A Stochastic Programming Approach to Planning of Offshore Gas Field Developments under Uncertainty in Reserves

Vikas Goel*and Ignacio E. Grossmann†
Department of Chemical Engineering, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Abstract

In this work, we consider the investment and operational planning of gas field developments under uncertainty in gas reserves. The resolution of uncertainty in gas reserves, and hence the shape of the scenario tree associated with the problem depends on the investment decisions. A novel stochastic programming model that incorporates the decision-dependence of the scenario tree is presented. A decomposition based approximation algorithm for the solution of this model is also proposed. We show that the proposed approach yields solutions with significantly higher Expected Net Present Value (ENPV) than that of solutions obtained using a deterministic approach. For a small sized example, the proposed approximation algorithm is shown to yield the optimal solution with more than one order of magnitude reduction in solution time, as compared to the full space method. "Good" solutions to larger problems, that require up to 165,000 binary variables in full space, are obtained in a few hours using the proposed approach.

1 Introduction

Oil and gas exploration and production is a highly capital-intensive industry. In the year 2001, the total earnings from oil and gas exploration and production alone, for one of the leading

*E-mail: vgoel@andrew.cmu.edu
†To whom all correspondence should be addressed. Tel.: (412) 268-2230. Fax: (412) 268-7139. E-mail: grossmann@cmu.edu.

petroleum companies, were in excess of $10.4 Billion (ExxonMobil, 2001). To maintain production levels that led to such high earnings, the company invested about $7.9 Billion in property acquisition, exploration and development during the year.

Facilities required for offshore exploration and production often remain in operation over the entire life-span of the project, typically 10-30 years. Therefore, decisions regarding investment in these facilities affect the profitability of the entire project. Given the large potential profits and high investments in each project, it is extremely important to follow an investment policy that is optimal with respect to the entire project life-span. This need has created significant interest in developing optimization models for planning in the oil and gas exploration and production industry.

Decision-makers in this industry, however, have to contend with a great deal of uncertainty. The most important sources of uncertainty associated with planning of oil or gas exploration and production are the future prices of oil and gas, and the quality of reserves. Oil and gas prices across the world are affected by a variety of reasons, from political to climatic, and it is obviously impossible to predict these with reasonable accuracy for horizons that span several years. Uncertainty in the quality of reserves, on the other hand, stems from a different reason. The existence of oil or gas at an offshore site is indicated by seismic surveys and preliminary exploration tests. However, the actual amount of oil or gas in these reserves, and the efficacy of extracting these remain largely uncertain until after the investments have been made. Both, future oil or gas price and the quality of reserves directly affect the profitability of the project. Hence it is very important to consider the impact of uncertainty in these parameters when formulating the decision policy. In this paper, we consider planning of gas field developments under uncertainty in gas reserves.

An offshore gas production site has a number of reservoirs of gas, known as fields. The typical infrastructure at an offshore site includes well platforms (WPs), production platforms (PPs) and connecting pipelines, as shown in Figure 1. To extract gas from a field, a dedicated well platform (WP) needs to be installed at the field. Each WP is connected to another WP, or to a production platform (PP), through pipelines. Gas produced at all WPs is sent to the PPs through this network of pipelines. The gas is compressed at the PPs before being sent to the shore for sales.

The typical duration of an offshore gas production project is 10-30 years. During this project horizon, investment and operation decisions have to be made. Investment decisions include

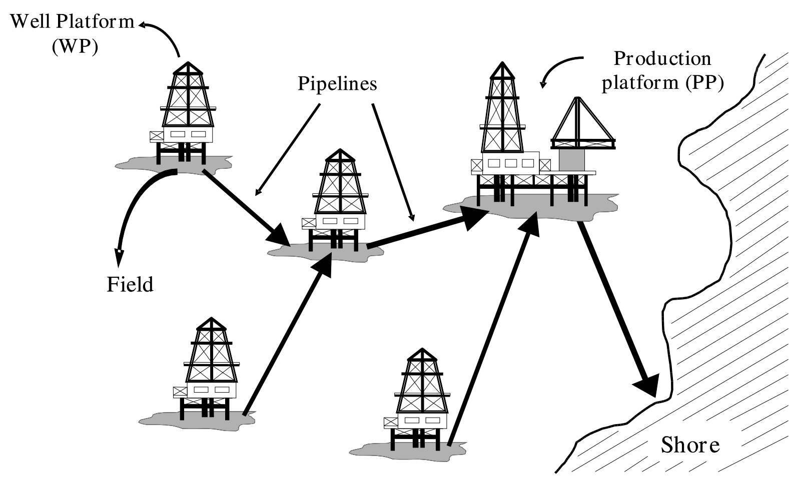


Figure 1: A typical offshore gas production site

selecting where and when to install the WPs and PPs, the capacities of these platforms, and the pipeline connections to be constructed. Operation decisions include determining the production profiles for the different fields over time.

To facilitate decision-making at the planning level, the project horizon is discretized into time periods and each of the above decisions have to be made for every period in the time horizon. Discrete decisions regarding which WPs, PPs and pipelines are to be installed, and when, are represented by binary (0-1) variables, while capacities of platforms and all operation related decisions are represented by continuous variables. This combination of discrete and continuous variables, along with the non-linear reservoir model that governs the gas production for a field leads to Mixed Integer Non-linear Programming (MINLP) models with the objective of maximizing or minimizing a specific value function. In this paper, we address the planning problem with uncertainty in the reservoir properties of the fields. Specifically, we present a multistage stochastic programming model and solution strategy for planning the development of a multi-field offshore site under uncertainty in the sizes and initial deliverabilities of the fields, where both, investment and operational decisions have to be made.

This paper is organized as follows. We first present a literature review investigating the previous work on oil and gas field planning under uncertainty, and the different approaches to handling uncertainty. We then present the problem statement followed by the description of the proposed model. Next, we propose an approximation solution strategy that relies on decomposition and involves solving scenario sub-problems and a sequence of two-stage stochastic programming problems. Finally, we illustrate the application of the proposed method with several example problems.

2 Literature Review

Most of the available literature for planning of oil and gas field infrastructures uses a deterministic approach (Ierapetritou *et al.* (1998), Iyer *et al.* (1998), Grothey and McKinnon (2000), Van den Heever and Grossmann (2000), Van den Heever and Grossmann (2001), Barnes *et al.* (2002), Kosmidis *et al.* (2002), Lin and Floudas (2002), Ortiz-Gomez *et al.* (2002)). For a recent review of the existing literature on deterministic approaches for these problems, please refer to Van den Heever and Grossmann (2001). Recently, there has been some work that considers uncertainty in this problem. Jonsbraten (1998a) presents an MILP model for optimizing the investment and operation decisions for an oil-field under uncertainty in oil prices. The author uses the Progressive Hedging Algorithm (PHA) to solve the problem. Jonsbraten (1998b) presents an implicit enumeration algorithm for the sequencing of oil wells under uncertainty in oil reserves. The author uses a Bayesian approach to represent the resolution of uncertainty with investments. The decision models for both these papers include investment and operation decisions for one field only. Jornsten (1992) uses Lagrangean relaxation of non-anticipativity constraints to solve a stochastic program for the sequencing of gas fields under uncertainty in future demands. The author assumes that production profiles and capacities of platforms have already been fixed. Haugen (1996) proposes a single parameter representation for uncertainty in the size of reserves and incorporates it into a Stochastic Dynamic Programming model for scheduling of petroleum fields. This work also assumes that the only decisions that need to be made are regarding the scheduling of fields. Meister *et al.* (1996) present a model to derive exploration and production strategies for one field under uncertainty in reserves and future oil price. The model is analyzed using stochastic control techniques. Lund (2000) presents a stochastic dynamic programming model for evaluating the value of flexibility in offshore development projects under uncertainty in future oil prices and in the reserves of one field. The author uses "simplified descriptions of the main variables in order to obtain a manageable model size". Jonsbraten (1998c) discusses an interesting though different problem associated with planning of oil field development. A situation where two surface lease owners with access to the same oil reservoir bargain their shares of production is considered. The author assumes a mixed-integer optimization model

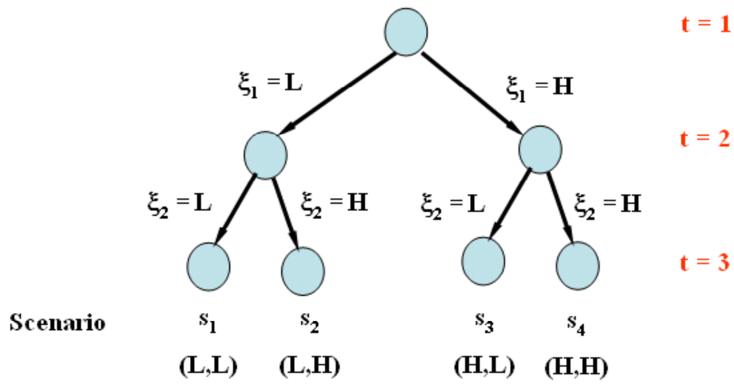


Figure 2: A fixed scenario tree for a problem with project exogenous uncertainty

and illustrates that there exists a unique Nash equilibrium for the non-cooperative normal form game. Recently, there has also been some work using real options based approaches (Dias and Rocha (1999), Dias (2001)) for planning of oil and gas field developments under uncertainty.

Based on the dependence of the stochastic process on the decisions, Jonsbraten (1998d) classifies uncertainty in planning problems into two categories: project exogenous uncertainty and project endogenous uncertainty.

Problems where the stochastic process is independent of the project decisions are said to possess project exogenous uncertainty. For these problems, the scenario tree is fixed and does not depend on the decisions. For example, Fig. 2 shows a scenario tree for a problem with three time periods ($t = 1, 2, 3$) and two random variables (ξ_1, ξ_2). Both ξ_1 and ξ_2 have two possible realizations, H(High) and L(Low). The uncertainty in ξ_1 is resolved after $t = 1$ while the uncertainty in ξ_2 is resolved after $t = 2$. The scenario tree in Fig. 2 is based on the assumption that uncertainty in each random variable is resolved at a pre-determined time, and this time, or the realizations of these random variables are not affected by the project decisions.

The uncertainty in gas price in a planning problem similar to the one described here is an example of project exogenous uncertainty. This is because the investment and operation decisions can neither affect when the uncertainty in future gas prices is resolved (the uncertainty associated with the gas price for a particular year is resolved no earlier and no later than that year) nor influence the values the gas prices will take (assuming a competitive market).

Most of the work in the stochastic programming literature deals with problems involving project exogenous uncertainty. For recent reviews on models and solution techniques for stochastic programs with project exogenous uncertainty, please refer to Kall and Wallace (1994), Birge (1997), Birge and Louveaux (1997), Stougie and van der Vlerk (1997), Haneveld and van der Vlerk (1999) and Sahinidis (2003). A "standard" stochastic programming model for a problem with project exogenous uncertainty is presented in Appendix A. A brief discussion on "non-anticipativity constraints", which will be used in our formulations, is also included along with the model. The most relevant characteristic of this standard stochastic programming model is that its formulation assumes a given scenario tree.

Problems where the project decisions influence the stochastic process are said to possess project endogenous uncertainty. A gas production planning problem with uncertainty in gas reserves is included in this category. This is because the uncertainty in gas reserves of a field is resolved only if, and when, exploration or investment is done at the field. If no action is taken, the uncertainty in the field does not resolve at all. For problems with project endogenous uncertainty, the scenario tree, and hence the non-anticipativity constraints are decision-dependent. This leads to severe difficulties in defining the model because, traditionally, the stochastic programming literature has relied on the assumption of fixed scenario trees. An example illustrating the dependence of the scenario tree on the decisions is included in Appendix B. This example also illustrates the fact that planning of gas field developments under uncertainty in reserves is a potentially multistage stochastic programming problem.

