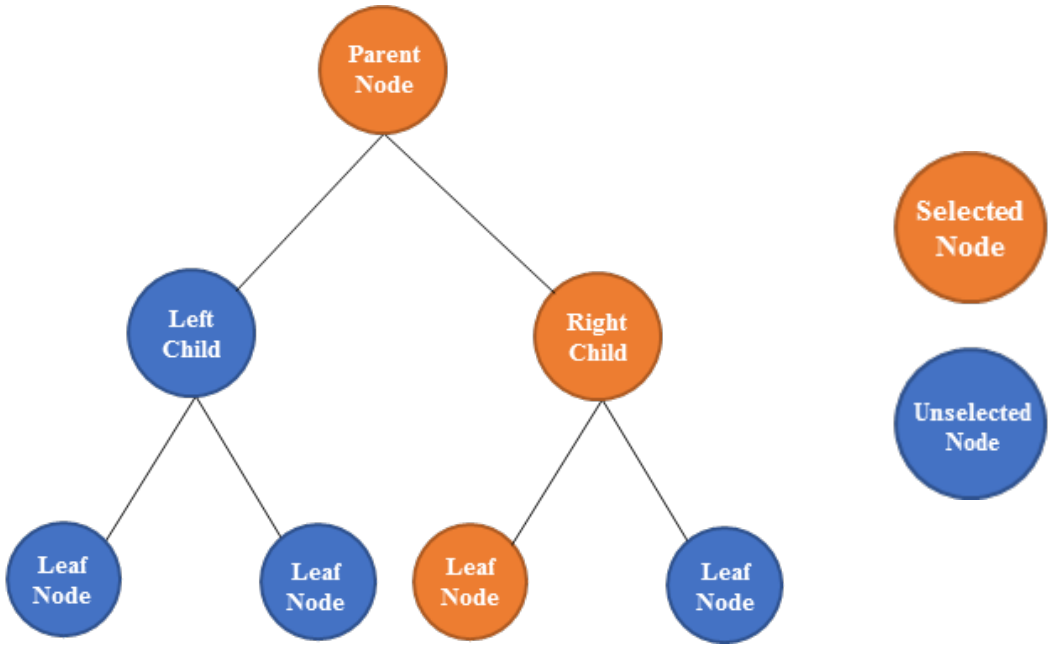

Figure 8.1: Random forest decision tree structure

The optimization problem for the RF objective function is formulated as follows:

$$\min_{x} \frac{1}{N_{trees}} \sum_{t}^{T} \sum_{l}^{L} c_{t,l} y_{t,l} \tag{8.1}$$

$$y_{t,1} = 1 \qquad \forall\, t \in T \tag{8.2}$$

$$y_{t,lc} + y_{t,rc} = y_{t,p} \qquad \forall\, (t, lc, rc, p) \tag{8.3}$$

$$\begin{bmatrix} y_{t,rc} \\ H_p \le x_i \end{bmatrix} \vee \begin{bmatrix} y_{t,lc} \\ x_i \le H_p \end{bmatrix} \qquad \forall\, (t, p, lc, i) \in Pl, (t, p, rc, i) \in Pr \tag{8.4}$$

The objective function, Eq. (8.1), of the formulated optimization problem is the output of the model, which is the expected value of the final leaf nodes selected by the model. In the formulation, $c_{t,l}$ is the value of a leaf node l in tree t, $y_{t,l}$ is a binary decision variable with a value of 1 if leaf node l is selected and 0 otherwise. H_p is the threshold value for parent node p, M and N are constants specific to the dataset, and each x_i is a continuous decision variable within a range specific to the dataset that represents an input to the random forest model. The index i for x_i indicates the input dimension. For example, x_1 is the input value for the first input to the model. T is the set of every tree in the random forest, and L is the set of every leaf node for each tree in the random forest. Pl is the set of grouping for tree, parent node, left child, and input dimension, while Pr represents the group sets of (tree, parent node, right child, input dimension). Sets T, L, Pl, Pr are constructed using the decision tree structure of each tree in a random forest.

Equation (8.2) forces the first node in every tree to be selected. Equations (8.3) enforces a constraint on the number of child nodes that can be selected for a parent node, with only one child being selected if the parent node is selected and neither being selected if the parent node is not. The disjunction of Equations (8.4) represent constraints that determine whether a left child node or right child node is selected based on the threshold value for their respective parent node. The disjunction structure can be formulated as Big-M formulation as Equations (8.5) and (8.6). The value of M in each of these constraints ensures that the threshold constraints are enforced if the parent node for that threshold is not selected.

$$x_i \leq H_p + M\left(1 - y_{t,lc}\right) \qquad \forall\, (t, p, lc, i) \in Pl \tag{8.5}$$

$$x_i \geq -M\left(1 - y_{t,rc}\right) + H_p + N y_{t,rc} \qquad \forall\, (t, p, rc, i) \in Pr \tag{8.6}$$

8.2 Parallels between MILP of random forests and two-stage stochastic program

In this section, we investigate the similarities between the MILP of random forest and the two-stage stochastic programming models. Figure 8.2.2 gives the MILP of random forests (RF) (1a – 1e) and the general form of a two-stage stochastic programming model (SP) (2a-2d), and their reformulations (R-RF and R-SP). In the general form of the two-stage stochastic programming model (SP), x denotes the here-and-now decisions, taken at the beginning of the planning horizon before uncertainty realizations, where S is the set of scenarios. y_s represents the wait-and-see decisions (recourse actions) for each individual scenario after the realizations of uncertainties. Equation (2b) in SP represents the first-stage constraints, while Equation (2c) represents scenario-specific constraints in the second stage. The non-anticipativity constraints (NACs) are implicitly represented by here-and-now decisions x in the first stage. pro_s represents probabilities for scenario s and A, b, c are known parameters. T_s, h_s, W_s can be regarded as uncertain parameters under scenario s at the second stage.

MILP of random forest (RF)		Two-stage stochastic program (SP)	
$\min_{x_i, y_{t,l}} \frac{1}{N_{trees}} \sum_{t}^{T} \sum_{l}^{L} c_{t,l} y_{t,l}$	(1a)	$\min_{x, y_s} cx + \sum_{s}^{\boldsymbol{S}} pro_s y_s$	(2a)
$y_{t,1} = 1 \quad \forall\, t \in T$	(1b)	$Ax \le b$	(2b)
$y_{t,lc} + y_{t,rc} = y_{t,p} \quad \forall\, (t, lc, rc, p)$	(1c)	$T_s x \le h_s - W_s y_s \quad \forall\, s \in \boldsymbol{S}$	(2c)
$x_i \le H_p + M(1 - y_{t,lc}) \quad \forall\, (t, p, lc, i)$	(1d)	$y_s \in Y$	(2d)
$x_i \ge -M(1 - y_{t,rc}) + H_p + N y_{t,rc} \quad \forall\, (t, p, rc, i)$	(1e)		
Reformulation of MILP of random forest (R-RF)		Reformulation of two-stage stochastic program (R-SP)	
$\min_{x_i, z_{i,t}, y_{t,l}} \frac{1}{N_{trees}} \sum_{t}^{T} \sum_{l}^{L} c_{t,l} y_{t,l}$	(3a)	$\min_{x, z_s, y_s} cx + \sum_{s}^{\boldsymbol{S}} pro_s y_s$	(4a)
$y_{t,1} = 1 \quad \forall\, t \in T$	(3b)	$Az_s \le b$	(4b)
$y_{t,lc} + y_{t,rc} = y_{t,p} \quad \forall\, t, lc, rc, p$	(3c)	$T_s z_s \le h_s - W_s y_s \quad \forall\, s \in \boldsymbol{S}$	(4c)
$z_{i,t} \le H_p + M(1 - y_{t,lc}) \quad \forall\, t, p, lc, i$	(3d)	$y_s \in Y$	(4d)
$z_{i,t} \ge -M(1 - y_{t,rc}) + H_p + N y_{t,rc} \quad \forall t, p, rc, i$	(3e)	$z_s = x \quad \forall\, s$	(4e)
$z_{i,t} = x_i \quad \forall\, t, i$	(3f)		

Figure 8.2.2: MILP of random forests (RF) (1a – 1e), the general form of a two-stage stochastic program model (SP) (2a-2d), Reformulation of RF and reformulation of SP (R-RF and R-SP)

The structure of the trees in the random forest model can be visualized as scenarios, where each individual tree represents one specific scenario, s, and corresponds to one of the realizations. In SP, following the non-anticipativity principle, all scenarios should have the same values for the

related first stage variables x. Similar to SP, the input values (x_i) with dimension i are enforced to have the same values for all trees in RF. Objective functions in RF (1a) and SP (2a) are to optimize (maximize or minimize) the expected value over the set of trees and scenarios. Both RF and SP have individual tree and scenario specific constraints, respectively. The complicating variables x_i in RF (1d and 1e) and x in SP (2c) are components causing inseparable formulations.

The reformulation in Figure 8.2.2 (R-RF and R-SP) both introduce splitting variables with equalities to explicitly represent the enforcing constraints (3f) in R-RF and NACs (4e) in R-SP. By removing the enforcing constraints and NACs, we can fully decompose the R-RF and R-SP into each tree and scenario subproblems that can be solved separately and utilize the parallelization to reduce the computational complexity.

8.3 Solution frameworks and developed toolbox

As shown in the previous section, the MILP of random forest model (RF) has a similar structure with a two-stage stochastic programming model (SP). The MILP of a random forest model contains the complicating variables x_i, which implicitly enforce input variable values to be equal for all individual tree subproblems. We introduce additional continuous variables $z_{i,t}$ as the input values for each tree t in dimension i with the equality $x_i = z_{i,t}$ as similar to non-anticipativity constraints in stochastic programming models. By removing the enforcing constraints, we can fully decompose the original MILP of the random forest into each tree MILP subproblems that can be solved simultaneously. In the following section, we explore a framework that utilizes both dual and primal bounding approaches for solving the MILP of a random forest model.

8.3.1 Lagrangian decomposition method with heuristic primal bounds (LD)

Lagrangian decomposition removes complicating constraints (enforcing constraints) and generates dual bounds by dualizing enforcing constraints to the objective function with Lagrangian multipliers $\lambda_{i,t} \in R$, which is shown in Equation (8.7).

$$\min_{x,y,z} \sum_{t}^{T} \sum_{l}^{L} c_{t,l} y_{t,l} * \frac{1}{No.Trees} + \sum_{t} \sum_{i} \lambda_{i,t}(x_i - z_{i,t}) \tag{8.7}$$

If complicating constraints are removed, the resulting problems can be solved independently. The scheme of Lagrangian decomposition-based algorithms is shown in Figure 8.2. After the initialization in the first iteration, a dual bound (Z_d^k) is obtained by solving the Lagrangian decomposition problems with fixed multipliers $\lambda_{i,t}^k$ at the k[th] iteration. The primal bound (Z_p^k) at the k[th] iteration is generated by using a heuristic that uses the solution of the Lagrangian problems. In the heuristic, we obtain $(x_i)^t$ value for different t subproblems, and use $(x_i)^t$ as input to trained random forest model to generate predicted objective values and obtain the best primal bound from all tree subproblem solutions.

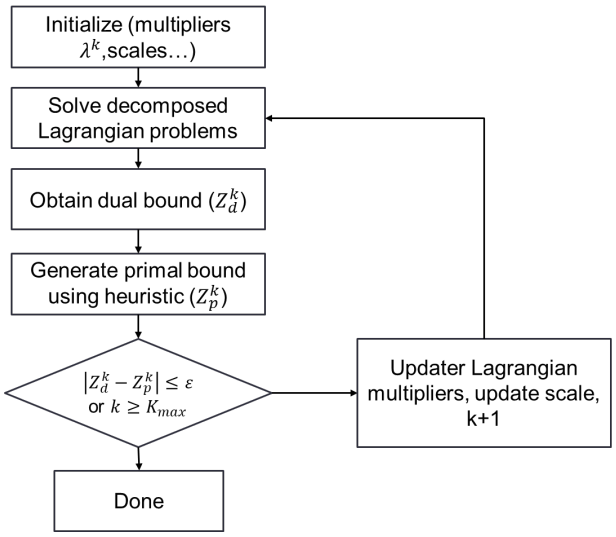

Figure 8.3: Lagrangian decomposition algorithm

At each iteration, the multipliers are updated, and the algorithm stops when either a maximum iteration/time limit is reached or the difference between the primal and dual bounds is less than a pre-specified tolerance. To obtain a tight dual bound, we proposed different ways to update Lagrangian multipliers as well as scale factor, which results in four alternatives, as shown below:

1) LD_d: Lagrangian decomposition with dynamically updating scale factor and subgradient method (Held et al. 1974) to update multipliers
2) LD_g: Lagrangian decomposition with golden ratio scheme to update scale factor and subgradient method to update multipliers
3) LDCP_d: Lagrangian decomposition with dynamically updating scale factor and cutting plane method (Oliveira et al. 2013) to update multipliers
4) LDCP_g: Lagrangian decomposition with golden ratio scheme to update scale factor and cutting plane method to update multipliers

Another alternative procedure considered is the use of the cutting plane method for solving the Lagrangian dual to update the Lagrangian multipliers. The solutions obtained from previous iterations are used to generate hyperplanes (cuts) to determine the next set of multipliers. At the k^{th} iteration, the Lagrangian multipliers can be obtained by solving the following problem:

$$max_{\eta,\lambda}\ \eta \tag{8.8}$$

$$\eta \le \sum_{t}^{T}\sum_{l}^{L} c_{t,l} y_{t,l}^{k} \frac{1}{No.Trees} + \sum_{t}\sum_{i} \lambda_{i,t}\left(x_i^k - z_{i,t}^k\right) \quad \forall k = 1, \dots, K \tag{8.9}$$

$$\begin{aligned} \lambda_{i,t}^{K-1} - scale\frac{Z_d^K - Z_p^K}{\sum_{i,t}\left(x_i^K - z_{i,t}^K\right)^2}\left|x_i^K - z_{i,t}^K\right| \le \lambda_{i,t} \\ \le \lambda_{i,t}^{K-1} + scale\frac{Z_d^K - Z_p^K}{\sum_{i,t}\left(x_i^K - z_{i,t}^K\right)^2}\left|x_i^K - z_{i,t}^K\right| \ \forall i \in I, t \in T \end{aligned} \tag{8.10}$$

Similar to Oliveira et al. (2013), we introduce constraints (8.10) as trust-region using subgradient information and dynamically update constraints to avoid the unboundedness issues. The scale value is also updated following two alternative strategies: LDCP_d and LDCP_g are defined to represent two Lagrangian decomposition algorithms using the cutting plane method with dynamic updating and golden ratio updating schemes, respectively.

8.3.2 Framework with Progressive hedging and Lagrangian decomposition (PHLD)

To further regulate the convergence of $x_i = z_{i,t}$ (complicating constraints), we introduce nonlinear terms $(x_i - z_{i,t})^2$ with penalty factor $\rho > 0$ in the objective function shown as Equation (8.11).

$$\min_{x,y,z} \sum_{t}^{T} \sum_{l}^{L} c_{t,l} y_{t,l} * \frac{1}{No.Trees} + \sum_{t} \sum_{i} \lambda_{i,t}(x_i - z_{i,t}) + \frac{\rho}{2} \sum_{t} \sum_{i} (x_i - z_{i,t})^2 \tag{8.11}$$

However, with nonlinear penalty terms, the mathematical formulation is not decomposable. Progressive Hedging (PH) was first proposed by Rockafellar and Wets (1991). It has been successfully applied as a heuristic for solving stochastic programs and is a decomposition technique to separate nonlinear terms in objective function. We developed a hybrid framework (PHLD) utilizing Progressive Hedging with Lagrangian decomposition, and the algorithm is summarized in Figure 8.4.

1. $k := 0$
2. For all tree indices $t \in T$

   $Z_{i,t}^0 = \underset{z,y}{arg\min} \sum_t^T \sum_l^L c_{t,l} y_{t,l} * \frac{1}{No.Trees} : (z,y) \in Q_t$
3. $x_i^0 := \frac{\sum_t Z_{i,t}^0}{|T|}$
4. $\lambda_{i,t}^0 := \rho(x_i^0 - Z_{i,t}^0)$
5. $k := k + 1$
6. Solve

   $Z_{i,t}^k := \arg\min_z L(y,z,\lambda) = \sum_t^T \sum_l^L c_{t,l} y_{t,l} * \frac{1}{No.Trees} + \sum_t \sum_i \lambda_{i,t}(-z_{i,t}) + \frac{\rho}{2} \sum_t \sum_i (-z_{i,t})^2 : y,z \in Q_t$

   $\lambda_{i,t}^k := \lambda_{i,t}^{k-1} + \rho(x_i^{k-1} - Z_{i,t}^{k-1})$

   $x_i^k := \frac{\sum_t Z_{i,t}^k}{|T|}$
7. Use solution x_i^k as input to Random Forest model, and generate primal bound $primal_{RF}^k$
8. Use solution $Z_{i,t}^k$ to generate heuristic primal bound $primal_{\text{he}}^k$
9. Generate dual bound using $\lambda_{i,t}^k$ by solving Lagrangian decomposition model
10. $primal_{best}^k = \min(primal_{RF}^k, primal_{\text{he}}^k)$
11. Update ρ
12. If the termination criteria are not met, then go to step 5.

Figure 8.4: The algorithm scheme of Progressive hedging with Lagrangian decomposition (PHLD)

At iteration zero, the value of $Z_{i,t}^0$ is initialized by solving the decomposable subproblems of the original problem without enforcing constraints. Besides the first iteration, $Z_{i,t}^k$ at the k[th] iteration, are generated solving decomposed subproblems without common decision variables x_i^{k}. Q_t represents the feasible sets for subproblem t. The common decision variable x_i^0 is taken as the average value of $z_{i,t}$ from each individual tree subproblems. Then the Lagrangian multipliers are updated using penalty factor ρ and the difference between x_i^0 and $Z_{i,t}^0$. The solutions x_i^k and $Z_{i,t}^k$ are used to generate the best primal bounds. The dual bounds are obtained by solving the Lagrangian decomposition model with calculated multipliers $\lambda_{i,t}^k$. At every iteration, the penalty factor ρ is updated and if the solutions have not converged sufficiently and the maximum wall time is not exceeded, the process starting with step 5 is repeated.

As mentioned in Watson and Woodruff (2011), the selection of ρ is important to convergence of PH, and the quality of generated bounds is sensitive to the choice of ρ. We propose three different strategies

for determining the value of ρ, which results in three alternative frameworks based on the calculation of ρ shown in Figure 8.5.

PHLD1: $\rho^k = \frac{\overline{c_{t,l}}}{(Z^{max} - Z^{min}+1)\times|T|\times|I|}$

PHLD2: $\rho^k = \frac{1}{(Z^{max} - Z^{min}+1)\times|T|\times|I|}$

PHLD3:

IF iteration<40:

IF iteration%2 == 0:

$\rho^{k+1} = 2\rho^k$

ELSE:

$\rho^{k+1} = \rho^k$

ELSE $\rho^{k+1} = \rho^k$

Figure 8.5: Three Alternative strategies to update penalty factor

PHLD1 and PHLD2 are variable-dependent strategies. After the k[th] iteration, for each variable $Z_{i,t}^k$, we define $Z^{max} = max_{i\in I, t\in T} Z_{i,t}^k$ and $Z^{min} = min_{i\in I, t\in T} Z_{i,t}^k$. PHLD1 use average leaf node value $\overline{c_{t,l}}$ to calculate ρ, while PHLD2 uses 1 in the numerator. PHLD3 is a heuristic space search strategy, where ρ^0 is given sufficient small value at initial iteration, and doubled by every two iterations.

8.3.3 Framework with alternating direction method of multipliers and Lagrangian decomposition (ADLD)

Alternating direction method of multipliers (ADMM) was first introduced by Gabay and Mercier (1976) and use partial updates to break optimization problem into smaller pieces. The hybrid framework (ADLD) is developed by incorporating ADMM with Lagrangian decomposition. ADLD and PHLD both allow separating the objective function (Equation (8.11)) into decomposable formulations. While PHLD fixes the x_i and obtains its value by enforcing $z_{i,t}$ from each subproblem shown in Figure 8.4, the ADLD generates and jointly solves two subproblems containing only variables x_i or variables $z_{i,t}$ in either subproblem. Both PHLD and

ADLD utilize LD to generate their dual bounds with fixed Lagrangian multipliers $\lambda_{i,t}$ calculated based on weight ρ. ADLD consists of the iterations shown in Figure 8.6.

$$y_t^{k+1}, Z_{i,t}^{k+1} := \arg\min_{y,z\in Q_t} L(x^k, y, z, \lambda^k) = \sum_t^T \sum_l^L c_{t,l} y_{t,l} \frac{1}{No.Trees} + \sum_t \sum_i \lambda_{i,t}(-z_{i,t}) + \frac{\rho}{2} \sum_t \sum_i (\bar{x}_i^k - z_{i,t})^2$$

$$x_i^{k+1} := \arg\min_{x\in Q_t} L(x, y^k, z^k, \lambda^k) = \sum_t^T \sum_l^L c_{t,l} y_{t,l} * \frac{1}{No.Trees} + \sum_t \sum_i \lambda_{i,t}(-z_{i,t}) + \frac{\rho}{2} \sum_t \sum_i (\bar{x}_i^k - z_{i,t})^2$$

$$\lambda_{i,t}^{k+1} := \lambda_{i,t}^k + \rho(x_i^{k+1} - Z_{i,t}^{k+1})$$

Figure 8.6: Main iterations of ADMM

The algorithm scheme of ADLD is similar to PHLD in ways of obtaining heuristic primal bounds. ADLD also requires updating the penalty term. We implement the same strategies show in Figure 8.4 in ADLD to update the penalty factor ρ and thus have three alternatives defined as ADLD1, ADLD2, and ADLD3.

8.3.4 Clustering strategies

To address and reduce the impact of each individual tree's error, we also investigate clustering strategies with proposed decomposition approaches. Instead of considering a splitting variable representation, clustering strategies consider a set of tree clusters when decomposing the original tree ensemble. In each tree cluster, an additional binary variable $z_{i,g}$ is defined for each tree group g, which implicitly keeps a subset of the enforcing constraints in the subproblems. Therefore, we can reduce the number of subproblems and additional binary variables. We apply clustering strategies in conjunction with the previously discussed three approaches, and we refer to the approaches with clustering strategies as LD-group, PHLD-group, ADLD-group.

8.4 Results and Discussion

8.4.1 Solution approach results and solution times

We applied the developed frameworks for solving the MILPs of RF models. The tested frameworks with combined dual and primal approaches are summarized in Table 8.1.

Table 8.1: Compile of tested frameworks

	PH +primal heuristic			PH_group +primal heuristic			ADMM +primal heuristic			ADMM_group +primal heuristic			Primal heuristic
	PH_1	PH_2	PH_3	PH_1	PH_2	PH_3	PH_1	PH_2	PH_3	PH_1	PH_2	PH_3	
LD (penalty term updating)	PHLD1	PHLD2	PHLD3	PHPLD1 (group)	PHPLD2 (group)	PHPLD3 (group)	ADLD1	ADLD2	ADLD3	ADLD1 (group)	ADLD2 (group)	ADLD3 (group)	
LD subgradient (dynamic updating scheme)													LD_d
LD subgradient (golden ration updating scheme)													LD_g
LD Cutting plane (dynamic updating scheme)													LDCP_d
LD Cutting plane (golden ration updating scheme)													LDCP_g
LD subgradient Group (dynamic updating scheme)													LD_d (group)
LD subgradient Group (golden ration updating scheme)													LD_g (group)
LD Cutting plane Group (dynamic updating scheme)													LDCP_d (group)
LD Cutting plane Group (golden ration updating scheme)													LDCP_g (group)

As shown in Table 8.1, we compile 20 proposed decomposition frameworks with different approaches combined with Lagrangian decomposition with cutting plane method (LD), Progressive Hedging (PH), and Alternating Direction Method of Multipliers (ADMM). Twenty frameworks are applied to solve the MILPs of 282 trained RF models with a total number of trees ranging from 10 to 100. These RF models were developed to approximate the test functions from the Virtual Library of Simulation Experiments (Surjanovi and Bingham, 2013).

The functions in the library are grouped by shape, which includes five main categories: multi-local minima (75 functions), bowl-shaped (90 functions), plate-shaped (27 functions), valley-shaped (36 functions), and other-shaped (54 functions), which contains functions that do not fit into the other four categories.