There is very little literature dealing with problems having process endogenous uncertainty. In fact, we know of only three papers (Pflug (1990), Jonsbraten *et al.* (1998), Jonsbraten (1998b)) that deal with such problems. Jonsbraten *et al.* (1998) propose an implicit enumeration algorithm for problems where the decisions controlling information can only be made in the first stage. Jonsbraten (1998b) applies a similar algorithm to a problem for the optimal sequencing of oil wells under uncertainty in reserves.

As can be seen from this review, most of the existing work that accounts for uncertainty in the gas field development planning problem deals with simplified cases, where either the investment or operation decisions have already been fixed, or the problem includes only one field. In this paper, we seek to develop a model and solution approach for a broader problem that deals with investment and operation decisions for a multi-field site.

3 Problem Statement

We consider the investment and operational planning for an offshore gas exploration and production site under uncertainty in gas reserves. It is assumed that a set of gas fields have been identified at the offshore site under consideration. Based on the accuracy of estimates of their gas reserves, these fields are classified into two categories: *Certain* and *Uncertain*. Fields for which estimates of gas reserves are known accurately are classified as *Certain*. The rest of the fields, gas reserves of which are relatively uncertain, are classified as *Uncertain*.

Some or all of these fields have to be exploited for gas over a project horizon of T years, which is discretized into T time periods of one year each. To produce gas from a field, a dedicated WP needs to be installed at the field. Potential pipeline connections from the WP at each field and potential locations for the PPs are given. Investment decisions for the project include selecting at which fields and in which periods should WPs be installed, at what locations and in which periods should PPs be installed, the capacities of WPs and PPs, and the pipeline connections to be installed. If more than one out-going pipeline connections from a WP are possible, exactly one has to be selected. Operation decisions include determining the production rates of each field for each time period.

It is assumed that investments are instantaneous and take place at the beginning of a time period, while operation takes place throughout the entire time period. Also, investments in a WP at any field, in a PP or in a pipeline can only occur once in the entire time horizon. Thus, the capacity of any platform cannot be expanded once it has been installed. Moreover, once installed, each of these units remains in existence for the remaining part of the project horizon.

The gas reserves of a field are characterized by the "size" and "deliverability" of the field. The size of a field refers to the total amount of gas that can be recovered from the field, while the deliverability of a field at any time is the maximum rate of gas production that can be obtained from the field. The deliverability of a field is highest (Initial Deliverability) when no gas has been recovered from the field, and decreases with increase in cumulative production from the field. When the cumulative production from the field equals the size of the field, the deliverability reduces to zero and hence, no more gas can be produced. In this paper, we assume that the deliverability of a field decreases linearly with the increase in cumulative production from the field. In other words, we assume a linear reservoir model, as shown in Fig. 3.

The uncertainty in gas reserves of a field is represented by uncertainty in the size and initial

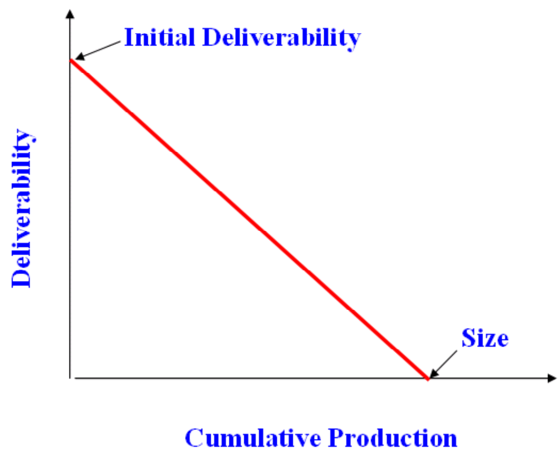


Figure 3: Linear Reservoir Model

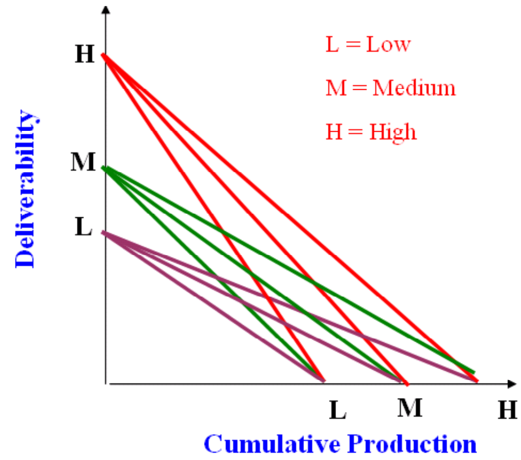


Figure 4: Representation of uncertainty in size and initial deliverability of a field

deliverability of the field. We use discrete probability distributions to represent the uncertainty in the size and the initial deliverability of each uncertain field. Thus, all future possibilities are represented by a set of "scenarios", where each scenario is a combination of the sizes and initial deliverabilities of the uncertain fields, and has a given probability. The objective is to find decisions that maximize the Expected Net Present Value (ENPV) of the project. Fig. 4 shows the nine scenarios associated with a problem with one uncertain field and three-point discrete distributions for the size and the initial deliverability of this field.

This problem is unique in the manner in which the uncertainty resolves. The uncertainty in reserves of an uncertain field persists until a WP is installed at the field. However, it is assumed that all uncertainty in the field is resolved completely as soon as a WP is installed at the field; i.e., the size and initial deliverability of the field are known deterministically once a WP is installed at the field.[1] Thus, investment at uncertain fields reduces the uncertainty. Operation decisions, on the other hand, do not have any effect on the level of uncertainty. This dependence of the resolution of uncertainty on the investment decisions implies that the scenario tree is not unique and depends on the decisions (see Appendix B for illustration). Another implication of this feature is that depending on when investments at the uncertain fields are made, the uncertainty resolves over the entire project horizon. Since both investment and operation decisions are spread over the project horizon, we consider recourse in investment and operation decisions.

[1]This is a simplification of reality because, in practice, installing a WP at a field yields significant amount of information about the field, but may not eliminate the uncertainty completely.

Thus, we consider recourse in both, integer and continuous variables.

4 Model

$$\text{(SPM)} \qquad \text{Max} \quad \sum_s p^s \left[\sum_t (c_{1t}^T q_t^s + c_{2t}^T d_t^s + c_{3t}^T y_t^s + \sum_{uf} c_{4t,uf}^T b_{uf,t}^s) \right] \tag{1}$$

$$A_t^s q_t^s \leq a_t^s \qquad \forall (t,s) \tag{2}$$

$$g_t(q_1^s, q_2^s, \ldots, q_t^s) \leq 0 \qquad \forall (t,s) \tag{3}$$

$$h_t(d_1^s, d_2^s, \ldots, d_t^s) \leq 0 \qquad \forall (t,s) \tag{4}$$

$$\sum_t b_{uf,t}^s \leq 1 \qquad \forall (uf,s) \tag{5}$$

$$r_{uf,t}(q_t^s, d_t^s, b_{uf,1}^s, b_{uf,2}^s, \ldots, b_{uf,t}^s, y_1^s, y_2^s, \ldots, y_t^s) \leq 0 \qquad \forall (uf,t,s) \tag{6}$$

$$\begin{bmatrix} & Z_t^{s,s'} & & \\ q_t^s & = & q_t^{s'} & \\ d_{t+1}^s & = & d_{t+1}^{s'} & \\ y_{t+1}^s & = & y_{t+1}^{s'} & \\ b_{uf,t+1}^s & = & b_{uf,t+1}^{s'} & \forall uf \end{bmatrix} \vee \begin{bmatrix} & \neg Z_t^{s,s'} & & \\ q_t^s, q_t^{s'} & \geq & 0 & \\ d_{t+1}^s, d_{t+1}^{s'} & \geq & 0 & \\ y_{t+1}^s, y_{t+1}^{s'} & \in & \{0,1\}^{dim(y)} & \\ b_{uf,t+1}^s, b_{uf,t+1}^{s'} & \in & \{0,1\} & \forall uf \end{bmatrix}$$

$$\forall (t,s,s'), s < s' \tag{7}$$

$$Z_t^{s,s'} \Leftrightarrow \bigwedge_{uf \in \mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t} \left(\neg b_{uf,\tau}^s \right) \right] \qquad \forall (t,s,s'), s < s' \tag{8}$$

$$b_{uf,1}^s = b_{uf,1}^{s'} \qquad \forall (uf,s,s'), s < s' \tag{9}$$

$$y_1^s = y_1^{s'} \qquad \forall (s,s'), s < s' \tag{10}$$

$$d_1^s = d_1^{s'} \qquad \forall (s,s'), s < s' \tag{11}$$

$$q_t^s,\ d_t^s \geq 0, \qquad b_{uf,t}^s \in \{0,1\},\ y_t^s \in \{0,1\}^{dim(y)}, \qquad Z_t^{s,s'} \in \{True, False\}$$

We propose the mathematical model[2] (SPM) where the multistage stochastic problem is formulated using conditional non-anticipativity constraints. (SPM) is a hybrid mixed-integer/disjunction

[2]Please refer to the Nomenclature section listed at the end of the paper.

optimization problem (see Raman and Grossmann (1994), Vecchietti and Grossmann (1999)).

Variable vectors q_t^s, d_t^s, $b_{uf,t}^s$ and y_t^s are defined for each time period t in scenario s. The variable vector q_t^s represents continuous operation variables such as the production rates at fields and the flow-rates through platforms and pipelines. The vector d_t^s represents continuous investment variables such as the capacities and the expansion decisions for each platform. $b_{uf,t}^s$ are binary investment variables representing whether or not a WP is installed at the uncertain field uf in time period t of scenario s. y_t^s is the vector of all other binary investment variables for time period t in scenario s. The elements of this vector represent decisions regarding whether or not the WPs at the *Certain* fields, the PPs and the pipelines are installed in period t of scenario s. The Boolean variables $Z_t^{s,s'}$ are used to represent whether or not scenarios s and s' are indistinguishable after investments have been made in time period t.

The objective, represented by (1), is to maximize the ENPV, which is the probability weighted average of the NPVs in each scenario. c_{1t}, c_{2t}, c_{3t}, $c_{4t,uf}$ represent discounted prices corresponding to variables q_t^s, d_t^s, y_t^s and $b_{uf,t}^s$ respectively. (2)-(6) represent constraints that govern investment and operation decisions for each individual scenario. (2) represents constraints that determine the operation decisions for period t. These include the linear reservoir model and mass balance constraints. The uncertainty in the problem is reflected by uncertainties in the coefficient matrix A_t, and the right-hand side vector a_t of these constraints. (3) represents linear constraints that link operation decisions for period t with operation decisions from previous periods. (4) represents linear constraints that link continuous investment decisions for period t with continuous investment decisions from previous periods. (5) represents the cardinality constraint that a WP can be installed at an uncertain field only once during the entire project horizon. Similar constraints that state that PPs, pipelines, and WPs at the certain fields can be installed only once, along with other logic, big-M and capacity constraints, are represented by (6). All of these constraints are linear. The specific constraints represented by (2)-(6) are given in Appendix C.

Decisions for different scenarios are linked to each other by "non-anticipativity" constraints ((7)-(11)). As explained before, we follow the convention that investments are made instantaneously at the beginning of each time period, while operation is performed during the entire period. Investment decisions in period t are made with information obtained from investments in period $t-1$. Operation decisions in period t, however, are made with the additional information obtained from investments done at the beginning of period t.