All approaches are implemented in Pyomo 5.6.6 and Python 3.6.4 and solved in Auburn University Hopper Cluster. The maximum number of iterations is set to 100. The results for the optimization performance of the 20 proposed approaches in 282 trained RF models are summarized in Table 8.2.

Table 8.2: Results of 282 trained RF model for twenty frameworks

Frameworks	Average gap from optimum (primal)	Average gap from optimum (dual)	Average relative gap	Number of cases (< 1% gap from optimum primal)	Number of cases (< 1% gap from optimum primal)	Number of cases (< 1% relative gap)
PHLD1	17.18%	29.64%	37.35%	87	43	39
PHLD2	17.66%	43.66%	51.12%	91	25	25
PHLD3	18.48%	40.35%	48.05%	85	28	28
ADLD1	12.32%	37.75%	42.23%	100	29	29
ADLD2	11.70%	37.34%	41.94%	108	28	27
ADLD3	17.91%	38.39%	46.24%	90	28	22
LD_d	32.18%	38.30%	53.75%	46	28	24
LD_g	32.44%	38.43%	53.89%	47	28	24
LDCP_d	31.84%	37.83%	53.18%	46	28	24
LDCP_g	32.18%	37.96%	53.43%	47	28	24
PHLD1 Group	16.25%	21.58%	30.29%	78	44	42
PHLD2 Group	14.46%	21.31%	29.14%	91	47	44
PHLD3 Group	7.07%	19.37%	22.72%	159	45	43
ADLD1 Group	10.42%	19.65%	26.17%	122	48	42
ADLD2 Group	12.21%	20.06%	26.51%	96	49	44
ADLD3 Group	6.87%	19.34%	22.65%	160	46	44
LD_d Group	21.47%	21.88%	34.81%	57	44	38
LD_g Group	21.81%	21.90%	35.06%	57	44	39
LDCP_d Group	21.31%	21.85%	34.76%	58	45	39
LDCP_g Group	21.64%	21.89%	34.96%	58	45	39

Table 8.2 summarizes the average optimality, dual, and relative gaps. The number of cases within 1% gap from optimum (primal, dual bounds) and relative gaps are also included in Table 8.2. The table shows that the updating schemes of penalty factor impact the performance of PHLD and ADLD alternative approaches, while the cutting plane method and subgradient method for

Lagrangian decomposition-based approaches have similar performances. Clustering strategies improve the performance of the proposed approaches such that the ADLD3 group obtains tighter average gap from optimum for primal bounds (6.87%) than ADLD3 (17.91%) and reduces the average gaps by 62%. Within 100 iterations, the ADLD3 group obtained the tightest relative gaps and average gap from the optimum for both primal and dual bounds. ADLD3 group was able to obtain feasible solutions within 1% average gap from optimum for primal bounds for 56.7% (160) trained models and within 1% relative gaps for 15.6% trained models.

The solution time of the proposed frameworks is dependent on the type of trained RF models and increases with larger input dimensions and the number of trees. Most cases are solved faster in CPLEX than the proposed approaches within 100 trees. There are still more than thirty cases that are solved faster using the proposed frameworks than directly solving MILP in CPLEX. We take the solutions of the ADLD3 group as an example. Table 8.3 summarizes the trained RF models that have less solution time in the ADLD3 group than MILP. From Table 8.3, the ADLD3 group is able to reduce solution times up to one order of magnitude for the 34 trained RF models. The solution times of CPLEX for trained models also depend on the function shape. For example, CPLEX took over 86 hours for Perm_db 8D, in which the ADLD3 group was able to obtain the optimum within 28 hours.

Table 8.3: Cases with less solution times in ADLD3 group compared to CPLEX

Number of trees	Function name	Dimension	Solution time (seconds)	
			MILP	ADLD3 group
10	Sphere	2D	8	5
10	Ellipsoid	2D	8	5
10	Sum of Squares	2D	7	4
10	Sum of Different Powers	2D	7	5
10	Trid	2D	6	5
10	Bohachevsky	2D	8	5
10	3Hump	2D	4	3
10	Ackley	2D	7	3
10	Griewank	2D	9	5
10	Rastrigin	2D	4	4
50	Sphere	2D	17	10
50	Ellipsoid	2D	22	11
50	Sum of Squares	2D	19	10
50	Sum of Different Powers	2D	18	10
50	Trid	2D	18	11
50	Bohachevsky	2D	21	10
50	3Hump	2D	17	10
50	Perm_db	8D	16607	2693
50	Ackley	2D	18	12
50	Griewank	2D	26	11
50	Schwefel	10D	5782	2312
50	Rastrigin	2D	18	15
100	Sphere	2D	40	21
100	Ellipsoid	2D	48	23
100	Sum of Squares	2D	43	22
100	Sum of Different Powers	2D	54	23
100	Trid	2D	40	32
100	Bohachevsky	2D	50	24
100	3Hump	2D	43	22
100	Perm_db	6D	8015	5260
100	Perm_db	8D	309650	100399
100	Ackley	2D	47	24
100	Griewank	2D	72	26
100	Schwefel	10D	102500	23483

8.4.2 Conclusions

The frameworks investigated here can be used to exploit the special structure of the RF model MILPs, which is similar to two-stage stochastic programs. The frameworks fully decompose the MILPs of the RF model to individual tree subproblems and update the dual and primal bounds

iteratively with different approaches. Twenty different frameworks are tested with 282 trained RF models with the number of trees ranging from 10 to 100. The performance of the frameworks depends on the function shape and the characteristics of trained RF models. The computational experiments revealed that the developed approaches were able to obtain optimality gaps for functions such as Sphere functions, Ellipsoid functions, and Trid functions within a few iterations. The approaches that utilize clustering strategies were able to obtain significantly smaller relative gaps compared to original approaches. Within 100 iterations, the computational experiments revealed that the proposed frameworks were able to obtain less than 1% gaps from optimum (feasible solutions) for 56.7% trained models and less than 1% relative gaps for 15.6% trained models.

CHAPTER 9

CONCLUSION AND FUTURE DIRECTION

In this dissertation, we developed models and solution approaches for large-scale MSSPs under endogenous uncertainties.

9.1 MSSP models for artificial lift infrastructure planning problem under production rate uncertainty

Chapter 3 presents and compares two discrete-time large-scale nonconvex mixed-integer nonlinear programming (MINLP) models (F0, F1) to solve the artificial lift infrastructure planning problem. Using the nonlinear terms' unique structure, we formulate two equivalent mixed-integer linear programming (MILP) models (LF0, LF1). Considering the well's production limitations, we develop a set of valid inequalities for tightening the original models, which results in four additional models (TF0, TF1, TLF0, and TLF1). The solution times of the tighter models are up to two orders of magnitude shorter than that of the original ones. To incorporate the impact of endogenous uncertainty in ALM-dependent production rates, we develop two stochastic MILP models (SMILP0 and SMILP1) based on TLF0 and TLF1. We also extend these models considering shale gas prices as a source of exogenous uncertainty. For a hypothetical case study under only endogenous uncertainties, the value of the stochastic solution (VSS) is $163,831, which is a 5% increase with respect to the deterministic solution.

9.2 Primal bounding approaches for MSSPs under endogenous uncertainties

Chapter 4 presents a generalized Knapsack-problem-based Decomposition Algorithm (GKDA) to efficiently obtain feasible solutions for large-scale MSSPs under endogenous and/or exogenous uncertainties. The GKDA decomposes the original MSSP into a series of knapsack

problems and solves these problems at appropriate decision points of the planning horizon. We applied GKDA to obtain feasible solutions for instances of four planning problems, which include continuous and/or discrete decision variables, and endogenous and/or exogenous uncertain parameters. The comparison of solutions obtained by the GKDA to the optimum solutions for these problems revealed that the GKDA, in most cases, yielded tight feasible solutions with solution times up to six orders of magnitude faster than that of the deterministic equivalents.

Chapter 5 develops a general primal-bounding framework based on extending the concepts of expected value solution and value of the stochastic solution from multistage stochastic programs under exogenous uncertainties. Under the traditional decision-making process with uncertainties, decision-makers utilize known information to-date along with the expected results for unrealized information to take action, realize some uncertainty, and repeat this process along the planning horizon. The bounding framework introduced here fits this decision-making process. It yields a tight feasible bound and an implementable solution for multistage stochastic programs under endogenous uncertainties, which we call the absolute expected value solution (AEEV). The framework is tested on three planning problems with up to 16,384 scenarios. It yielded primal bounds within 1% of the true solutions for all tested cases and generated these implementable solutions up to four orders of magnitude faster than solving the original multistage stochastic programs.

9.3 Dual bounding approaches for MSSPs under endogenous uncertainties

In Chapter 6, we present two dual bounding approaches and utilize GKDA to efficiently generate dual bounds for large-scale MSSPs under endogenous uncertainties. The first approach is a new Lagrangian relaxation, which regards NACs as complicating constraints and utilizes the unique structure of the NACs to formulate a tighter dual problem, which also eliminates half of

the Lagrange multipliers. We also developed two different multiplier updating schemes to update multipliers. The first scheme dynamically updated the multipliers based on the relative gap of the Lagrangian dual solutions between two iterations. The second scheme utilized the ratio of dualized constraints in the objective function to the original objective function value. We applied the new Lagrangian relaxation to solve instances of artificial lift infrastructure planning (ALIP) and the clinical trial planning problems. The computational results reveal that the new Lagrangian relaxation generates tight dual bounds compared to the existing Lagrangian relaxation formulation. Both proposed multiplier updating schemes reduce the zigzagging behavior of the Lagrangian dual solutions as iterations progress.

The second approach presents a Relaxed Knapsack-problem-based decomposition Algorithm (RKDA)and employs GKDA to solve a relaxed version of the original MSSP. The relaxed MSSP is obtained by removing its resource constraints. We applied GKDA and RKDA to generate initial primal and dual bounds for the pharmaceutical clinical trial planning problem with two-, three-, four-, five-, six-, seven- and ten-products and three clinical trials with different planning horizon lengths. The results revealed that the resulting relative gaps ranged from 2.10% to 4.66%, and GKDA and RKDA yielded these gaps in 0.1 to 2377.0 CPU seconds.

9.4 A decomposition framework for MSSP under endogenous uncertainty

In Chapter 7, the development of a decomposition framework (LGKDA) for solving MSSPs has shown convergence in small cases. The dual bounds are generated by Lagrangian decomposition with relaxing all NACs for parallelization. The LGKDA generates primal bounds by GKDA, which are iteratively improved with fixed items using solutions from the Lagrangian decompositions. We applied LGKDA to generate solutions for the pharmaceutical clinical trial

planning problem. The relative gaps obtained by LGKDA is from 0.021% to 7.15% within 100 iterations.

9.5 Decomposition frameworks for surrogate-based optimization with random forests

We proposed different frameworks as well as clustering approaches to solve the MILP of the random forest model. The MILP of random forest is similar to stochastic programs, where each the inputs to individual tree subproblems are enforced to be equal to each other and generated the expected objective values. The proposed frameworks fully decompose the MILPs of the RF model to individual tree subproblems and updated the dual and primal bounds iteratively with different approaches. Two hundred eighty-two trained models from five main categories with the number of trees ranging from 10 to 100 are solved with proposed frameworks. The computational experiments revealed that 56.7% of total trained models can be obtained feasible solutions within 1% gaps from optimum with the proposed frameworks, and 15.6% trained models can be converged within 1% relative gaps.

9.6 Recommendations for future work

In this dissertation, the optimal development of ALIP is based on the assumption that the uncertainty in the production rate is realized as soon as an ALM is installed at the well. In reality, the uncertainty can be learned gradually and related to the cumulative production at the well. Besides, artificial lift infrastructure for multiple wells under the production rate should be considered.

As shown in the review of MSSP models under endogenous uncertainty, current schemes of scenario generation are limited to the Cartesian product of outcomes of uncertainties (9 of 11 models) and arbitrary scenario sets (2 of 11 models). Unlike the fixed scenario trees with

exogenous uncertainties, the structure of scenario trees under endogenous uncertainties are decision dependent. In the future, an analysis of the impact of different sampling methods to generate scenario sets on MSSPs under endogenous uncertainties may be useful for addressing the computational complexities as well as generating improved solutions for the MSSPs.

GKDA and AEEV are both process-based approaches, which only generate and solve an initial subproblem based on expected information and obtain tight feasible solutions to instances with a significantly large number of scenarios (over one million scenarios). GKDA and AEEV do not suffer from space and time complexity, and hence, the proposed approaches are initial steps towards solving large-scale MSSP for real-world sized problems. It is recommended to develop iterative schemes for GKDA and AEEV to converge to optimums as future work.

Two approaches (RKDA and mLR) have been shown to generate tight dual bounds in Chapter 6. However, proposed dual bounding approaches for MSSP under endogenous uncertainties are still limited to a moderate number of scenarios and certain types of problems. A modified Lagrangian relaxation formulation has been developed to reduce the total number of dualized constraints and multipliers and obtains tight dual bound by utilizing a disjunctive structure. However, the inseparable formulation of mLR limits its applications to cases with a large number of scenarios. The decomposable formulation of mLR is recommended to utilize parallelization in future work.

We have developed an RKDA, which implicitly incorporates the NACs in the framework and relax the resource constraints to generate a dual bound for resource-constraints multiple project scheduling with stochastic task success. As mentioned in Chapter 7, since RKDA generates tighter initial dual bounds than approaches relaxing the NACs and does not suffer from the computational complexity associated with updating Lagrangian multipliers, we recommend exploring the

process-based framework similar to RKDA for general MSSP under endogenous uncertainties as to future work.

As the NACs contribute to the main computation complexities, most decomposition frameworks relax the initial and conditional NACs and decompose original MSSP to individual scenario subproblems. However, with a significantly large number of scenarios (over one billion scenarios in a 15-Drug case of clinical trial planning), the CPU times required to solve the dual problem and the updating procedure between iterations are critical. We recommend investigating the process-based framework, incorporating the NACs without computational complexity, and relaxing scenario-specific constraints to form a tighter form.

REFERENCES

Ansotegui, C., M. Bofill, M. Palahi, J. Suy and M. Villaret (2011). Satisfiability modulo theories: An efficient approach for the resource-constrained project scheduling problem. Proceedings of the Ninth Symposium on Abstraction, Reformulation and Approximation, Parador de Cardona, Cardona, Catalonia, Spain, The AAAI Press.

Apap, R.M. and Grossmann, I.E., Models and computational strategies for multistage stochastic programming under endogenous and exogenous uncertainties. Computers & Chemical Engineering, 103, 2017; pp.233-274.

Ardalan, Z., Karimi, S., Naderi, B., & Khamseh, A. A. (2016). Supply chain networks design with multi-mode demand satisfaction policy. Computers & Industrial Engineering, 96, 108-117.

Breiman L (2001) Random forests. Mach Learn 45:5–32.

Birge JR, Louveaux F (2011). Introduction to stochastic programming (Springer Science & Business Media)

Bartusch, M., R. H. Möhring and F. J. Radermacher (1988). "Scheduling project networks with resource constraints and time windows." Annals of Operations Research 16(1): 199-240.

Bondurant, A. B., Dotson, B. D., and Oyewole, P. O. (2007). “Getting the Last Gasp: Deliquification of Challenging Gas Wells,” SPE paper 11651 presented at The International Petroleum Technology Conference, Dubai.

Boland, N., Dumitrescu, I., & Froyland, G. (2008). A multistage stochastic programming approach to open pit mine production scheduling with uncertain geology. Optimization online, 1-33.

Boyd, Stephen P.; Vandenberghe, Lieven (2004). Convex Optimization. Cambridge University Press. p. 216. ISBN 978-0-521-83378-3.

Conejo, A. J., Castillo, E., Mínguez, R., & García-Bertrand, R. (2006). Decomposition in Linear Programming: Complicating Constraints. Decomposition Techniques in Mathematical Programming: Engineering and Science Applications, 67-106.

Christian, B., & Cremaschi, S. (2015). Heuristic solution approaches to the pharmaceutical R&D pipeline management problem. Computers & Chemical Engineering, 74, 34-47.

Christian, B., & Cremaschi, S. (2017). Variants to a knapsack decomposition heuristic for solving R&D pipeline management problems. Computers & Chemical Engineering, 96, 18-32.

Christian, B., & Cremaschi, S. (2017). A branch and bound algorithm to solve large-scale multistage stochastic programs with endogenous uncertainty. AIChE Journal.

Christian, B. (2017). Solution Approaches to Large Scale Multistage Stochastic Programs with Endogenous and Exogenous Uncertainty.

Christian, B., & Cremaschi, S. (2018). A multistage stochastic programming formulation to evaluate feedstock/process development for the chemical process industry. Chemical Engineering Science, 187, 223-244.

Cafaro, Diego C., Markus G. Drouven, and Ignacio E. Grossmann. "Optimization models for planning shale gas well refracture treatments." AIChE Journal 62.12 (2016): 4297-4307

Clegg, J. D.* SPE, Shell Oil Co. (1988). High-Rate Artificial Lift, 17638-PA SPE Journal Paper

Clegg, J. D., Bucaram, S. M., and Jr., N. W. H. (1993). Recommendations and Comparisons for Selecting Artificial-Lift Methods (Includes Associated Papers 28645 and 29092). SPE Journal of Petroleum Technology 45 (12)

Colvin, M., & Maravelias, C. T. (2008). A stochastic programming approach for clinical trial planning in new drug development. Computers & Chemical Engineering, 32(11), 2626-2642.

Colvin, M., & Maravelias, C. T. (2009). Scheduling of testing tasks and resource planning in new product development using stochastic programming. Computers & Chemical Engineering, 33(5), 964-976.

Colvin, M., & Maravelias, C. T. (2010). Modeling methods and a branch and cut algorithm for pharmaceutical clinical trial planning using stochastic programming. European Journal of Operational Research, 203, 205-215.

Christofides, N., R. Alvarez-Valdes and J. M. Tamarit (1987). "Project scheduling with resource constraints: A branch and bound approach." European Journal of Operational Research 29(3): 262-273.

Donald, J., Marino, M., and Veeken, K. (2014). “Standardized Mobile Wellhead Compressor for Onshore Gas Wells in the Netherlands,” Gas Well Deliquification Workshop, Denver, 23 – 26 February

Drouven, Markus G., and Ignacio E. Grossmann. "Multi-period planning, design and strategic models for long-term, quality-sensitive shale gas development." AIChE Journal (2016)

Escudero, L. F., Garín, A., Merino, M., & Pérez, G. (2007). The value of the stochastic solution in multistage problems. Top, 15(1), 48-64.

Escudero, L. F., Garín, M. A., & Unzueta, A. (2016). Cluster Lagrangean decomposition in multistage stochastic optimization. Computers & Operations Research, 67, 48-62.

Fahmi, I., Nuchitprasittichai, A., & Cremaschi, S. (2014). A new representation for modeling biomass to commodity chemicals development for chemical process industry. Computers & Chemical Engineering, 61, 77-89

FDA. FDA Drug Approval Process—US Food and Drug Administration. 2017. Available online: http://www.fda.gov/downloads/Drugs/ResourcesForYou/Consumers/UCM284393.pdf (accessed on 28 August 2017).

FDA. FDA Drug Approval Process—US Food and Drug Administration. Available online: https://www.fda.gov/forpatients/approvals/drugs/ucm405622.htm (accessed on 21 August 2017).

Fetkovich, M. J., Fetkovich, E. J., & Fetkovich, M. D. (1996). Useful Concepts for Decline Curve Forecasting, Reserve Estimation, and Analysis. Society of Petroleum Engineers

Fisher, M. L. (1981). The Lagrangian relaxation method for solving integer programming problems. Management science, 27(1), 1-18.

Gatica, G.; Papageorgiou, L.G.; Shah, N. Capacity planning under uncertainty for the pharmaceutical industry. Chem. Eng. Res. Des. 2003, 81, 665–678.

Goel, V., & Grossmann, I. E. (2006). A class of stochastic programs with decision dependent uncertainty. Mathematical programming, 108(2), 355-394.

Goel, V., & Grossmann, I. E. (2004). A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. Computers & chemical engineering, 28(8), 1409-1429.

Goel, V., Grossmann, I. E., El-Bakry, A. S., & Mulkay, E. L. (2006). A novel branch and bound algorithm for optimal development of gas fields under uncertainty in reserves. Computers & chemical engineering, 30(6-7), 1076-1092.

Gupta, V., & Grossmann, I. E. (2011). Solution strategies for multistage stochastic programming with endogenous uncertainties. Computers & Chemical Engineering, 35(11), 2235-2247.

Gupta, V., & Grossmann, I. E. (2014a). A new decomposition algorithm for multistage stochastic programs with endogenous uncertainties. Computers & Chemical Engineering, 62, 62-79.

Gupta, V., & Grossmann, I. E. (2014b). Multistage stochastic programming approach for offshore oilfield infrastructure planning under production sharing agreements and endogenous uncertainties. Journal of Petroleum Science and Engineering, 124, 180-197.b

Gao, Jiyao, and Fengqi You. "Design and optimization of shale gas energy systems: Overview, research challenges, and future directions." Computers & Chemical Engineering (2017)

Giannelos, S., Konstantelos, I., & Strbac, G. (2018). Option value of demand-side response schemes under decision-dependent uncertainty. IEEE Transactions on Power Systems, 33(5), 5103-5113.

Gabay, D., & Mercier, B. (1976). A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Computers & Mathematics with Applications, 2, 17-40.

Heinze, L. R., Thornsberry, K., Witt, L. D. (1989). AL: An Expert System for Selecting the Optimal Pumping Method, SPE 18872, SPE Production Operations Symposium, Oklahoma City, Oklahoma

Hart, W.E.; Laird, C.; Watson, J.P.; Woodruff, D.L. Pyomo–Optimization Modeling in Python; Springer Science & Business Media: New York, NY, USA , 2012; Volume 67

Held, M., Wolfe, P., & Crowder, H. (1974). Validation of subgradient optimization. Mathematical Programming, 6(1), 62–88.

Hooshmand Khaligh, F., & MirHassani, S. A. (2016). A mathematical model for vehicle routing problem under endogenous uncertainty. International Journal of Production Research, 54(2), 579-590.

IMS. Institute Global Use of Medicines: Outlook through 2016; Future Prescr; IMS: Parsippany, NJ, USA , 2013; Volume 14, pp. 2–10

Jonsbraten, T.W., Wets, R.J. B., Woodruff, D.L., 1998. A class of stochastic programs with decision dependent random elements. Annals of Operations Research 82, 83-106.

Koulinas, G., L. Kotsikas and K. Anagnostopoulos (2014). "A particle swarm optimization based hyper-heuristic algorithm for the classic resource constrained project scheduling problem." Information Sciences 277: 680-693.

Liess, O. and P. Michelon (2008). "A constraint programming approach for the resource-constrained project scheduling problem." Annals of Operations Research 157(1): 25-36.

Misener, R.; Thompson, J. P.; Floudas, C. A. APOGEE: Global optimization of standard, generalized, and extended pooling problems via linear and logarithmic partitioning schemes. Computers & Chemical Engineering. 2011, 35(5), 876-892

Naber, A. and R. Kolisch (2014). "MIP models for resource-constrained project scheduling with flexible resource profiles." European Journal of Operational Research 239(2): 335-348.

Oral, M., & Kettani, O. (1992). A linearization procedure for quadratic and cubic mixed-integer problems. Operations Research, 40(1-supplement-1), S109-S116.

Oliveira, F., et al. (2013). A Lagrangean decomposition approach for oil supply chain investment planning under uncertainty with risk considerations. Computers & Chemical Engineering, 50, 184-195.

PhRMA. Biopharmaceutical Research Industry Profile, Pharmaceutical Research and Manufacturers of America; PhRMA: Washington, DC, USA, 2016

Solak, S., Clarke, J. P. B., Johnson, E. L., & Barnes, E. R. (2010). Optimization of R&D project portfolios under endogenous uncertainty. European Journal of Operational Research, 207(1), 420-433

Sundaramoorthy, A.; Evans, J.M.B.; Barton, P.I. Capacity planning under clinical trials uncertainty in continuous pharmaceutical manufacturing, 1: Mathematical framework. Ind. Eng. Chem. Res. 2012, 51, 13692–13702

Surjanovic, S. & Bingham, D. (2013). Virtual Library of Simulation Experiments: Test Functions and Datasets. Retrieved November 23, 2020, from http://www.sfu.ca/~ssurjano.