The Boolean variable $Z_t^{s,s'} \in \{True, False\}$ is $True$ if and only if scenarios s and s' are indistinguishable after investments have been made in period t. Non-anticipativity requires that if scenarios s and s' are indistinguishable after investment in period t , then operation decisions in period t, and investment decisions in period $t+1$ should be the same for scenarios s and s' (constraint (7)).

The logic expression (8) relates the logic variable $Z_t^{s,s'}$ with the investment decisions for the uncertain fields. We define the set $\mathcal{D}(s,s') \subseteq UF$ such that $uf \in \mathcal{D}(s,s')$ if and only if the uncertain properties (size and initial deliverability) of uf are not the same in s and s'. For example, consider a problem with two uncertain fields, A and B, where only the sizes of these fields are uncertain. Consider two realizations, H(High) and L(Low), for the size of each uncertain field. We represent the four scenarios in the problem by $s_1 = (L,L); s_2 = (L,H); s_3 = (H,L); s_4 = (H,H)$, where $s = (\xi_{size,A}, \xi_{size,B})$, and $\xi_{size,A}$ and $\xi_{size,B}$ represent the sizes of fields A and B respectively. Now, since scenarios s_1 and s_2 differ only in the size of uncertain field B, hence $\mathcal{D}(s_1,s_2) = \{B\}$. On the other hand, sizes of both A and B in scenario s_1 are different from their sizes in scenario s_4. Hence, $\mathcal{D}(s_1,s_4) = \{A,B\}$.

Constraint (8) states that if WPs have not been installed in or before period t, at any of the uncertain fields which cause s and s' to be different scenarios, (i.e. $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b_{uf,\tau}^{s}\right)\right] = True$), then no information is available that can help us to differentiate between scenarios s and s'. Thus $Z_t^{s,s'} = True$, and non-anticipativity constraints linking decisions for these scenarios need to be applied. Since no information is available before investments are made in the first period, non-anticipativity requires that investment decisions at $t=1$ be the same for all scenarios (constraints (9)-(11)).

Note that among constraints (7)-(11), only the right hand side of constraint (8) is not symmetric with respect to the scenario pair, (s,s'). However, in the context of this model, we can prove that $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b_{uf,\tau}^{s}\right)\right] \Leftrightarrow \bigwedge_{uf\in\mathcal{D}(s',s)} \left[\wedge_{\tau=1}^{t}\left(\neg b_{uf,\tau}^{s'}\right)\right]$ (see Appendix D for proof). Thus, constraints (7)-(11) are "symmetric" with respect to the pair of scenarios (s,s'). To avoid the duplication of non-anticipativity constraints for pairs of scenarios, we assume that the set of scenarios forms an ordered set, and constraints (7)-(11) are applied for (s,s') only if $s < s'$, i.e. s precedes s' in this ordering.

Appendix E includes an explanation of the application of the non-anticipativity constraints for the two-uncertain field example introduced above.

5 Solution Strategy

In its current form, model (SPM) involves Boolean variables and disjunctions. This model can be converted to a standard MILP form by replacing the Boolean variables by binary (0-1) variables, and using big-M constraints or a convex-hull transformation (Balas (1985), Turkay and Grossmann (1996)) to convert the disjunctions to linear constraints. However, (SPM) suffers from the "curse of dimensionality". The number of Boolean variables and the number of non-anticipativity constraints are both quadratic in the number of scenarios. Also, the number of scenarios in the problem is exponential ($m^{2|UF|}$; $m =$ Number of discretization points for each uncertain parameter, $|UF| =$ Number of uncertain fields) in the number of uncertain fields. The size of the model thus explodes in the number of uncertain fields. Therefore, it is clear that a specialized solution strategy capable of dealing with the size of the problem is needed.

We propose an approximation algorithm, based on decomposition, that aims at finding "good" solutions to (SPM). The proposed algorithm restricts the search to solutions that satisfy

$$b^s_{uf,t} = b^{s'}_{uf,t} \qquad \forall (uf, t, s, s'), s < s' \tag{12}$$

in addition to constraints (2)-(11). Hence, the proposed algorithm searches in a sub-space of the feasible region of (SPM). The restriction that (12) hold implies that variables $b^s_{uf,t}$ be independent of scenario, and hence that for any time period t, there be no recourse in decisions regarding whether or not we invest in an uncertain field. $(CSPM)$ is the constrained stochastic programming model obtained from (SPM) by dropping index s from the variables $b^s_{uf,t}$ (which is equivalent to adding the constraint (12)).

It should be noted that since the non-anticipativity constraints on $b^s_{uf,t}$ are automatically satisfied by (12), these are not included in $(CSPM)$. The proposed algorithm aims at finding the optimal solution of problem $(CSPM)$ rather than that of (SPM). Note that since the feasible region of $(CSPM)$ is a subset of the feasible region of (SPM), any solution that is feasible in $(CSPM)$ is also feasible in (SPM). However, the optimal solution to $(CSPM)$ need not be optimal to (SPM).

The proposed algorithm proceeds by searching in the space of variables $b_{uf,t}$. A schematic outline of the algorithm is presented in Fig. 5. An upper bound to the optimal solution of $(CSPM)$ is first obtained by solving a relaxed problem. In $(CSPM)$, constraints (19)-(22) and variables $b_{uf,t}$ couple the scenarios together. A valid relaxation $(RCSPM)$ to $(CSPM)$ is obtained by removing constraints (19)-(22) and disaggregating variables $b_{uf,t}$ into $b^s_{uf,t}$. The relaxed model

$$\text{(CSPM)} \qquad \text{Max} \quad \sum_s p^s \left[\sum_t (c_{1t}^T q_t^s + c_{2t}^T d_t^s + c_{3t}^T y_t^s) \right] + \sum_t \sum_{uf} c_{4t,uf}^T b_{uf,t} \tag{13}$$

$$A_t^s q_t^s \leq a_t^s \qquad \forall (t,s) \tag{14}$$

$$g_t(q_1^s, q_2^s, \ldots, q_t^s) \leq 0 \qquad \forall (t,s) \tag{15}$$

$$h_t(d_1^s, d_2^s, \ldots, d_t^s) \leq 0 \qquad \forall (t,s) \tag{16}$$

$$\sum_t b_{uf,t} \leq 1 \qquad \forall uf \tag{17}$$

$$r_{uf,t}(q_t^s, d_t^s, b_{uf,1}, b_{uf,2}, \ldots, b_{uf,t}, y_1^s, y_2^s, \ldots, y_t^s) \leq 0 \qquad \forall (uf,t,s) \tag{18}$$

$$\begin{bmatrix} & Z_t^{s,s'} & \\ q_t^s & = & q_t^{s'} \\ d_{t+1}^s & = & d_{t+1}^{s'} \\ y_{t+1}^s & = & y_{t+1}^{s'} \end{bmatrix} \vee \begin{bmatrix} & \neg Z_t^{s,s'} & \\ q_t^s, q_t^{s'} & \geq & 0 \\ d_{t+1}^s, d_{t+1}^{s'} & \geq & 0 \\ y_{t+1}^s, y_{t+1}^{s'} & \in & \{0,1\}^{dim(y)} \end{bmatrix} \qquad \forall (t,s,s'), s < s' \tag{19}$$

$$Z_t^{s,s'} \Leftrightarrow \bigwedge_{uf \in \mathcal{D}(s,s')} \left[\wedge_{\tau=1}^t (\neg b_{uf,\tau}) \right] \qquad \forall (t,s,s'), s < s' \tag{20}$$

$$y_1^s = y_1^{s'} \qquad \forall (s,s'), s < s' \tag{21}$$

$$d_1^s = d_1^{s'} \qquad \forall (s,s'), s < s' \tag{22}$$

$$q_t^s,\ d_t^s \geq 0, \qquad b_{uf,t} \in \{0,1\},\ y_t^s \in \{0,1\}^{dim(y)}, \qquad Z_t^{s,s'} \in \{True, False\}$$

decomposes into one sub-problem for each scenario s, each of which is solved independently. The solution of $(RCSPM)$ gives an upper bound, UB, on the optimal solution of $(CSPM)$.

Lower bounds to the optimal solution are obtained by generating feasible solutions for $(CSPM)$. The expected value problem (EVP) is first solved and the solution is used to fix variables $b_{uf,t}$ in $(CSPM)$. (EVP) is simply the deterministic approximation of the problem with all uncertain parameters $(A_t^{(.)}, a_t^{(.)}$ in (14)) replaced with their expected (or mean) values $(\bar{A}_t, \bar{a}_t)$. Fixing variables $b_{uf,t}$ fixes the logic variables $Z_t^{s,s'}$ (by (20)) and the non-anticipativity conditions to be applied, which in turn defines the scenario tree for model $(CSPM)$. The multistage stochastic program associated with this fixed scenario tree is then solved to optimize decisions q_t^s, d_t^s, y_t^s. The solution obtained, along with the values for variables $b_{uf,t}$ constitutes a feasible solution to

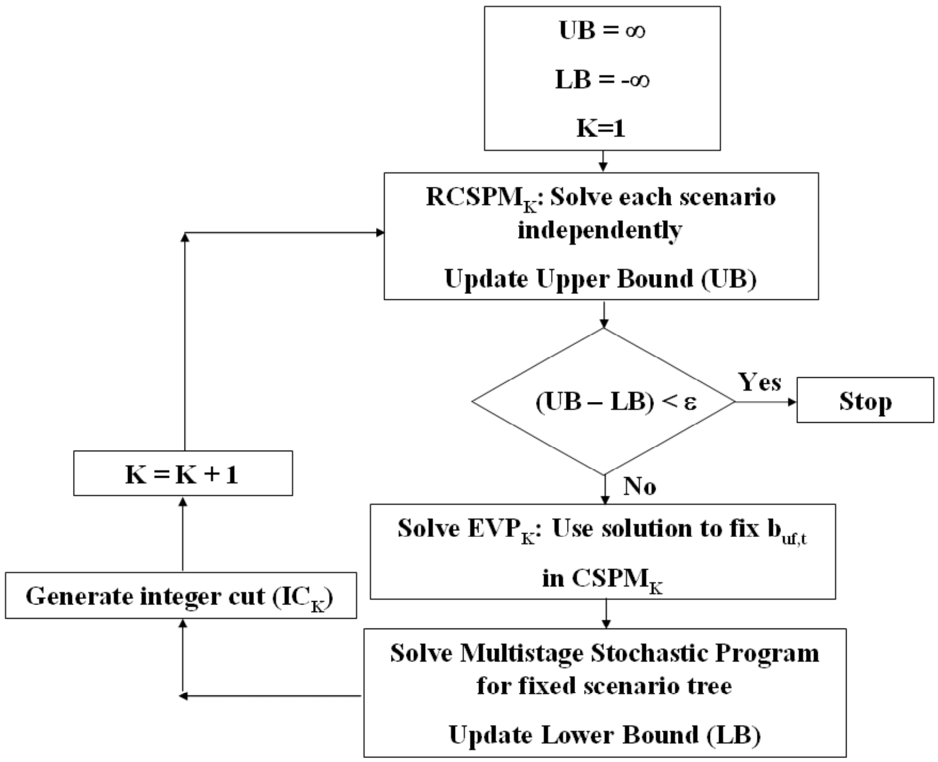


Figure 5: Schematic outline of proposed solution strategy

$(CSPM)$, and hence gives a lower bound, LB, to the optimal solution.