Tarhan, B., & Grossmann, I. E. (2008). A multistage stochastic programming approach with strategies for uncertainty reduction in the synthesis of process networks with uncertain yields. Computers & Chemical Engineering, 32(4), 766-788.

Tarhan, B., Grossmann, I. E., & Goel, V. (2009). Stochastic programming approach for the planning of offshore oil or gas field infrastructure under decision-dependent uncertainty. Industrial & Engineering Chemistry Research, 48(6), 3078-3097

Tarhan, B., Grossmann, I. E., & Goel, V. (2013). Computational strategies for non-convex multistage MINLP models with decision-dependent uncertainty and gradual uncertainty resolution. Annals of Operations Research, 203(1), 141–166. http://doi.org/10.1007/s10479-011-0855-x

Misener, R., Thompson, J. P., & Floudas, C. A. (2011). APOGEE: Global optimization of standard, generalized, and extended pooling problems via linear and logarithmic partitioning schemes. Computers & Chemical Engineering, 35(5), 876-892.

Mercier L., Van Hentenryck P. (2008) Amsaa: A Multistep Anticipatory Algorithm for Online Stochastic Combinatorial Optimization. In: Perron L., Trick M.A. (eds) Integration of AI and OR

Techniques in Constraint Programming for Combinatorial Optimization Problems. CPAIOR 2008. Lecture Notes in Computer Science, vol 5015. Springer, Berlin, Heidelberg

Medal, H. R., Pohl, E. A., Rossetti, M. D., Medal, H. R., Pohl, E. A., & Allocating, M. D. R. (2016). Allocating Protection Resources to Facilities When the Effect of Protection is Uncertain. IIE Transactions, 48(3), 220–234. **http://doi.org/10.1080/0740817X.2015.1078013**

Robbins, K. (2013). Awakening the Slumbering Giant: How horizontal drilling technology brought the endangered species act to bear. 63 Cas. W. Res. L. Rev

Rockafellar, R. T., & Wets, R. J. B. (1991). Scenarios and Policy Aggregation in Optimization under Uncertainty. Mathematics of Operations Research, 16, 119-147.

Valbuena, J. D. (2015). Defining the artificial lift system selection guidelines in horizontal wells. The University of Tulsa.

Weatherford, 2013, Introduction to Artificial Lift Systems <http://www.alrdc.com/recommendations/Gas%20Well%20Deliquification/artificial%20lift%20selection/WFT%20Intro%20to%20Artificial%20Lift%20Systems.pdf> (accessed 04/14/2017)

Watson, J.-P., & Woodruff, D. L. (2011). Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. Computational Management Science, 8, 355-370.

Zapata, J. C., B. M. Hodge and G. V. Reklaitis (2008). "The multimode resource constrained multiproject scheduling problem: Alternative formulations." AIChE Journal 54(8): 2101-2119.

Zeng, Z., & Cremaschi, S. (2017a). Artificial lift infrastructure planning for shale gas producing horizontal wells. Proceedings of the FOCAPO/CPC, Tuscan, AZ, USA, 8-12.

Zeng, Z., & Cremaschi, S. (2017b). Multistage Stochastic Programming Models for Pharmaceutical Clinical Trial Planning. Processes, 5(4), 71.

Zeng, Z., Christian, B., & Cremaschi, S. (2018a). A generalized knapsack-problem based decomposition heuristic for solving multistage stochastic programs with endogenous and/or exogenous uncertainties. Ind. Eng. Chem. Res., 2018, 57 (28), pp 9185–9199.

Zeng, Z., & Cremaschi, S. (2018b). Multistage Stochastic Models for Shale Gas Artificial Lift Infrastructure Planning. In Computer Aided Chemical Engineering (Vol. 44, pp. 1285-1290). Elsevier.

Zeng, Z., & Cremaschi, S. (2018c). A Relaxed Knapsack-Problem Based Decomposition Heuristic for Large-Scale Multistage Stochastic Programs. Computer Aided Chemical Engineering, vol. 43. Elsevier, pp. 519–524.

Zeng, Z., & Cremaschi, S. (2019a). A general primal bounding framework for large-scale multistage stochastic programs under endogenous uncertainties. Chemical Engineering Research and Design, 141, 464-480.

Zeng, Z., & Cremaschi, S. (2019b). A New Framework to Primal Bounding Multistage Stochastic Programs Under Endogenous and Exogenous Uncertainties. Chemical Engineering Transactions, 74, 631-636.

Zeng, Z., & Cremaschi, S. (2019c). A primal bounding approach for multistage stochastic programs of resource-constrained planning and scheduling with stochastic task success. In Computer Aided Chemical Engineering (Vol. 46, pp. 925-930). Elsevier.

Zeng, Z., & Cremaschi, S. (2020). A New Lagrangian Relaxation Approach for Multistage Stochastic Programs under Endogenous Uncertainties. In Computer Aided Chemical Engineering (Vol. 48, pp. 1855-1860). Elsevier.

APPENDIX A1: NOMENCLATURE OF ALIP

Sets

$i, j \in I$	Set of ALMS
$r, t, p, l, k \in T$	Sets of operating months

Parameters

$ArmaticContent$	Aromatic content
b	Decline exponent constant
BHP	Bottomhole Pressure
$CasingDiameter$	Casing diameter
Ce_i	Equipment cost of method *i*
Cm_i	Operating cost of method *i*
Co_i	Equipment and installation cost of method *i*
CGA	Compressed Gas Availability
$CICHP$	Closed in casing head pressure
$CITHP$	Closed in tubing head pressure
D	Nominal decline rate
DLS	Dogleg Severity
EPA	Electrical power availability
$EPEF$	Electrical power economic feasibility
FLP	Flowline pressure
FT	Federal tax rate
$FTHP$	Flowing tubing head pressure
GPA	Gas power availability
$GPEF$	Gas Power economic feasibility
GVF	Gas Void Fraction
LF	Load Factor
LT	Local tax rate
M	Upper bound of the taxable income at month r
$MARR$	Minimum acceptable rate of return
MCF	A machine casing
$Offshore$	The well offshore
$Packer$	A packer Present
Pg	Price of gas
Png	Price of natural gas liquid
Po	Price of oil
PTD	Production Tubing Diameter
Qrc_i	Flowrate change ratio when ALM i is installed

RT	Royalty tax rate
TAC	Tubing Annulus Communication
$Temp$	Temperature of well
TDS	Total Dissolved Solids
$TbingUniformity$	The tubing diameter uniform
$WellboreAccess$	Wellbore access
$WellDept$	Depth of well
$WellDevialtion$	Well deviation above end of tubing
$WaterCut$	Water Cut
WI	Working interest

Continuous Variables (0 to + ∞)

CC	Total equipment and installation cost
Dep_r	Depreciation at period r
GI_r	Gross income at period r
LFR_r	Liquid flow rate at month r
Qg_r	Gas flow rate at month r
$Qg_{i,t,p,r}$	Gas flow rate at month r with method i installed on month p and uninstalled at month t
Qng_r	Natural gas liquid flow rate at month r
$Qng_{i,t,p,r}$	Natural gas flow rate at month r with method i installed on month p and uninstalled at month t
Qo_r	Oil flow rate at month r
$Qo_{i,t,p,r}$	Oil flow rate at month r with method i installed on month p and uninstalled at month t
TI_r	Taxable income at month r

Binary Variables

$w_{i,p}$	Active if method i installed at month p
x_r	Active if taxable income is negative at month r

$y_{i,t,p}$	Active if method i installed at month p and uninstalled at month t
$z_{i,t}$	Active if method i uninstalled at month t

Auxiliary Variables (0 to +∞)

$LFRy_{i,t,p,r}$	The product of $y_{i,t,p}$ and LFR_r
$LFRy1_{i,t,p,r}$	Auxiliary variable for linearization
$LFRw_{i,p,r}$	The product of $w_{i,p}$ and LFR_r
$LFRw1_{i,p,r}$	Auxiliary variable for linearization
$wzQg_{i,t,p,r}$	The multiples of $w_{i,p}$, $z_{i,t}$ and Qg_r
$wQg_{i,p,r}$	Auxiliary variable for linearization
$wQg1_{i,p,r}$	Auxiliary variable for linearization
$wzQg1_{i,t,p,r}$	Auxiliary variable for linearization
$wzQo_{i,t,p,r}$	The multiples of $w_{i,p}$, $z_{i,t}$ and Qg_r
$wQo_{i,p,r}$	Auxiliary variable for linearization
$wQo1_{i,p,r}$	Auxiliary variable for linearization
$wzQo1_{i,t,p,r}$	Auxiliary variable for linearization
$wzQng_{i,t,p,r}$	The multiples of $w_{i,p}$, $z_{i,t}$ and Qg_r
$wQng_{i,p,r}$	Auxiliary variable for linearization
$wQng1_{i,p,r}$	Auxiliary variable for linearization
$wzQng1_{i,t,p,r}$	Auxiliary variable for linearization
XTI_r	The product of x_r and TI_r
$XTI1_r$	Auxiliary variable for linearization

$yQg_{i,t,p,r}$	The product of $y_{i,t,p}$ and $Qg_{i,t,p,r}$
$yQo_{i,t,p,r}$	The product of $y_{i,t,p}$ and $Qo_{i,t,p,r}$
$yQng_{i,t,p,r}$	The product of $y_{i,t,p}$ and $Qng_{i,t,p,r}$
$yQg1_{i,t,p,r}$	Auxiliary variable for linearization
$yQo1_{i,t,p,r}$	Auxiliary variable for linearization
$yQng1_{i,t,p,r}$	Auxiliary variable for linearization
$zw_{i,t,p}$	The product of $z_{i,t}$ and $w_{i,p}$

APPENDIX A2: PHYSICAL LIMITATION CONSTRAINTS FOR ALL ARTIFICIAL LIFT METHODS

Plunger lift, $i = 1$

$$LFR_r \cdot y_{1,t,p} \leq 200 \;\; \forall t, p, r \in T, p \leq r \leq t \tag{A1}$$

IF *TubingSize* < 2.875:

$$LFR_r \cdot y_{1,t,p} \leq 165 \;\; \forall t, p, r \in T, p \leq r \leq t \tag{A2}$$

IF *TubingSize* < 2.375:

$$LFR_r \cdot y_{1,t,p} \leq 110 \;\; \forall t, p, r \in T, p \leq r \leq t \tag{A3}$$

IF *TubingSize* < 2.0:

$$LFR_r \cdot y_{1,t,p} \leq 50, \;\; \forall t, p, r \in T, p \leq r \leq t \tag{A4}$$

IF *TubingSize* < 1.5:

$$LFR_r \cdot y_{1,t,p} \leq 35 \;\; \forall t, p, r \in T, p \leq r \leq t \tag{A5}$$

IF *Packer* = 1:

$$Qg_r \geq 1000 \cdot LFR_r \cdot y_{1,t,p} \;\; \forall t, p, r \in T, p \leq r \leq t \tag{A6}$$

IF *Packer* = 1:

$$Qg_r \geq 400 \cdot LFR_r \cdot y_{1,t,p} \;\; \forall t, p, r \in T, p \leq r \leq t \tag{A7}$$

$$LF \cdot y_{1,t,p} \leq 0.5 \ \ \forall t, p \in T \quad \text{(A8)}$$

$$WellDeviation \cdot y_{1,t,p} \leq 70 \ \ \forall t, p \in T \quad \text{(A9)}$$

$$DLS \cdot y_{1,t,p} \leq 6 \ \ \forall t, p \in T \quad \text{(A10)}$$

$$PTD \cdot y_{1,t,p} \leq 3.5 \ \ \forall t, p \in T \quad \text{(A11)}$$

$$TubingUniformity \geq \ y_{1,t,p} \ \ \forall t, p \in T \quad \text{(A12)}$$

$$y_{1,t,p} \leq 1 - TAC \quad \forall t, p \in T \quad \text{(A13)}$$

$$y_{1,t,p} \leq 1 - MCF \quad \forall t, p \in T \quad \text{(A14)}$$

$$WellboreAccess \geq y_{1,t,p} \quad \forall t, p \in T \quad \text{(A15)}$$

$$WellDepth \cdot y_{1,t,p} \leq 19000 \quad \forall t, p \in T \quad \text{(A16)}$$

$$Temp \cdot y_{1,t,p} \leq 110 \quad \forall t, p \in T \quad \text{(A17)}$$

$$y_{1,t,p} \leq 1 - Offshore \quad \forall t, p \in T \quad \text{(A18)}$$

Foam Lift, $i = 2$

$$WaterCut \geq 0.75 \cdot y_{2,t,p} \ \ \forall t, p \in T \quad \text{(A19)}$$

IF Packer = 1:

$$CITHP \geq \ 1.2 \cdot FLP \ y_{2,t,p} \ \ \forall t, p \in T \quad \text{(A20)}$$

ELSE Packer =0:

$$CITHP \geq 2.5 \cdot FLP\ y_{2,t,p} \quad \forall t,p \in T \tag{A21}$$

$$Qg_r \geq 430 \cdot LFR_r \cdot y_{2,t,p} \quad \forall t,p,r \in T, p \leq r \leq t \tag{A22}$$

$$TDS \cdot y_{2,t,p} \leq 0.05 \quad \forall t,p \in T \tag{A23}$$

$$LFR_r \cdot y_{2,t,p} \leq 500 \quad \forall t,p,r \in T, p \leq r \leq t \tag{A24}$$

$$WellDepth \cdot y_{2,t,p} \leq 22000 \quad \forall t,p \in T \tag{A25}$$

$$Temp \cdot y_{2,t,p} \leq 400 \quad \forall t,p \in T \tag{A26}$$

Well Head Compression, *i* =3

$$Qg_r \cdot y_{3,t,p} \leq 0.3 \quad \forall t,p,r \in T, p \leq r \leq t \tag{A27}$$

$$LFR_r \cdot y_{3,t,p} \leq 100 \quad \forall t,p,r \in T, p \leq r \leq t \tag{A28}$$

$$FTHP \cdot y_{3,t,p} \leq 175 \quad \forall t,p \in T \tag{A29}$$

$$FTHP \geq 30 \cdot y_{3,t,p} \quad \forall t,p \in T \tag{A30}$$

$$CITHP \cdot y_{3,t,p} \leq 1100 \quad \forall t,p \in T \tag{A31}$$

Velocity Strings, *i* = 4

$$LFR_r \cdot y_{4,t,p} \leq 300 \quad \forall t,p \in T \tag{A32}$$

$$BHP \geq 500 \cdot y_{4,t,p} \quad \forall t,p \in T \tag{A33}$$

Gas Lift, *i* =5

$$LFR_r \geq 150 \cdot y_{5,t,p} \quad \forall t,p,r \in T, p \leq r \leq t \tag{A34}$$

$$CGA \geq y_{5,t,p} \quad \forall t,p \in T \tag{A35}$$

$$y_{5,t,p} \leq 1 - MC \quad \forall t,p \in T \tag{A36}$$

$$WellDepth \cdot y_{5,t,p} \leq 18000 \quad \forall t,p \in T \tag{A37}$$

$$Temp \cdot y_{5,t,p} \leq 450 \quad \forall t,p \in T \tag{A38}$$

Electrical Submersible Pump, *i* = 6

$$y_{6,t,p} \cdot GVF \leq 0.3 \quad \forall t,p \in T \tag{A39}$$

$$WellDepthy_{6,t,p} \leq 15000 \quad \forall t,p \in T \tag{A40}$$

$$Temp \cdot y_{6,t,p} \leq 400 \quad \forall t,p \in T \tag{A41}$$

$$LFR_r \geq 150 \cdot y_{6,t,p} \quad \forall t,p,r \in T, p \leq r \leq t \tag{A42}$$

$$LFR_r \cdot y_{6,t,p} \leq 30000 \quad \forall t,p,r \in T, p \leq r \leq t \tag{A43}$$

$$CasingDiameter \geq 5.5 \cdot y_{6,t,p} \quad \forall t,p \in T \tag{A44}$$

$$DLS \cdot y_{6,t,p} \leq 6 \quad \forall t,p \in T \tag{A45}$$

$$EPA + EPEF \geq y_{6,t,p} \quad \forall t,p \in T \tag{A46}$$

Sucker Rod Pump, *i* =7

IF DownholeSeparator = 0

$$y_{7,t,p} \cdot GVF \leq 0.4 \quad \forall t,p \in T \tag{A47}$$

$$WellDepth \cdot y_{7,t,p} \leq 16000 \quad \forall t,p \in T \tag{A48}$$

$$Temp \cdot y_{7,t,p} \leq 550 \quad \forall t,p \in T \tag{A49}$$

$$LFR_r \cdot y_{7,t,p} \leq 6000 \quad \forall t,p,r \in T, p \leq r \leq t \tag{A50}$$

$$LFR_r \geq 10 \cdot y_{7,t,p} \quad \forall t,p,r \in T, p \leq r \leq t \tag{A51}$$

$$PTD \geq 2.375 \cdot y_{7,t,p} \quad \forall t,p \in T \tag{A52}$$

$$WellDeviation \cdot y_{7,t,p} \leq 80 \quad \forall t,p \in T \tag{A53}$$

$$DLS \cdot y_{7,t,p} \leq 15 \quad \forall t,p \in T \tag{A54}$$

$$y_{7,t,p} \leq 1 - Offshore \quad \forall t,p \in T \tag{A55}$$

Progressing Cavity Pump, $i = 8$

IF DownholeSeparator = 0:

$$y_{8,t,p} \cdot GVF \leq 0.5 \quad \forall t,p \in T \tag{A56}$$

IF WellDepth <= 6000:

$$LFR_r \geq 10 \cdot y_{8,t,p} \quad \forall t,p,r \in T, p \leq r \leq t \tag{A57}$$

IF WellDepth >= 6000:

$$LFR_r \geq 60 \cdot y_{8,t,p} \quad \forall t,p,r \in T, p \leq r \leq t \tag{A58}$$

$$WellDepth \cdot y_{8,t,p} \leq 8600 \quad \forall t,p \in T \tag{A59}$$

$$LFR_r \cdot y_{8,t,p} \leq 4500 \quad \forall t,p,r \in T, p \leq r \leq t \tag{A60}$$

$$CasingDiameter \geq 4.5 \cdot y_{8,t,p} \quad \forall t,p \in T \tag{A61}$$

$$Temp \cdot y_{8,t,p} \leq 250 \quad \forall t,p \in T \tag{A62}$$

$$AromaticContent \cdot \ y_{8,t,p} \leq 0.03 \quad \forall t,p \in T \tag{A63}$$

$$DLS \cdot y_{8,t,p} \leq 15 \quad \forall t,p \in T \tag{A64}$$

$$y_{8,t,p} \leq 1 - Offshore \quad \forall t,p \in T \tag{A65}$$

$$GPA + GPEF \geq y_{8,t,p} \quad \forall t,p \in T \tag{A66}$$

$$EPA + EPEF \geq y_{8,t,p} \quad \forall t,p \in T \tag{A67}$$

Jet Pump, $i = 9$

$$DLS \cdot y_{9,t,p} \leq 24 \quad \forall t,p \in T \tag{A68}$$

$$WellDepth \cdot y_{9,t,p} \leq 15000 \quad \forall t,p \in T \tag{A69}$$

$$LFR_r \geq 200 \cdot y_{9,t,p} \quad \forall t,p,r \in T, p \leq r \leq t \tag{A70}$$

$$LFR_r \cdot y_{9,t,p} \leq 35000 \quad \forall t,p,r \in T, p \leq r \leq t \tag{A71}$$

$$y_{9,t,p} \cdot GVF \leq 0.5 \quad \forall t,p \in T, \tag{A72}$$

$$TubingSize \geq 2.875 \cdot \ y_{9,t,p} \quad \forall t,p \in T \tag{A73}$$

$$y_{9,t,p} \leq 1 - MCF \quad \forall t, p \in T \tag{A74}$$

$$y_{9,t,p} \leq 1 - TAC \quad \forall t, p \in T \tag{A75}$$

$$WellboreAccess \geq y_{9,t,p} \quad \forall t, p \in T \tag{A76}$$

$$Temp \cdot y_{9,t,p} \leq 550 \quad \forall t, p \in T \tag{A77}$$

$$y_{9,t,p} \leq 1 - Offshore \quad \forall t, p \in T \tag{A78}$$

$$EPA + EPEF \geq y_{9,t,p} \quad \forall t, p \in T \tag{A79}$$

APPENDIX B1: NOMENCLATURE FOR THE CLINICAL TRIAL PLANNING MSSP MODEL

Indices/Sets

$i \in I$	set of drugs
$j \in J$	set of clinical trials; J = {PI, PII, PIII}
$r \in R$	set of resource type
$s,s' \in S$	set of scenarios
$p \in \{1,2,...T\}$	set of time periods
$t \in \{1,2,...T+max_{ij}(\tau_{ij})\}$	set of time periods

Indices/Sets

B pairs of scenarios that differ in the outcome of one clinical trials *(i, j)*; non-anticipativity constraints are expressed only for (s,s') ϵ *B*.

Parameters

C_{ij}	cost of trial time periods
Cd_t	discounting factor for time value of money
f_{ij}	discounting factor for open revenue
P_{ij}	probability of trial *(i, j)* being successful
p_s	probability of scenario *s*
rev_i^{max}	maximum possible revenue for drug *i*
rev_{ij}^{open}	estimated revenue realized for drug i, if trial *(i, j-1)* completed while trial *(i, j)* not started, is successfully developed beyond the end of the time horizon

γ_i^D	loss coefficient – late completion
γ_i^L	loss coefficient – loss in active patent life
$\rho_{i,j,r}$	resources of type r required to start trial *(i, j)*
$\tau_{i,j}$	duration of trial *(i, j)*
ρ_r^{max}	maximum resources available of type *r*

Variables

Cst_s	total development cost in scenario s
Rv_s	revenue of scenario s
FRv_s	free revenue of scenario s
ENPV	expected net present value
$X_{i,j,p,s} \in \{0, 1\}$	1 if clinical trial *(i, j)* starts at the time p for scenario s

APPENDIX B2: MSSP FORMULATION OF COLVIN AND MARAVELIAS (2008)

$$\mathrm{ENPV} = \sum_{s} p_s(\mathrm{Rv}_s + \mathrm{FRev}_s - \mathrm{Cst}_s) \tag{B1}$$

$$V_{i,j,t,s} = V_{i,j,t-1,s} + X_{i,j,t-\tau_{i,j},s} \qquad \forall i, j, t, s \tag{B2}$$

$$Z_{i,1,1,s} = 1 - X_{i,1,t,s} \qquad \forall i, s \tag{B3}$$

$$Z_{i,1,t,s} = Z_{i,1,t-1,s} - X_{i,j,t,s} \qquad \forall i, t > 1, s \tag{B4}$$

$$Z_{i,j,t,s} = Z_{i,j,t-1,s} + X_{i,j-1,t-\tau_{i,j-1},s} - X_{i,j,t,s} \; \forall i, j > 1, t, s \tag{B5}$$

$$\sum_{t} X_{i,j,t,s} \leq 1 \qquad \forall i, j, s \tag{B6}$$

$$\sum_{t' \leq t} X_{i,j,t',s} \leq V_{i,j-1,t,s} \quad \forall i, j > 1, t, s \tag{B7}$$

$$\sum_{i} \sum_{j} \sum_{t' > t-\tau_{i,j}}^{t' \leq t} \rho_{i,j,r} X_{i,j,t',s} \leq {\rho_r}^{\max} \forall r, t, s \tag{B8}$$

$$X_{i,1,1,s} = X_{i,1,1,1} \quad \forall i, s \tag{B9}$$

$$-V_{i^{s,s'},j^{s,s'},t,s} \leq X_{i,j,t,s} - X_{i,j,t,s'} \leq V_{{i^{s,s'}}_{,j^{s,s'}},t,s} \quad \forall i, j, (s, s') \in \Psi, t > 1 \tag{B10}$$

$$\mathrm{Cst}_s = \sum_{i,j,t} \mathrm{cd}_t C_{i,j} X_{i,j,t,s} \; \forall s \tag{B11}$$

$$\mathrm{Rv}_s = \sum_{i} \sum_{t} \mathrm{rev}_i^{\max} X_{i,\mathrm{PII},t,s} - {\gamma_i}^{D}\left(Z_{i,\mathrm{PII},t,s} + Z_{i,\mathrm{PIII},t,s}\right) - {\gamma_i}^{L}(t + \tau_{i,\mathrm{PIII}}) X_{i,\mathrm{PIII},t,s} \; \forall s \tag{B12}$$

$$FRev_s = \sum_i \sum_j rev_{i,j}^{open} f_{i,j} Z_{i,j,|T|,s} + \sum_i \sum_{j \in \{PI,PII\}} \sum_{t > |T| - \tau_{i,j}} rev_{i,j,t}^{run} f_{i,j+1} X_{i,j,t,s} \quad \forall s \tag{B13}$$

$$rev_{i,j}{}^{open} = rev_i{}^{max} - \gamma_i{}^{L} \left(|T| + \sum_{j' \geq j} \tau_{i,j'} \right) \tag{B14}$$

$$rev_{i,j}{}^{run} = rev_i{}^{max} - \gamma_i{}^{L} \left(t + \sum_{j' \geq j} \tau_{i,j'} \right) \tag{B15}$$

$$f_{i,j} = 0.9 \left[\frac{rev_i{}^{max} - \gamma_i{}^{L}|T| - \sum_{j' \geq j} C_{i,j}}{rev_i{}^{max} - \gamma_i{}^{L}|T|} \right] \tag{B16}$$

$X_{i,j,t,s} \in \{0,1\}$ 1 if trial (i,j) starts at the beginning of time t for scenario s (B17)

$V_{i,j,t,s} \in [0,1]$ 1 if trial (i,j) is completed at the beginning of time t for scenario s (B18)

$Z_{i,j,t,s} \in [0,1]$ 1 if trial (i,j) can be started at the beginning of time t for scenario s (B19)

APPENDIX C: THE PARAMETERS OF FIVE BASE CASE

Table C1. Parameters of the two-drug case study.