The current choice of values for $b_{uf,t}$ is then excluded from the feasible space by adding an integer cut to $(CSPM)$. (IC_K) represents the integer cut added at iteration K of the algorithm.

$$\sum_{(uf,t)\in B_K} b_{uf,t} - \sum_{(uf,t)\in N_K} b_{uf,t} \leq |B_K| - 1 \qquad (IC_K)$$

$B_K = \{(uf,t)|\hat{b}_{uf,t,K} = 1\}$ and $N_K = \{(uf,t)|\hat{b}_{uf,t,K} = 0\}$ where $\hat{b}_{uf,t,K}$ is the set of values for variables $b_{uf,t}$ used to fix the scenario tree at iteration K.

Since we use the solution of (EVP) to fix variables $b_{uf,t}$ in $(CSPM)$ at each iteration, the integer cuts are included in (EVP) also to avoid fixing variables $b_{uf,t}$ to the same set of values in the next iteration. Models $(CSPM)$, $(RCSPM)$ and (EVP) at iteration K are represented by $(CSPM_K)$, $(RCSPM_K)$ and (EVP_K) respectively, and are included in Appendix F.

$$\text{(RCSPM)} \quad \text{UB} = \quad \text{Max} \quad \sum_s p^s \left[\sum_t (c_{1t}^T q_t^s + c_{2t}^T d_t^s + c_{3t}^T y_t^s + \sum_{uf} c_{4t,uf}^T b_{uf,t}^s) \right] \tag{23}$$

$$A_t^s q_t^s \leq a_t^s \qquad \forall (t,s) \tag{24}$$

$$g_t(q_1^s, q_2^s, \ldots, q_t^s) \leq 0 \qquad \forall (t,s) \tag{25}$$

$$h_t(d_1^s, d_2^s, \ldots, d_t^s) \leq 0 \qquad \forall (t,s) \tag{26}$$

$$\sum_t b_{uf,t}^s \leq 1 \qquad \forall (uf,s) \tag{27}$$

$$r_{uf,t}(q_t^s, d_t^s, b_{uf,1}^s, b_{uf,2}^s, \ldots, b_{uf,t}^s, y_1^s, y_2^s, \ldots, y_t^s) \leq 0 \qquad \forall (uf,t,s) \tag{28}$$

$$q_t^s,\ d_t^s \geq 0, \qquad b_{uf,t}^s \in \{0,1\},\ y_t^s \in \{0,1\}^{dim(y)}$$

Note that $(RCSPM_K)$ is a relaxation of $(CSPM_K)$. Hence, the solution of $(RCSPM_K)$ gives an upper bound to all feasible solutions in $(CSPM_K)$. Thus, if at iteration K, the upper bound obtained from the solution of $(RCSPM_K)$ is smaller than the current lower bound, then none of the remaining set of choices for variables $b_{uf,t}$ can provide a feasible solution better than the current best, and hence the algorithm is terminated.

Solution of Multistage Stochastic Programs with Mixed-Integer Recourse

The multistage stochastic programs obtained after fixing variables $b_{uf,t}$ have both continuous (q_t^s, d_t^s) and binary (y_t^s) variables, and thus have mixed-integer recourse. Because of the large number of scenarios, the solution of these problems can be computationally expensive. Hence, we use a moving-shrinking horizon approach to solve these multistage stochastic problems approximately. Suppose fixing variables $b_{uf,t}$ in $(CSPM)$ results in an n-stage stochastic program. We approximate this n-stage stochastic program by a two-stage program simply by relaxing the non-anticipativity constraints for all stages but the 1^{st} stage. This two-stage approximation problem is solved and the solution is used to fix the 1^{st} stage decisions in the n-stage stochastic program. Fixing these decisions decomposes the n-stage stochastic program in to a set of $(n-1)$-stage stochastic programs. Two-stage approximations are developed for each of these $(n-1)$-stage stochastic programs and the solutions of these approximation problems are used to fix decisions in the 2^{nd} stage of the n-stage stochastic program. This procedure is repeated until decisions in all stages of the n-stage stochastic program have been fixed. The solution obtained

$$\text{(EVP)} \qquad \text{Max} \quad \left[\sum_t (c_{1t}^T q_t + c_{2t}^T d_t + c_{3t}^T y_t)\right] + \sum_t \sum_{uf} c_{4t,uf}^T b_{uf,t} \tag{29}$$

$$\bar{A}_t q_t \leq \bar{a}_t \qquad \forall t \tag{30}$$

$$g_t(q_1, q_2, \ldots, q_t) \leq 0 \qquad \forall t \tag{31}$$

$$h_t(d_1, d_2, \ldots, d_t) \leq 0 \qquad \forall t \tag{32}$$

$$\sum_t b_{uf,t} \leq 1 \qquad \forall uf \tag{33}$$

$$r_{uf,t}(q_t, d_t, b_{uf,1}, b_{uf,2}, \ldots, b_{uf,t}, y_1, y_2, \ldots, y_t) \leq 0 \qquad \forall (uf, t) \tag{34}$$

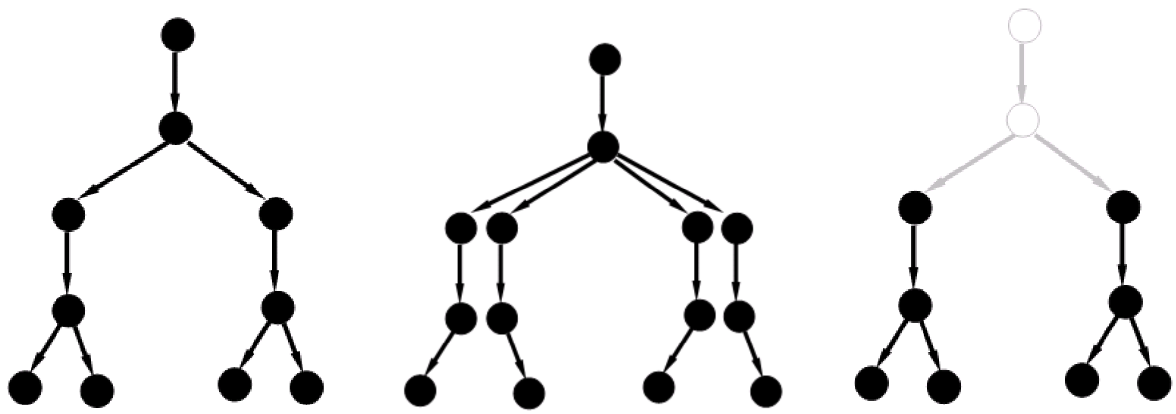
$$q_t,\ d_t \geq 0, \qquad b_{uf,t} \in \{0,1\},\ y_t \in \{0,1\}^{dim(y)}$$

is a feasible solution for $(CSPM)$.

Solution of two-stage approximation problems

If the number of scenarios is large, the solution of the two-stage stochastic programs can be computationally expensive due to the increased number of binary variables (y_t^s). We solve these two-stage stochastic programs in full space with all binary variables treated as 1^{st} stage variables. Note that if due to the large number of scenarios, these two-stage problems become too large to be solved in full space, then specialized methods such as L-shaped method (Benders decomposition) (Van Slyke and Wets (1969), Geoffrion (1972)) or Stochastic decomposition (Higle and Sen (1991)) could be used.

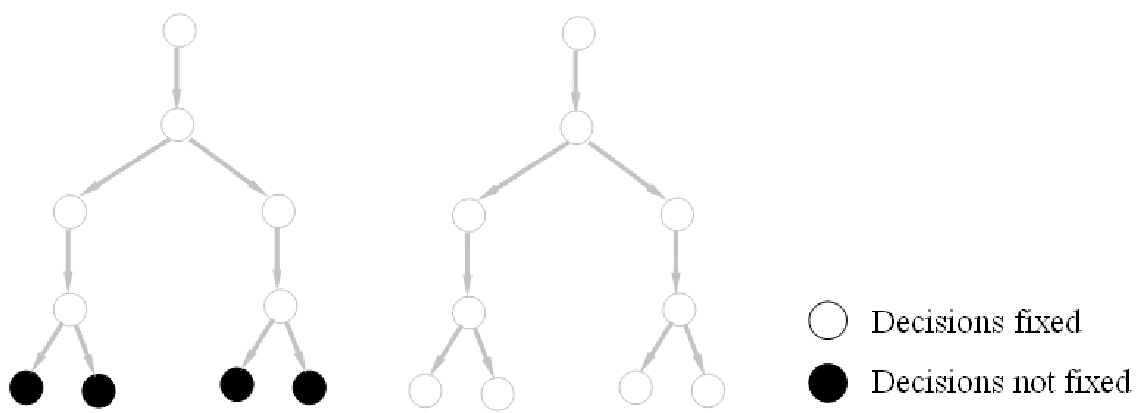
Fig. 6 illustrates the application of the proposed moving-shrinking horizon approach to the solution of a three-stage stochastic problem (Fig. 6(a)). The two-stage approximation, obtained by relaxing the non-anticipativity constraints in the 2^{nd} stage, is shown in Fig. 6(b). The solution of this two-stage approximation is used to fix the first stage decisions in the three-stage problem. As a result, the three-stage problem decomposes into two two-stage problems (Fig. 6(c)), which are then solved independently. The 1^{st} stage part of the solutions of these two-stage problems are used to fix the 2^{nd} stage decisions in the three-stage problem. Four more one-stage problems (Fig. 6(d)) are then solved and their solutions used to fix decisions in the third-stage of the three-stage problem (Fig. 6(e)). Note that the last step is needed because, while solving the two-stage problems, we treat all binary variables as first-stage variables, whereas in reality, recourse in



(a) Three-stage problem

(b) Two-stage approximation obtained by relaxing non-anticipativity constraints in 2^{nd} stage

(c) Solution of two-stage approximation used to fix 1^{st} stage decisions; Three-stage problem decomposes into two two-stage problems



(d) Solution of two-stage problems used to fix 2^{nd} stage decisions; Four one-stage problems solved

(e) Solution used to fix decisions in 3^{rd} stage. Feasible solution to three-stage problem found

Figure 6: Illustration of the moving-shrinking horizon approach for a three-stage problem

these variables is allowed.

Regarding the main steps of the proposed algorithm, the following can be stated:

1. Since the number of possible combinations for variables $b_{uf,t}$ is finite, the algorithm is guaranteed to stop after a finite number of iterations. However, in some cases, a large number of iterations may be needed for the bounds to cross each other. Also, if the number of scenarios is large, the completion of each iteration could be very time consuming. In such cases, we terminate the algorithm if the bounds do not improve for a pre-specified number of iterations.

2. In Appendix G, we prove that if the problem involves only one uncertain field, then $(SPM) \equiv (CSPM)$. Thus in this case, if the stochastic programs corresponding to the fixed scenario trees at each iteration are solved to optimality, then the proposed approach will give the optimal solution to the original problem, (SPM).

3. Problem $(RCSPM)$ is a relaxation of the model (SPM). Therefore, the upper bound obtained from the solution of $(RCSPM_K)$ for $K = 1$ gives an upper bound to the optimal solution of (SPM). Thus, the difference between the solutions of $(RCSPM_K)$ for $K = 1$ and the optimal solution of $(CSPM)$ is itself an upper bound on the "compromise" due to solving $(CSPM)$ rather than (SPM).

4. With regard to obtaining upper bounds from $(RSCPM_{K+1})$, if the optimal solution of a scenario sub-problem at iteration K is not made infeasible by the integer cut (IC_K) included at the end of iteration K, then the optimal solution for this scenario sub-problem remains unchanged and it need not be solved again while computing the upper bound at iteration $K + 1$.