Drug	Duration		Probability of Success		Cost ($M)		Resource 1 (Max = 2)		Resource 2 (Max = 3)		rev^{max}	γ^L	γ^D
	PI	PII	PI	PII	PI	PII	PI	PII	PI	PII			
D1	2	4	0.3	0.5	10	90	1	1	1	2	3100	19.2	44
D2	2	3	0.4	0.6	10	80	1	2	1	1	3250	19.6	56

Clinical trial plan for a 15-month planning horizon divided into five equal time periods.

Table C2. Parameters of the three-drug case study.

Drug	Duration			Probability of Success			Trial Cost ($M)			Resource 1 (Max = 2)			Resource 1 (Max = 3)			rev^{max}	γ^L	γ^D
	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII			
D1	2	4	4	0.3	0.5	0.8	10	90	220	1	1	2	1	2	3	3100	19.2	44
D2	2	3	5	0.4	0.6	0.8	10	80	200	1	2	2	1	1	3	3250	19.6	56
D3	2	3	4	0.3	0.6	0.9	10	90	180	1	1	2	1	1	3	3300	20	52

Clinical Trial plan for a 36-month planning horizon divided into 12 equal time periods.

Table C3. Parameters of the four-drug case study.

Drug	Duration			Probability of Success			Trial Cost ($M)			Resource 1 (Max = 4)			Resource 1 (Max = 3)			rev^{max}	γ^L	γ^D
	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII			
D1	1	1	3	0.3	0.5	0.8	10	90	220	1	1	2	1	2	3	3100	19.2	22
D2	1	2	2	0.4	0.6	0.8	10	80	200	1	2	2	1	1	3	3250	19.6	28
D3	1	1	3	0.3	0.6	0.9	10	90	180	1	1	2	1	1	3	3300	20	26
D4	1	2	2	0.4	0.6	0.8	10	100	170	1	1	2	1	2	3	3000	19.4	24

Clinical trial plan for an 18-month planning horizon divided into six equal time periods.

Table C4. Parameters of the five-drug case study.

Drug	Duration			Probability of Success			Trial Cost ($M)			Resource 1 (Max = 4)			Resource 1 (Max = 3)			rev^{max}	γ^L	γ^D
	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII			
D1	1	1	3	0.3	0.5	0.8	10	90	220	1	1	2	1	2	3	3100	19.2	22
D2	1	2	2	0.4	0.6	0.8	10	80	200	1	2	2	1	1	3	3250	19.6	28
D3	1	1	3	0.3	0.6	0.9	10	90	180	1	1	2	1	1	3	3300	20	26
D4	1	2	2	0.4	0.6	0.8	10	100	170	1	1	2	1	2	3	3000	19.4	24
D5	1	2	3	0.35	0.5	0.9	10	70	210	1	1	2	1	1	3	3150	19.6	24

Clinical trial plan for an 18-month planning horizon divided into six equal time periods.

Table C5. Parameters for a six-drug case study.

Drug	Duration			Probability of Success			Trial Cost ($M)			Resource 1 (Max = 4)			Resource 1 (Max = 3)			rev^{max}	γ^L	γ^D
	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII	PI	PII	PIII			
D1	1	1	3	0.3	0.5	0.8	10	90	220	1	1	2	1	2	3	3100	19.2	22
D2	1	2	2	0.4	0.6	0.8	10	80	200	1	2	2	1	1	3	3250	19.6	28
D3	1	1	3	0.3	0.6	0.9	10	90	180	1	1	2	1	1	3	3300	20	26
D4	1	2	2	0.4	0.6	0.8	10	100	170	1	1	2	1	2	3	3000	19.4	24
D5	1	2	3	0.35	0.5	0.9	10	70	210	1	1	2	1	1	3	3150	19.6	24
D6	1	2	3	0.45	0.45	0.8	10	85	195	1	2	2	2	1	3	3050	19	25

Clinical trial plan for an 18-month planning horizon divided into six equal time periods.

Table C6. Parameter perturbation results for CM1, CM2 and CM3.

Numbers	Case Name	ENPV	Solving Times (s)		
			CM1	CM2	CM3
1	5-Drug-GD75%	1360.78	332	426	4972
2	5-Drug-GD90%	2127.53	536	551	3870
3	5-Drug-GD110%	2123.03	477	638	5759
4	5-Drug-GD125%	2127.56	489	532	5656
5	5-Drug-GL75%	2083.53	260	277	5650
6	5-Drug-GL90%	2083.4	442	440	4380
7	5-Drug-GL110%	2081.13	396	418	4393
8	5-Drug-GL125%	2080.27	476	452	6214
9	5-Drug-RV75%	2111.88	404	412	6021
10	5-Drug-RV90%	2093.38	544	554	5912
11	5-Drug-RV110%	2070.03	427	573	4137
12	5-Drug-RV125%	2051.49	477	479	4249
13	5-Drug-TC75%	1443.15	347	390	3556
14	5-Drug-TC90%	1826.37	175	214	3511
15	5-Drug-TC110%	2337.79	485	500	6824
16	5-Drug-TC125%	2721.78	671	602	4451
17	5-Drug-TL1	2003	38	45	84
18	5-Drug-TL2	1862	9	14	23

APPENDIX D1: NOMENCLATURE OF SYNTHESIS OF PROCESS NETWORKS

Sets

$(s, s') \in D$	pairs of scenarios that differ in the outcome of yields of processes
$i \in I$	set of processes in the process network
k $\in K$	set of streams in the process network
$s, s' \in S$	set of scenarios
$t \in T$	set of periods in time horizon

Binary variables (or equivalent Boolean variables)

$Z_t^{s,s'}$	whether or not scenarios s, s' are indistinguishable at time period t
$y_{i,t}^{exp,s}$	whether or not process i is expanded in time period t in scenario s
$y_{i,t}^{oper,s}$	whether or not process i is operated in time period t in scenario s

Continuous variables

$NPVs$	net present value of project under scenario s
$x_s^{purch,s}$	amount of purchases of final product at time period t in scenario s
$x_s^{sales,s}$	amount of sales of final product at time period t in scenario s
$w_{i,t}^{cap,s}$	capacity of process i at time period t in scenario s
$w_{i,t}^{QE,s}$	amount of capacity expansion of process i at time period t in scenario s
$w_{k,t}^{rate,s}$	flowrate of stream k at time period t in scenario s

Parameters

d_t	demand for final product at time period t

$FE_{i,t}$	fixed expansion cost for process i at time period t
$FO_{i,t}$	fixed operating cost for process i at time period t
$VO_{k,t}$	variable operating cost for stream k at time period t
$VE_{i,t}$	variable expansion cost for process i at time period t
α_t	purchase price for final product at time period t
β_t	sales price for final product at time period t
γ_t	maintaining inventory cost at time period t
θ_i^s	yield of process i in scenario s

APPENDIX D2: An MSSP Formulation of Synthesis of Process Networks

$$\max ENPV = \sum_{s} p^{s} NPV^{s} \tag{D1}$$

$$\text{s.t. } NPV^{s} = -\sum_{t}\sum_{i}\left(FE_i y_{i,t}^{exp,s} + FO_i y_{i,t}^{oper,s} + VE_i w_{i,t}^{QE,s}\right) - \sum_{t}\sum_{k} VO_k w_{k,t}^{rate,s} - \sum_{t}\left(\alpha_t x_t^{purch,s} - \beta_t x_t^{sales,s} - \gamma_t w_t^{inv,s}\right) \tag{D2}$$

$$w_{3,t}^{rate,s} = \theta_1^s w_{1,t}^{rate,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D3}$$

$$w_{4,t}^{rate,s} = \theta_2^s w_{2,t}^{rate,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D4}$$

$$w_{8,t}^{rate,s} = \theta_3^s w_{7,t}^{rate,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D5}$$

$$w_{5,t}^{rate,s} = w_{3,t}^{rate,s} + w_{4,t}^{rate,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D6}$$

$$w_{7,t}^{rate,s} = w_{5,t}^{rate,s} + w_{6,t}^{rate,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D7}$$

$$w_t^{inv,s} = w_{t-1}^{inv,s} + (w_{8,t}^{rate,s} + x_t^{purch,s} - x_t^{sales,s}) \qquad \forall s \epsilon S, t \epsilon T \tag{D8}$$

$$x_t^{sales,s} = d_t \qquad \forall s \epsilon S, t \epsilon T \tag{D9}$$

$$w_{3,t}^{rate,s} \leq w_{1,t}^{cap,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D10}$$

$$w_{4,t}^{rate,s} \leq w_{2,t}^{cap,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D11}$$

$$w_{8,t}^{rate,s} \leq w_{3,t}^{cap,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D12}$$

$$w_{i,t}^{cap,s} = w_{i,t-1}^{cap,s} + w_{i,t}^{QE,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D13}$$

$$L_i^{QE} y_{i,t}^{exp,s} \leq \quad {}_{i,t}^{QE,s} \leq U_i^{QE} y_{i,t}^{exp,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D14}$$

$$L_1^{inflow} y_{1,t}^{oper,s} \leq w_{3,t}^{rate,s} \leq U_1^{inflow} y_{1,t}^{oper,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D15}$$

$$L_2^{inflow} y_{2,t}^{oper,s} \leq w_{4,t}^{rate,s} \leq U_2^{inflow} y_{2,t}^{oper,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D16}$$

$$L_3^{inflo} y_{3,t}^{oper,s} \leq w_{8,t}^{rate,s} \leq U_3^{inflow} y_{3,t}^{oper,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D17}$$

$$\sum_{\tau}^{t} y_{i,\tau}^{exp,s} \geq y_{i,t}^{oper,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D18}$$

$$y_{i,t}^{oper,s} \geq y_{i,t}^{exp,s} \qquad \forall s \epsilon S, t \epsilon T \tag{D19}$$

$$y_{i,1}^{oper,s} = y_{i,1}^{oper,s'} \qquad \forall i \in I, s \epsilon S, t \epsilon T \tag{D20}$$

$$y_{i,1}^{exp,s} = y_{i,1}^{exp,s'} \qquad \forall i \in I, s \epsilon S, t \epsilon T \tag{D21}$$

$$w_{i,1}^{QE,s} = w_{i,1}^{QE,s'} \qquad \forall i \in I, s \epsilon S, t \epsilon T \tag{D22}$$

$$w_{k,1}^{rate,s} = w_{k,1}^{rate,s'} \qquad \forall k \in K, s \epsilon S, t \epsilon T \tag{D23}$$

$$Z_t^{s,s'} \Leftrightarrow \left[\bigwedge_{\tau=1}^{t} \left(\neg y_{i,\tau}^{oper,s} \right) \right] \qquad \forall (s, s') \in S, \{i\} = D(s, s'), t \epsilon T \tag{D24}$$

$$\begin{bmatrix} Z_t^{s,s'} \\ x_t^{purch,s} = x_t^{purch,s'} \\ x_t^{sales,s} = x_t^{sales,s'} \\ y_{i,t+1}^{oer,s} = y_{i,t+1}^{oper,s'} \quad t \le T-1 \\ y_{i,t+1}^{exp,s} = y_{i,t+1}^{exp,s'} \quad t \le T-1 \\ w_{i,t+1}^{QE,s} = w_{k,t+1}^{QE,s'} \quad t \le T-1 \\ w_{k,t+1}^{rate,s} = w_{k,t+1}^{rate,s'} \quad t \le T-1 \end{bmatrix} \vee \left[\neg Z_t^{s,s'}\right] \quad \forall (s,s') \in D, t \epsilon T \tag{D25}$$

$$y_{i,t}^{exp,s} \in \{0,1\} \quad \forall s \in S, t \epsilon T, i = 1,2,3 \tag{D26}$$

$$y_{i,t}^{ope,s} \in \{0,1\} \quad \forall s \in S, t \epsilon T, i = 1,2,3 \tag{D27}$$

APPENDIX E1: NOMENCLATURE FOR THE NEW TECHNOLOGY INVESTMENT PLANNING MSSP MODEL

Indicies/Sets

$i \in \boldsymbol{I}$	Technologies
$sg \in \boldsymbol{SG}$	Technology stages
$n \in \boldsymbol{N}$	Chemicals
$t \in \boldsymbol{T}$	Time Periods
$s, s' \in \boldsymbol{S}$	Scenarios
$np \in \boldsymbol{NP}$	partitions

Subsets

$\boldsymbol{\phi}_k$	Pairs of scenarios s and s' which differ in the realized value of the uncertain parameter k

Parameter

$\gamma_{i,n,n'}$	Stoichiometric ratio between component n and n' in technology i
p_s	The probability that scenario s will occur
$MCst_{n,t}$	The per tonne purchase cost of chemical n at time t
INV^{Max}	The maximum invest available for capacity expansion and research at each time period
$CC0_i$	The initial capacity expansion cost for technology i
β_i^0	The assumed value for the parameter β before investment
$\beta_{i,s}^1$	The realized value of β for technology i in scenario s
α_i^0	The assumed value for the parameter α before investment
$\alpha_{i,s}^1$	The realized value of α for technology i in scenario s
$D_{n,t,s}$	The demand of chemical n at time s in scenario s
$CX_{i,sg}^{Min}$	The minimum capacity of stage sg for technology i
$CX_{i,sg}^{Max}$	The maximum capacity of stage sg for technology i
$t_{s,s'}^{diff}$	The time of differentiation between scenarios s and s'
V^{α}	The number of investments in research required to realize the true value of the parameter α

V^{β}	The number of investments in capacity expansion required to realize the true value of β
Decision Variables	
$CX_{i,t,s}$	The installed capacity of technology i at time t in scenario s
$RD_{i,t,s}$	The research investment in technology i at time t in scenario s
$F_{n,t,s}$	The amount of chemical n purchased at time t in scenario s
$M_{i,n,t,s}$	The amount of chemical n produced in technology i at time t in scenario s
Variables	
$MatCst_s$	The total material cost for scenario s
$CapExCst_s$	The total amount spent on capacity expansion in scenario s
$RDCst_s$	The total amount spent on research in scenario s
ETC	The expected total cost
$CC_{i,t,s}$	The capacity expansion cost of technology i at time t in scenario s
$X_{i,t,s}$	The capacity expansion of technology i at time t in scenario s
$\beta_{i,t,s}$	The elasticity parameter for capacity expansion for technology i at time t in scenario s
$\alpha_{i,t,s}$	The elasticity parameter for research expansion for technology i at time t in scenario s
$G_{n,t,s}$	The net production of chemical n at time t in scenario s
$\chi_{i,t,s}$	The yield of technology i at time t in scenario s
Binary Variables	
$NN^{\beta}_{i,t,s}$	Takes a value of 1 if there have been V^{β} investments in capacity expansion for technology i at time t in scenario s, and 0 otherwise
$NN^{\alpha}_{i,t,s}$	Takes a value of 1 if there have been V^{α} investments in research expansion for technology i at time t in scenario s, and 0 otherwise
$N^{\beta}_{i,t,s}$	Takes a value of 1 if there was an investment in capacity expansion in technology i at time t in scenario s and 0 otherwise
$N^{\alpha}_{i,t,s}$	Takes a value of 1 if there was an investment in research expenditures for technology i at time t for scenario s and 0 otherwise
$Z_{i,sg,t,s}$	Takes a value of 1 if technology i is currently in stage sg at time t in scenario s and 0 otherwise

$Y_{i,sg,t,s}$	Takes a value of 1 if technology i has completed stage sg at time t in scenario s and 0 otherwise

APPENDIX E2: An MSSP Formulation of Synthesis of Process Networks

$$\max: TC = \sum_s p_s(MatCst_s + CapExCst \quad + RDCst_s) \tag{A1}$$

s.t.

$$CapExCst_s = \sum_i \sum_t CC_{i,t,s} X_{i,t,s} \tag{A2}$$

$$MatCst_s = \sum_n \sum_t MCst_{n,t} F_{n,t,s} \tag{A3}$$

$$RDCst_s = \sum_i \sum_t RD_{i,t,s} \tag{A4}$$

$$CC_{i,t,s} = CC_{i,0,s} \left(\frac{CX_{i,t,s}\, MCS_{i,0}}{CX_{i,0} MCS_{i,t}} \right)^{\beta_{i,t,s}} \left(\frac{RD_{i,t,s}\, MRDS_{i,0}}{RD_{i,0} MRDS_{i,t}} \right)^{\alpha_{i,t,s}} \tag{A5}$$

$$0 = F_{n,t,s} - D_{n,t,s} + P_{n,t,s} \tag{A6}$$

$$P_{n,t,s} = \sum_i \sum_{n'} \gamma_{i,n,n'} \chi_{i,s} M_{i,t,s} \tag{A7}$$

$$P_{n,t,s} = \sum_i \gamma_{i,n} \chi_{i,s} M_{i,t,s} \tag{A8}$$

$$CX_{i,t,s} \geq {CX_{i,sg}}^{min} Y_{i,sg,t,s} \tag{A9}$$

$$CX_{i,t,s} \leq \sum_{sg>1} \left({CX_{i,sg}}^{max} - {CX_{i,sg-1}}^{max} \right) Y_{i,sg,t,s} \tag{A10}$$

$$Y_{i,sg,t,s} \leq Y_{i,sg-1,t,s} \tag{A11}$$

$$Y_{i,sg,t,s} \leq Y_{i,sg,t-1,s} \tag{A12}$$

$$M_{i,t,s} \leq Y_{i,3,t,s} CX_{i,t,s} \tag{A13}$$

$$Z_{i,sg,t,s} = Y_{i,sg,t,s} - Y_{i,sg+1,t,s} \tag{A14}$$

$$\chi_{i,t,s} = \sum_{sg} Z_{i,sg,t,s} \chi_{i,sg,s}{}^{0} \tag{A15}$$

$$CX_{i,t,s} - CX_{i,t,s} \leq N_{i,t,s}{}^{\beta} X_i{}^{max} \tag{A16}$$

$$\sum_{t'<t} N_{i,t',s}{}^{\beta} \leq NN_{i,t,s}{}^{\beta} \tag{A17}$$

$$RD_{i,t,s} - RD_{i,t,s} \leq N_{i,t,s}{}^{\alpha} RD^{max} \tag{A18}$$

$$\sum_{t'<t} N_{i,t',s}{}^{\alpha} \leq NN_{i,t,s}{}^{\alpha} \tag{A19}$$

$$\alpha_{i,t,s} = \alpha_i{}^{0}\left(1 - NN_{i,t,s}{}^{\alpha}\right) + \alpha_{i,s}{}^{1} NN_{i,t,s}{}^{\alpha} \tag{A20}$$

$$\alpha_{i,t,s} = \sum_{k} \alpha_i^k NN_{i,t,s}^{\alpha,k} \tag{A21}$$

$$\beta_{i,t,s} = \beta_i{}^{0}\left(1 - NN_{i,t,s}{}^{\beta}\right) + \beta_{i,s}{}^{1} NN_{i,t,s}{}^{\beta} \tag{A22}$$

$$\beta_{i,t,s} = \sum_{k} \beta_i^k NN_{i,t,s}^{\beta,k} \tag{A23}$$

$$-CX_{i,3}^{max} Z_{iss',sgss',t,s} \leq CX_{i,t,s} - CX_{i,t,s'} \leq Z_{iss',sgss',t,s} CX_{i,3}^{max} \tag{A24}$$

$$-CX_{i,3}^{max} NN_{iss',t,s}{}^{\beta} \leq CX_{i,t,s} - CX_{i,t,s'} \leq NN_{iss',t,s}{}^{\beta} CX_{i,3}^{max} \tag{A25}$$

$$-CX_{i,3}^{ma} NN_{iss',t,s}{}^{\alpha} \leq CX_{i,t,s} - CX_{i,t,s'} \leq NN_{iss',t,s}{}^{\alpha} CX_{i,3}^{max} \tag{A26}$$

$$-RD^{max} Z_{iss',sgss',t,s} \leq RD_{i,t,s} - RD_{i,t,s'} \leq Z_{iss',sgss',t,s} RD^{max} \quad \text{(A27)}$$

$$-RD^{max} {NN_{iss',t,s}}^{\beta} \leq RD_{i,t,s} - RD_{i,t,s'} \leq {NN_{iss',t,s}}^{\beta} RD^{max} \quad \text{(A28)}$$

$$-RD^{max} {NN_{iss',t,s}}^{\alpha} \leq RD_{i,t,s} - RD_{i,t,s'} \leq {NN_{iss',t,s}}^{\alpha} RD^{max} \quad \text{(A29)}$$

$$-CX_{i,sg}^{max} Z_{iss',sgss',t,s} \leq M_{i,n,t,s} - M_{i,n,t,s'} \leq Z_{iss',sgss',t,s} CX_{i,sg}^{max} \quad \text{(A30)}$$

$$-CX_{i,sg}^{max} {NN_{iss',t,s}}^{\beta} \leq M_{i,n,t,s} - M_{i,n,t,s'} \leq {NN_{iss',t,s}}^{\beta} CX_{i,sg}^{max} \quad \text{(A31)}$$

$$-CX_{i,sg}^{max} {NN_{iss',t,s}}^{\alpha} \leq M_{i,n,t,s} - M_{i,n,t,s'} \leq {NN_{iss',t,s}}^{\alpha} CX_{i,sg}^{max} \quad \text{(A32)}$$

$$-Pd_i^{max} Z_{iss',sgss',t,s} \leq F_{i,t,s} - F_{i,t,s'} \leq Z_{iss',sgss',t,s} Pd_i^{max} \quad \text{(A33)}$$

$$-Pd_i^{max} {NN_{iss',t,s}}^{\beta} \leq F_{i,t,s} - F_{i,t,s'} \leq {NN_{iss',t,s}}^{\beta} Pd_i^{max} \quad \text{(A34)}$$

$$-Pd_i^{max} {NN_{iss',t,s}}^{\alpha} \leq F_{i,t,s} - F_{i,t,s'} \leq {NN_{iss',t,s}}^{\alpha} Pd_i^{max} \quad \text{(A35)}$$

$$CX_{i,t,s} = CX_{i,t,s'} \qquad \forall i, t < t^{diff}, (s, s') \in \boldsymbol{\phi} \quad \text{(A36)}$$

$$RD_{i,t,s} = RD_{i,t,s'} \qquad \forall i, t < t^{diff}, (s, s') \in \boldsymbol{\phi} \quad \text{(A37)}$$

$$M_{i,n,t,s} = M_{i,n,t,s'} \qquad \forall i, n, t < t^{diff}, (s, s') \in \boldsymbol{\phi} \quad \text{(A38)}$$

$$F_{i,t,s} = F_{i,t,s'} \qquad \forall i, t < t^{diff}, (s, s') \in \boldsymbol{\phi} \quad \text{(A39)}$$