5. The proposed algorithm can also be viewed as a search in the space of scenario trees. The variables $b^s_{uf,t}$ determine the non-anticipativity constraints to be applied, and hence the scenario tree. By restricting our search to solutions that satisfy (12), we restrict the search to the space of symmetric scenario trees only. Note that symmetry here refers to symmetry in information rather than symmetry simply in the shape of the scenario trees.

   At each iteration, one of the symmetric scenario trees is selected from the feasible space of scenario trees, by fixing variables $b_{uf,t}$. The multistage stochastic program associated with this scenario tree is then solved to optimize decisions q^s_t, d^s_t, y^s_t. The solution obtained is a feasible solution to the problem and hence gives a lower bound to the optimal solution. The integer cut on variables $b_{uf,t}$ serves to exclude this symmetric scenario tree from the feasible space of scenario trees.

The relaxed model $(RCSPM_K)$ is the "wait and see" model for all symmetric scenario trees that have not yet been investigated. Therefore, the solution of $(RCSPM_K)$ gives an upper bound to the optimal solutions associated with all symmetric scenario trees remaining in the feasible space. The algorithm terminates when none of the scenario trees remaining in the feasible space can provide a solution better than the current best.

6. An important feature of this algorithm is that it is amenable to parallelization (although this was not implemented in this work). While computing upper bounds, the relaxed model $(RCSPM_K)$ decomposes into scenario sub-problems, each of which can be solved in parallel. While solving the multistage stochastic programs to obtain lower bounds, we use a moving-horizon approach where solutions of two-stage approximations are used to fix 1^{st} stage decisions in an n-stage stochastic program. As a result, the n-stage program decomposes in to a set of $(n-1)$-stage stochastic programs. The moving-horizon approach can then be applied to these $(n-1)$-stage stochastic programs in parallel.

Next, we present a set of examples that demonstrate the value of the stochastic programming model and the effectiveness of the proposed solution strategy. To demonstrate the importance of considering uncertainty in this problem, we compare the solution obtained by the proposed approach with that obtained using a deterministic approach. The effectiveness of the solution strategy is demonstrated by a comparison of the solution time of the proposed approach with that of the full space method. It should be noted that the full space method was computationally more efficient when applied to the big-M formulation of (SPM), as compared to when applied to the convex-hull reformulation of (SPM) (see Raman and Grossmann (1994), Lee and Grossmann (2000)). Thus, we used the results for the big-M formulation of (SPM) for this comparison. All models are written in GAMS and solved using CPLEX 7.5 on a Pentium III, 667 MHz, 256 MB RAM machine.

Example 1

The problem under consideration is the investment and operational planning for an offshore site with six fields, denoted by the letters A-F, and one PP, as shown in Fig. 7. Investment

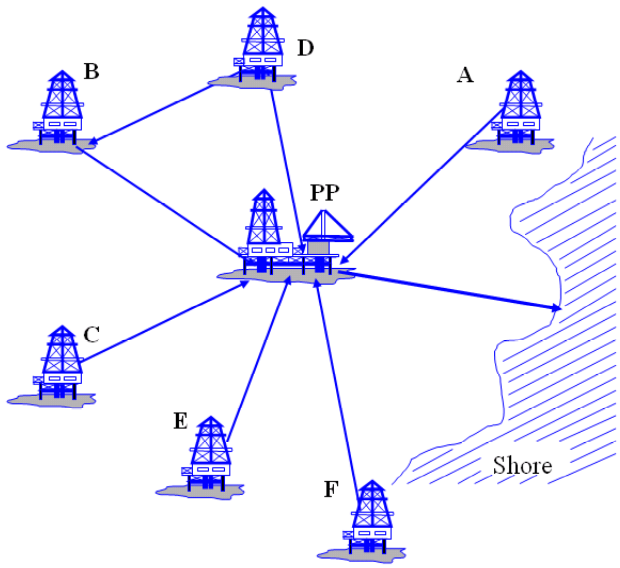


Figure 7: Superstructure for Examples 1, 2 and 3

and operation decisions for this offshore site have to be made for a horizon of 15 years (15 time periods). Among fields A-F, field F is uncertain while fields A-E are certain. For ease of illustration in this example, we assume that only the size of field F is uncertain, and that it has three possible realizations, represented by L(Low), M(Medium) and H(High). The problem thus has three scenarios. The initial deliverabilities (in MSCF/D)[3] and sizes (in BSCF)[4] of the fields, along with the probabilities of the respective sizes for field F, are shown in Table 1.

	A	B	C	D	E	F					
						Low	p^{Low}	Med.	$p^{Med.}$	High	p^{High}
Size (BSCF)	400	400	350	200	290	130	0.3	300	0.4	470	0.3
Initial Deliverability (MSCF/D)	130	200	100	100	130	130					

Table 1: Field properties for Example 1

[3]MSCF/D = Million Standard Cubic Feet per Day
[4]BSCF = Billion Standard Cubic Feet

A deterministic approach to solving this problem could be to solve the expected value problem (EVP) where mean (or expected) values for the size of field F are assumed. The solution of (EVP) proposes that the PP, and the WPs at the fields A, B, C and F be installed in year 1, while investments at fields E and D be made in years 5 and 7 respectively. The uncertainty in the size of field F would be resolved as soon as the WP is installed at this field. The model would then be resolved to find the optimal future decisions, given that investments in the first year have been fixed using the solution of (EVP). The investment years for the platforms in each scenario, proposed by the solutions thus obtained, along with the investment decisions for year 1 that were obtained from the solution of (EVP), and the associated NPV for each scenario are listed in Table 2. The ENPV of this solution over the three scenarios is \$94.56 Million. Note that scenarios 1, 2 and 3 correspond to low, medium and high sizes, respectively, for field F. In the remaining portion of this paper, we refer to a solution obtained using this approach as a "deterministic solution". Note that this solution has in-built recourse because of the re-optimization performed for each scenario, and is not simply the solution of (EVP). Thus, we find that the deterministic solution proposes that investment in the uncertain field be done in year 1.

Using the proposed approach, we obtain a solution with ENPV = \$99.55 Million. This solution proposes that the PP, and WPs at fields A, B, C and E be installed in year 1, while the WP at uncertain field F be installed later, in year 5. Investments at field D are proposed to be made in years 6 or 7 depending on the size of field F. The proposed investments, along with the NPV for each scenario are listed in Table 3. Thus, for this example, the Value of the Stochastic Solution (VSS) is \$4.99 Million (= 99.55 − 94.56).

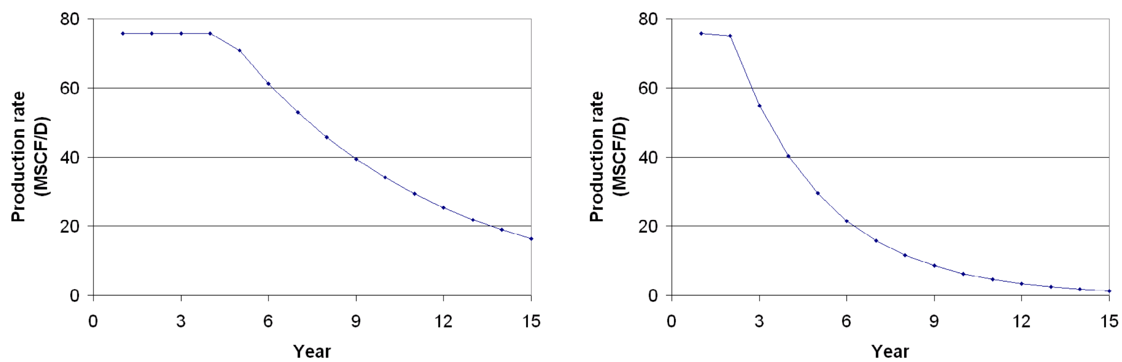
The large difference in ENPV obtained from the two approaches can qualitatively be explained as follows. Each investment decision leads to high costs and hence has a strong impact on the NPV. Thus, if solution I can maintain the same revenues as solution II but delay the investments, then solution I will have a higher NPV than solution II. By assuming a mean value for the size of field F, the deterministic approach assumes that field F would produce to full capacity for four years (as shown in Fig. 8(a); proposed WP capacity for field F = 76 MSCF/D). And at the beginning of year 5, when the production plateau for field F ends, a WP would be installed at field E to offset the reduction in total production rate. However, if the size of field F turns out to be low (scenario 1), then as shown in Fig. 8(b), field F can produce to full capacity for two years only and the investment at field E has to be done earlier, in year 3 (see Table 2). This preponement of investment, that is worth \$111 Million, leads to a significant reduction in NPV for

	Scenario		
	1	2	3
Year 1	A, B, C, F, PP		
Year 3	E	-	-
Year 5	-	E	E
Year 6	D	-	-
Year 7	-	D	-
Year 8	-	-	D
NPV ($ Million)	-24.18	117.26	183.03
ENPV ($ Million)	94.56		

Table 2: Deterministic solution (with recourse) for Example 1

	Scenario		
	1	2	3
Year 1	A, B, C, E, PP		
Year 5	F		
Year 6	D	-	-
Year 7	-	D	D
NPV ($ Million)	15.73	115.97	161.48
ENPV ($ Million)	99.55		

Table 3: Proposed solution for Example 1



(a) Predicted assuming mean size for field F: Production Plateau = 4 years

(b) Actual production rate if field F has low size: Production Plateau = 2 years

Figure 8: Production profile of Field F for deterministic solution

this scenario. The solution obtained from the proposed approach, on the other hand, proposes investment only in the certain fields in year 1, thereby ensuring that no more investments have to be made until year 5 (Table 3). Hence the large difference in NPV for the two solutions in scenario 1. The investment decisions and the NPV of the two solutions in scenario 2 are similar. In scenario 3, the deterministic solution is able to delay investment in field D compared to the proposed solution and hence lead to a higher NPV. However, this difference cannot offset the larger difference in NPVs in scenario 1. Hence the higher ENPV for the proposed solution. Besides the large difference in the ENPV, it is also important to note that the solution obtained using the proposed approach leads to a positive NPV in all scenarios, whereas the solution from the deterministic approach results in a large negative NPV in scenario 1 (see Tables 2 and 3).

The progression of bounds, the ENPV of the feasible solution obtained at every iteration, along with the year of investment for field F used to obtain the solution are given in Table 4. Note that at the first iteration, the year of investment for field F is the same as that proposed by the deterministic solution. This is because both, the deterministic approach, and the first iteration of the proposed algorithm use the solution of (EVP) to fix the investment year for uncertain field F.

Iteration Number	UB ($ Million)	Soln. obtained ENPV($ Million)	LB ($ Million)	Investment year: Field F
1	110.49	94.82	94.82	1
2	103.99	99.55	99.55	5
3	101.39	93.55	99.55	6
4	100.89	91.71	99.55	4
5	99.08	*	*	*

Table 4: Progression of bounds for Example 1

	Full space	Proposed Strategy
ENPV ($ Million)	99.55	99.55
Solution time (CPU seconds)	1,776	105

Table 5: Comparison of full space solution and proposed solution for Example 1

The algorithm terminates at the 5^{th} iteration when the upper bound obtained is smaller than the current lower bound. The total solution time required is 105 CPU seconds. Table 5 shows that the big-M formulation of (SPM), when solved in full space, returns the same solution but requires significantly more CPU time. Thus, for this example, the proposed solution strategy obtains the optimal solution of (SPM) with more than one order of magnitude reduction in solution time, as compared to the full space method.

Example 2

We extend Example 1 by including uncertainty in initial deliverability of field F, in addition to the uncertainty in size. The three possible realizations for initial deliverability of field F, along with their probabilities are given in Table 6. Note that the size and initial deliverability of a field are independent of each other. Hence, there are 9 scenarios in the problem.

	Initial Deliverability (MSCF/D)	Probability
Low	80	0.3
Medium	130	0.4
High	180	0.3

Table 6: Initial deliverabilities for field F in Example 2

The investment decisions obtained using the deterministic approach remain the same as in Example 1. The ENPV of the deterministic solution, computed over the 9 scenarios, is $84.51 Million. Table 7 shows the progression of bounds obtained using the proposed approach. The best solution has an ENPV = $92.06 Million and is obtained at the second iteration. Thus, including uncertainty in the initial deliverability of field F increases the VSS to $7.55 Million. The years of investment for the various platforms proposed by this solution are the same as those proposed by the solution in Example 1. The algorithm terminates after 7 iterations and the total solution time is 488 CPU seconds. The full space method, on the other hand, could only find a solution with ENPV = $38.91 Million (upper bound = 134.17) in 58 hours of CPU time. Table 8 compares the size of this full space model with that of the largest problem solved using the proposed approach.

An important point to be noted in Table 7 is that the solution obtained at the first iteration of the proposed approach has a significantly higher ENPV than the deterministic solution, although both these solutions were obtained with investment in uncertain field F fixed to year 1. This is because realizing the uncertainty in field F, the proposed approach chooses a lower capacity than that chosen by the deterministic solution (68 MSCF/D rather than 76 MSCF/D) for the WP at field F. As a result, the plateau length for field F is increased to 3 years (rather than 2 years for the deterministic solution), and the investment in field E has to be done in year 4, even

Iteration Number	UB ($ Million)	Soln. obtained ENPV($ Million)	LB ($ Million)	Investment year: Field F
1	104.75	87.69	87.69	1
2	98.24	92.06	92.06	5
3	95.86	85.57	92.06	6
4	95.28	85.18	92.06	4
5	93.59	82.20	92.06	7
6	92.69	77.89	92.06	3
7	92.25	*	*	*

Table 7: Progression of bounds for Example 2

	Full space	Proposed Strategy
Constraints	43,684	12,464
Continuous Variables	11,932	9,027
Binary Variables	1,782	198

Table 8: Comparison of size of full space model for Example 2 with that of largest problem solved in proposed approach

if both the size and the initial deliverability of F turn out to be "Low". Hence, the investment in field E, that is worth \$111 Million, is delayed by atleast one year compared to the deterministic solution.

Besides the large VSS of \$7.55 Million, the solutions also differ in the NPVs for each scenario. While the deterministic solution results in a negative NPV in three of the nine scenarios (total probability = 0.3), the solution obtained from the proposed approach leads to a positive NPV in all scenarios.

It is also important to note here that feasible solutions obtained at the first 4 iterations of the algorithm have a higher ENPV than the solution obtained by the deterministic approach. In gas field planning problems, it is always preferable to have multiple "good" solutions rather than "one" optimal solution, for two reasons. Firstly, not all logistic and investment related constraints can be represented in the model. Thus one investment policy, though with a slightly smaller ENPV, may be preferable to another because it is easier to implement. The second

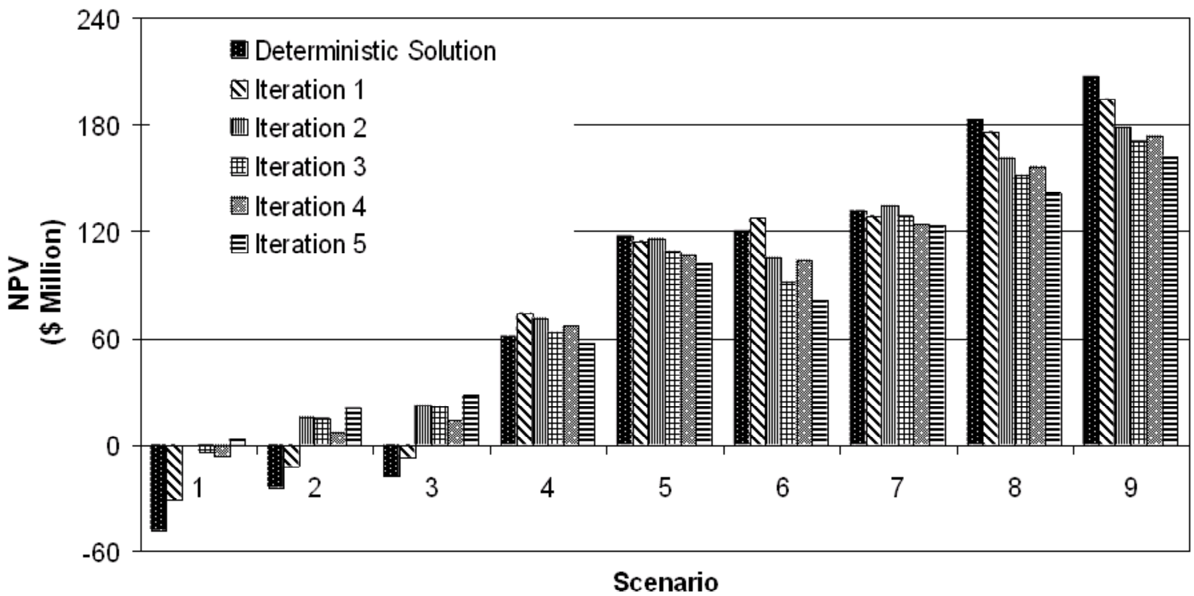


Figure 9: NPV in each scenario for the deterministic solution and the solutions obtained at iterations 1 to 5 of proposed approach

reason is motivated by the need for less "risky" solutions. Here we choose to maximize the ENPV of the project. However, from the point of view of the decision-maker, it is extremely helpful to have alternative solutions that may not have the highest ENPV but involve less risk (see Barbaro and Bagajewicz (2002)).

Fig. 9 compares the deterministic solution and the feasible solutions obtained at the first five iterations of the proposed approach, in terms of their NPVs in the nine scenarios. Note that although the solution obtained at iteration 5 has the lowest ENPV among these six solutions, it is the only solution that leads to a positive NPV in each scenario (the solution obtained at iteration 2 has NPV = 0 in scenario 1). Also, it is important to note here that solutions obtained at iterations 3 and 4 are "stochastically dominated[5]" by the solution obtained at iteration 2. Therefore, unless these solutions have a special advantage in terms of ease of implementation, they would not be preferred over the solution obtained at iteration 2.

An important question faced by decision-makers in this industry is that, given a set of potential

[5]Solution I is stochastically dominated by solution II if for every scenario, the NPV of solution II is at least as large as that of solution I, and strictly greater for at least one scenario (Barbaro and Bagajewicz (2002)).

fields, whether to invest in small fields that are "certain", or fields that are large but have significant uncertainty associated with them. In the following examples, we try to address this problem. It should be noted here that while the same set of data was used for examples 1 and 2, the following examples are based on independent sets of data.

Example 3

In this example, the superstructure and the project horizon is the same as that in example 1. However, here we treat fields D and E as uncertain while fields A, B, C and F are certain. Table 9 gives the sizes and initial deliverabilities of all the fields. We assume that the probabilities of size or initial deliverability of an uncertain field having high, medium or low values are 0.3, 0.4 and 0.3 respectively. Note that the size and initial deliverabilities of the uncertain fields are independent of each other. Hence there are 81 scenarios in the problem. Also, notice that on the basis of mean values, fields D and E are larger than the certain fields F and C respectively, but they also have a large amount of uncertainty associated with them.

	A	B	C	D			E			F
				Low	Med.	High	Low	Med.	High	
Size (BSCF)	400	400	250	200	320	440	200	350	500	250
Initial Deliverability (MSCF/D)	130	200	130	190	120	150	100	130	160	120

Table 9: Sizes and Initial deliverabilities for fields in Example 3

The deterministic solution proposes that the PP, and WPs at certain fields A, B, F, and at uncertain field E be installed in year 1; WP at uncertain field D be installed in year 4, 5 or 7 depending upon the size and initial deliverability of field E; finally, WP at field C be installed in years 5, 6 or 7 depending upon the sizes and initial deliverabilities of fields D and E. The ENPV of this solution is \$143.43 Million.

The full space model of the big-M formulation of (SPM), with more than 16,000 binary variables and 2.7 Million constraints, was intractable. Table 10 shows the progression of bounds for the proposed approach. The algorithm was terminated after 6 iterations (solution time = 4,990 CPU seconds, gap = 8.9 %) because the bounds did not show significant improvement for three successive iterations. The best solution obtained has an ENPV = \$146.32 Million, leading to a VSS = \$2.89 Million. This solution differs from the deterministic solution because here, the proposed investment policy consists of installing the PP and WPs at fields A, B, C and F, all of which are certain fields, in year 1. Investment in the uncertain fields D and E is proposed for years 7 and 5 respectively.

Iteration Number	UB ($ Million)	Soln. obtained ENPV($ Million)	LB ($ Million)	Investment year	
				D	E
1	162.74	143.65	143.65	5	1
2	162.27	142.78	143.65	7	1
3	161.92	146.32	146.32	7	5
4	161.57	143.95	146.32	5	7
5	160.83	143.45	146.32	5	6
6	160.61	143.35	146.32	6	5

Table 10: Progression of bounds for Example 3

Solutions	ENPV ($ Million)	p(NPV<0) (%)	Std.Dev.(NPV) ($ Million)
Deterministic	143.43	8.2	82.35
Iteration 1	143.65	6.0	82.84
Iteration 2	142.78	6.0	81.64
Iteration 3	146.32	0.8	69.06
Iteration 4	143.95	0.8	67.20
Iteration 5	143.45	0.8	65.67
Iteration 6	143.35	0.8	66.28

Table 11: Comparison of solutions for Example 3 based on ENPV, “risk” and “robustness”

In this example too, we find that the proposed approach gives multiple alternative solutions that have ENPV similar to or greater than that of the deterministic solution. Table 11 compares the solutions obtained at the first 6 iterations of the proposed algorithm with the deterministic solution in terms of the ENPV, the probability of incurring a negative NPV, and the standard deviation in NPV. It can be seen that besides having similar or higher ENPV than the deterministic solution, these solutions are associated with significantly lower risk of leading to a negative NPV, and have less variance (more robust).

Regarding the question whether to invest in large uncertain fields or small certain fields, in this example the best solution obtained proposes investment in the certain fields in year 1, and postpones investment in the uncertain fields to later years. However, since the algorithm was terminated without cross-over of bounds, we cannot conclude that it is actually preferable to invest in the certain fields as opposed to the uncertain fields. To answer this question we fixed the investment time of both the uncertain fields to year 1 and solved the multistage stochastic program. The solution obtained had an ENPV = \$133.81 Million (< \$146.32 Million). Thus, in this case, it is actually preferable to invest in the smaller certain fields early and to postpone investment in the larger uncertain fields to later years.

From the results of Examples 1, 2 and 3, it is clear that the solution obtained using the proposed approach not only has significantly higher ENPV than the deterministic solution, but also involves much lower risk in terms of the probability of incurring a negative NPV. Moreover, the proposed approach gives alternative feasible solutions that may have slightly lower ENPVs, but involve significantly lower risk.

Also, in all examples presented above, the proposed approach found it preferable to invest in the certain fields during the early part of the project horizon, and postpone investment in the uncertain fields for later years. Finally, we present an example where this is not true.

Example 4

The superstructure with 7 fields (A-G), where fields D, E and G are uncertain, and one PP, is shown in Fig. 10. Investment and operation decisions for this offshore site have to be made for a horizon of 15 years (15 time periods). As in the previous examples, we consider three possible sizes and initial deliverabilities (low, medium and high with probabilities 0.3, 0.4 and 0.3 respectively) for each of the uncertain fields. Thus, the problem consists of 729 scenarios. The field data is shown in Table 12. Note that in this example too, fields D and E are larger than the certain fields C and F on the basis of mean values, but they have significant uncertainty associated with them.

The full space model of the big-M formulation of (SPM), with more than 165,000 binary

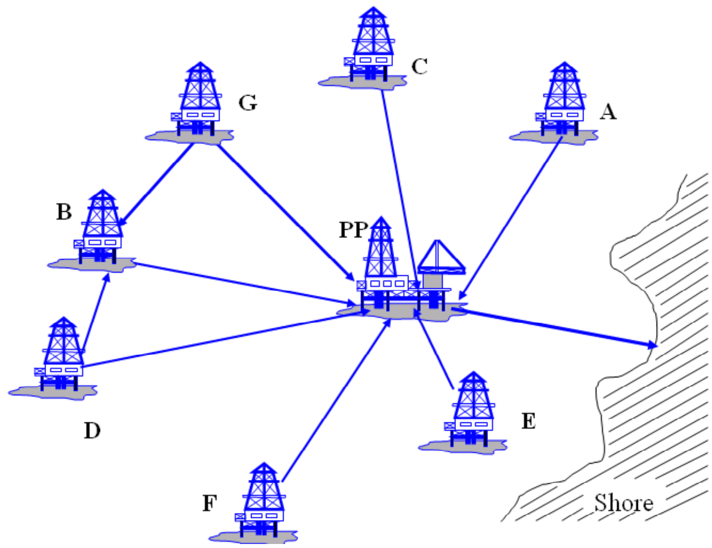


Figure 10: Superstructure for Example 4

variables, was intractable. The proposed approach, on the other hand, yielded a solution with ENPV = \$79.13 Million. The algorithm was terminated after four iterations (2.8 % gap), because the lower bound did not improve for three successive iterations. The total solution time for this example was 33,059 CPU seconds. Table 13 shows the progression of bounds, the ENPV of the feasible solution obtained at every iteration, and the years of investment for the uncertain fields used to obtain the solution.

It should be noted that, due to the large number of scenarios, the two-stage approximation models obtained as explained before, are very large. For this example, we used an alternative method of obtaining these approximations. Instead of relaxing non-anticipativity constraints for all stages after the 1^{st} stage, we assumed mean values for all uncertain parameters resolved in the 2^{nd} stage and later. For example, at iteration 1, fixing investment years for the uncertain fields from the solution of (EVP) gives us a three-stage problem. The uncertainty in fields D and E is resolved at the end of the 1^{st} stage, while uncertainty in field G is resolved at end of the 2^{nd} stage. A two-stage approximation to this three-stage problem is developed by assuming mean values for the sizes and initial deliverabilities of field G. The solution of this approximation is used to fix decisions in the 1^{st} stage of the three-stage problem. And the resulting 81 two-stage problems are then solved independently.

	A	B	C	D			E			F	G		
				Low	Med.	High	Low	Med.	High		Low	Med.	High
Size (BSCF)	440	412	250	230	290	350	230	280	330	250	150	200	250
Initial Deliverability (MSCF/D)	150	200	130	115	135	155	115	125	135	120	115	135	155

Table 12: Sizes and Initial deliverabilities for fields in Example 4

As shown in Table 14, both, the proposed solution (iteration 1) and the deterministic solution propose investment in the certain fields A and B, and the uncertain fields D and E, in year 1. However, the ENPV of the deterministic solution is only \$73.53 Million, compared to \$79.13 Million for the proposed solution. This is because, as shown in Table 15, compared to the deterministic solution, the proposed solution selects smaller capacities for WPs at the uncertain fields D and E. As a result, even if these fields have low sizes and low initial deliverabilities, their plateau length is 4 years. For the deterministic solution, on the other hand, the plateau lengths for the uncertain fields would be only 2 years if they have low sizes and low initial deliverabilities, in which case the investment in field F will have to be made earlier (see Table 14).

Iteration Number	UB (\$ Million)	Soln. obtained ENPV(\$ Million)	LB (\$ Million)	Investment year		
				D	E	G
1	85.31	79.13	79.13	1	1	10
2	82.47	74.49	79.13	5	1	10
3	82.08	74.07	79.13	1	5	10
4	81.44	70.22	79.13	7	1	10

Table 13: Progression of bounds for Example 4

Fig. 11 compares the deterministic and proposed solutions in terms of the cumulative probability distribution of their NPVs. The cumulative probability of incurring an NPV less than a target NPV is widely used as a measure of financial risk associated with that target NPV (Barbaro and Bagajewicz, 2002). Based on this definition, these curves demonstrate that in

Platforms	Investment Year	
	Proposed Solution	Deterministic Solution
A, B, D, E, PP	Year 1	
F	Year 5	Year 3 or 4 or 5
C	Year 7 or 8	Year 7 or 8
G	Year 10	Year 10
ENPV ($ Million)	79.13	73.53

Table 14: Proposed and deterministic solutions for Example 4

	Platform Capacity (MSCF/D)				
	A	B	D	E	PP
Deterministic Solution	81	117	76	76	350
Proposed Solution	101	117	66	66	350

Table 15: Comparison of proposed capacities for platforms to be installed in year 1: Example 4

the lower and medium NPV ranges, the proposed solution has significantly lower financial risk than the deterministic solution, while in the higher NPV ranges, the two solutions have similar performance.

Finally, we must emphasize that the proposed model and algorithm are not intended for one-time use in order to plan for the next 15 years. Instead, the model will be updated and resolved periodically as more information becomes available. This will serve to undo the effect of restricting the solution search to a sub-space of the feasible region, as proposed by our approach.

6 Conclusions

A stochastic programming approach for investment and operational planning of gas field developments under uncertainty in gas reserves has been presented. The uncertainty in the reserves of a gas field has been represented by uncertainty in the size and the initial deliverability of the field. This problem falls under the category of problems where the scenario tree depends on the project decisions. A novel stochastic programming model, where the decision-dependence of the scenario tree is represented by applying non-anticipativity constraints through disjunctions,

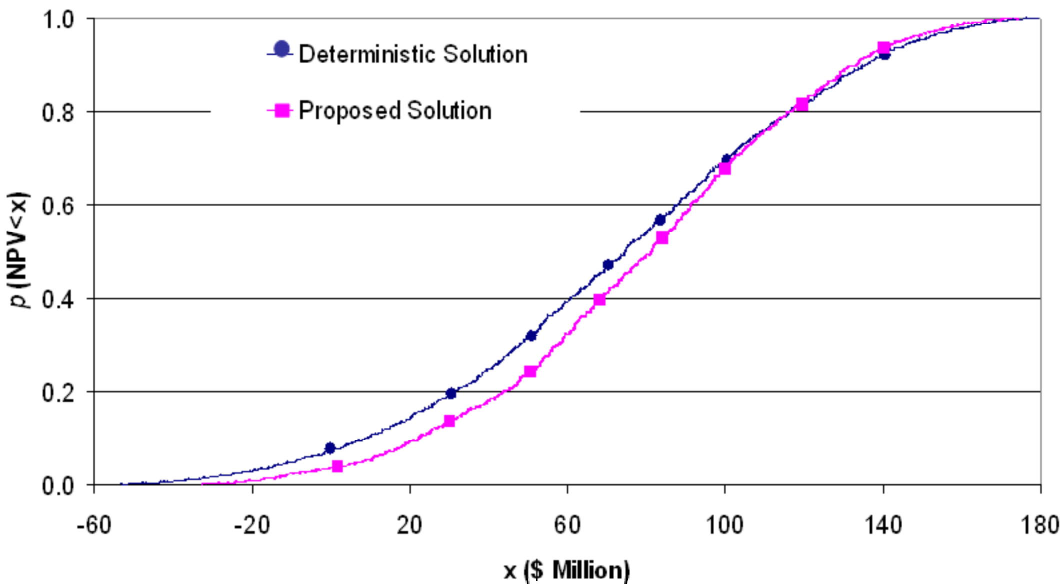


Figure 11: Cumulative probability distributions of NPV for the deterministic and proposed solutions in Example 4

has been proposed. A decomposition based approximation algorithm to address the exponential increase in size of this model with the number of uncertain fields has also been proposed.

This approach has been applied to a set of examples and the results show significant improvements in ENPV over those of solutions obtained using a deterministic approach. Furthermore, it has been demonstrated that the proposed approximation algorithm yields multiple solutions that not only have higher ENPV than solutions obtained by a deterministic approach, but are also associated with significantly lower risk in terms of leading to a negative NPV. For the smallest example presented (Example 1), the proposed approximation algorithm is shown to yield the exact optimal solution with more than one order of magnitude reduction in solution time as compared to the full space method. For all larger examples (Examples 2, 3 and 4), where full space models require up to 165,000 binary variables, the approximation algorithm is shown to yield “good” solutions in reasonable time.

Acknowledgment

The authors acknowledge financial support from ExxonMobil Upstream Research Company for this research. The authors also acknowledge valuable comments from Dr. Andrew Schaefer (University of Pittsburgh).

7 Nomenclature

Sets and Indices

τ, t = time periods, $t \in \{1, 2, \ldots, T\}$

WP = set of WPs

wp = well platform $wp \in WP$

PP = set of PPs

pp = production platform $pp \in PP$

F = set of fields

f = field $f \in F$

UF = set of uncertain fields

uf = uncertain field $uf \in UF$

CF = set of certain fields, $UC \cup CF = F$

cf = certain field $cf \in CF$

K = current iterations in algorithm

k = index used for previous iterations in algorithm

s,s' = scenario

$\mathcal{D}(s, s')$ = set of uncertain fields whose properties (size and initial deliverability) is s are not the same as those in s'. $\mathcal{D}(s, s') \subseteq UF$

Note: The indices f and wp can be used interchangeably since there is one WP per field

Subscript Combinations denoted by (.)

$(f) or (wp)$ = variable associated with field f or wp

(pp) = variable associated with production platform pp

(wp, wp') = variable associated with pipeline between wp and wp'

(wp, pp) = variable associated with pipeline between wp and pp

Note: All variables are defined for scenario s

Continuous Variables

$q_{f,t}^{prod,s}$ = Gas production rate at field f in period t

$q_{f,t}^{deliv,s}$ = Deliverability of field f in period t

$q_{f,t}^{cum,s}$ = Cumulative production from field f up to period t

$q_{(.),t}^{out,s}$ = Gas flow-rate out of structure in period t

$q_t^{shr,s}$ = Gas flow-rate at shore in period t

q_t^s = Vector representing all operation variables for period t

$Cap_{wp,t}^s$ = Capacity of well platform wp in period t

$Cap_{pp,t}^s$ = Capacity of production platform pp in period t

$Expan_{wp,t}^s$ = Expansion in capacity of well platform wp in period t

$Expan_{pp,t}^s$ = Expansion in capacity of production platform pp in period t

d_t^s = Vector representing all continuous investment variables for period t

Binary Variables

$b_{(.),t}^s$ = 1 if (.) is installed in period t, 0 otherwise

y_t^s = Vector of $b_{(.),t}$ for $f \in CF$, PP, pipe connections

Parameters

c_{1t} = Discounted costs corresponding to operation variables q_t^s

c_{2t} = Discounted costs corresponding to continuous investment variables d_t^s

c_{3t} = Discounted costs corresponding to binary investment variables y_t^s

$c_{4t,uf}$ = Discounted costs corresponding to binary investment variables $b_{uf,t}^s$ for uncertain field uf

$Size_f^s$ = Size of field f in scenario s

$InitDeliv_f^s$ = Initial deliverability of field f in scenario s

Δ_t = number of days in period t

$shrink$ = Shrinkage factor for flow in pipeline towards shore

$M_{(.)}$ = Upper bound on capacity of (.)

m = Number of realizations for each uncertain parameter

$\hat{b}_{uf,t,K}$ = Value of $b_{uf,t}$ in solution of (EVP_K). These values are used to fix the scenario tree while obtaining lower bounds at iteration K in algorithm

References

Balas, E. (1985) Disjunctive Programming and a Hierarchy of Relaxations for Discrete Optimization Problems. *SIAM Journal of Algebraic and Discrete Methods*,**v 6 n 3**, 466.

Barbaro, A.; Bagajewicz, M.J. (2002) Managing Financial Risk in Planning Under Uncertainty. Submitted for publication.

Barnes, R.J.; Linke, P.; Kokossis, A. (2002) Optimization of Oil-field Development Production Capacity. *European Symposium on Computer Aided Process Engineering*, **12**, 631.

Birge, J.R. (1997) Stochastic Programming Computation and Applications. *INFORMS Journal on Computing*, **v 9 n 2**, 111.

Birge, J.R.; Louveaux, F. (1997) Introduction to Stochastic Programming. Springer, New York, NY.

Dias, M.A.G.; Rocha, K.M.C. (1999) Petroleum Concessions with Extendible Options Using Mean Reversion with Jumps to model Oil Prices. *Proceedings of* 3^{rd} *Annual International Conference on Real Options*, Wassenaar/Leiden, Netherlands.

Dias, M.A.G. (2001) Selection of Alternatives of Investment in Information for Oil-field Development using Evolutionary Real Options Approach. *Proceedings of* 5^{th} *Annual International Conference on Real Options*, UCLA, Los Angeles.

ExxonMobil Financial and Operating Review, 2001.

Geoffrion, A.M. (1972) Generalized Benders' Decomposition. *Journal of Optimization Theory and Applications*, **v 10 n 4**, 237.

Grothey, A.; McKinnon, K. (2000) Decomposing the Optimization of a Gas Lifted Oil Well Network. Technical Report, MS 00-005, Department of Mathematics and Statistics, University of Edinburgh.

Haneveld W.K.K.; van der Vlerk, M.H. (1999) Stochastic Integer Programming: General Models and Algorithms. *Annals of Operations Research*, **85**, 39.

Haugen, K.K. (1996) A Stochastic Dynamic Programming Model for Scheduling of Offshore Petroleum Fields with Resource Uncertainty. *European Journal of Operational Research*, **88**, 88.

Higle, J.L.; Sen, S. (1991) Stochastic Decomposition: An Algorithm for Two-stage Linear Programs with Recourse. *Mathematics of Operations Research* **v 16 n 3**, 650.

ILOG CPLEX 7.0. *User's Manual*; ILOG Inc.

Ierapetritou, M.G.; Floudas, C.A.; Vasantharajan, S.; Cullick, A.S. (1998) Optimal Location of Vertical Wells: A Decomposition Approach. *AIChE Journal*, **45**, 844.

Iyer, R.R.; Grossmann, I.E.; Vasantharajan, S.; Cullick, A.S. (1998) Optimal Planning and Scheduling of Offshore Oil Field Infrastructure Investment and Operations. *Industrial and Engineering Chemistry Research*, **37**, 1380.

Jonsbraten, T.W. (1998a) Oil-field optimization under Price Uncertainty. *Journal of the Operational Research Society*, **49**, 811.

Jonsbraten, T.W. (1998b) Optimal Selection and Sequencing of Oil Wells under Reservoir Uncertainty. *Phd Thesis*, Department of Business Administration, Stavanger College, Norway.

Jonsbraten, T.W. (1998c) Nash Equilibrium and Bargaining in an Oil Reservoir Management Game. *Ph. D. Thesis*, Department of Business Administration, Stavanger College, Norway.

Jonsbraten, T.W. (1998d) Optimization Models for Petroleum Field Exploitation, Ph. D. Thesis, Department of Business Administration, Stavanger College, Norway.

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.

Jornsten, K.O. (1992) Sequencing Offshore Oil and Gas Fields under Uncertainty. *European Journal of Operational Research*, **58**, 191.

Kall, P.; Wallace, S.W. (1994) Stochastic Programming, Wiley, New York.

Kosmidis, V.D.; Perkins, J.D.; Pistikopoulos, E.N. (2002) A Mixed Integer Optimization Strategy for Integrated Gas/Oil Production. *European Symposium on Computer Aided Process Engineering*, **12**, 697.

Lee, S.; Grossmann, I.E. (2000) New Algorithms for Generalized Disjunctive Programming. *Computers and Chemical Engineering*, **v 24 n 9-10**, 2125.

Lin, X.; Floudas, C.A. (2002) A Novel Continuous-Time Modeling and Optimization Framework for Well Platform Planning Problems. *Optimization and Engineering*, in press.

Lund, M.W. (2000) Valuing Flexibility in Offshore Petroleum Projects. *Annals of Operations Research*, **99**, 325.

Meister, B.; Clark, J.M.C.; Shah, N. (1996) Optimization of Oil-field Exploitation Under Uncertainty. *Computers and Chemical Engineering* , **20 ser. B**, S1251.

Ortiz-Gomez, A.; Rico-Ramirez, V.; Hernandez-Castro, S. (2002) Mixed-integer Multi-period Model for the Planning of Oil-field Production. *Computers and Chemical Engineering* , **v 26 n 4-5**, 703.

Pflug G. (1990) Online Optimization of Simulated Markov Processes. *Mathematics of Operations Research*, **v 15 n 3**, 381.

Raman, R.; Grossmann, I.E. (1994) Modeling and Computational Techniques for Logic Based Integer Programming. *Computers and Chemical Engineering*, **v 18 n 7**, 563.

Sahinidis, N.V. (2003) Optimization under Uncertainty: State-of-the-Art and Opportunities. To appear in *Proceedings of FOCAPO, 2003*.

Stougie, L.; van der Vlerk, M.H. (1997) Stochastic Integer Programming. In *Annotated Bibliographies in Combinatorial Optimization*, Dell'Amico, Maffioli and Martello (Eds.), (1997), Wiley, Chapter 9, 127-141.

Turkay, M.; Grossmann, I.E. (1996) Disjunctive Optimization Techniques for the Optimization of Process Systems with Discontinuous Investment Costs-Multiple Size Regions. *Industrial and Engineering Chemistry Research*, **v 35 n 8**, 2611.

Van den Heever, S.A.; Grossmann, I.E. (2000) An Iterative Aggregation/Disaggregation Approach for the Solution of a Mixed Integer Nonlinear Oil-field Infrastructure Planning Model. *Industrial and Engineering Chemistry Research*, **v 39 n 6**, 1955.

Van den Heever, S.A.; Grossmann, I.E. (2001) A Lagrangean Decomposition Heuristic for the Design and Planning of Offshore Hydrocarbon Field Infrastructures with Complex Economic Objectives. *Industrial and Engineering Chemistry Research*, **v 40 n 13**, 2857.

Van Slyke, R.; Wets, R.J.-B. (1969) L-shaped Linear Programs with Applications to Optimal Control and Stochastic Programming. *SIAM Journal on Applied Mathematics*, **17**, 638.

Vecchietti, A.; Grossmann, I.E. (1999) LOGMIP: A Disjunctive 0-1 Non-linear Optimizer for Process Synthesis. *Computers and Chemical Engineering*, **v 23 n 4-5**, 555.

A Standard Stochastic Programming Model for problems with Project Exogenous Uncertainty

(SP) is a 'standard' stochastic programming model for a problem with time periods $t = 1, 2, \ldots, T$ involving project exogenous uncertainty.

$$\text{(SP)} \qquad \text{Max} \quad \sum_s p^s c^{sT} X^s \tag{35}$$

$$A^s X^s \leq a^s \qquad \forall s \tag{36}$$

$$X \in \mathcal{N}_{\mathcal{S}} \tag{37}$$

Variables $X^s = (x_1^s, x_2^s, \ldots, x_T^s)$ represent decision variables for scenario s, where x_t^s represents decision variables for time period t in scenario s. The objective is to maximize the expected profit (35). (36) represents constraints that govern the feasible space for decisions associated with a particular scenario. These include single period and period linking constraints typical in multiperiod problems.

Decisions for different scenarios are linked together by "non-anticipativity" or "implementability" constraints (37). At time t, scenarios s and s' are said to be "indistinguishable" if, based on the information obtained until t, one cannot infer which one of scenarios s and s' will occur in the future. Non-anticipativity requires that if scenarios s and s' are indistinguishable in time period t, then decisions for time period t have to be the same in scenarios s and s'. Thus, non-anticipativity constraints state that at any time decisions cannot be based on knowledge that will be revealed in the future. $\mathcal{N}_{\mathcal{S}}$ represents the set of non-anticipative solutions for the given scenario tree $\mathcal{S}$.

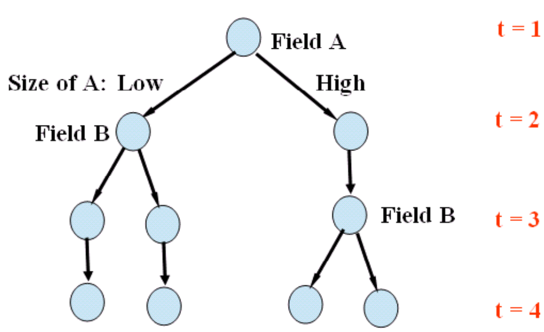
For the scenario tree presented in Fig. 2, all scenarios are indistinguishable from each other at $t = 1$. At $t = 2$, s_1 is indistinguishable from s_2 while s_3 is indistinguishable from s_4. Thus, the non-anticipativity constraints corresponding to the scenario tree in Fig. 2 are as follows:

$$\begin{aligned} x_1^s &= x_1^{s'} && \forall(s, s') \\ x_2^{s1} &= x_2^{s2} && \text{and} \\ x_2^{s3} &= x_2^{s4} && \end{aligned}$$

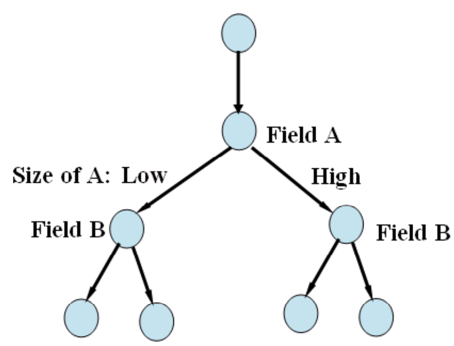
As can be seen, the "non-anticipativity" or "implementability" constraints to be applied, and hence the 'standard' stochastic programming model, depends directly on the structure of the scenario tree.

B Decision-dependent Scenario Trees in a problem with Project Endogenous Uncertainty

We consider a small example with two uncertain fields, A and B, and four time periods. For sake of simplicity, we assume that only the sizes of these fields are uncertain and that the size of each field has two possible realizations, H(High) and L(Low).



(a) Investment policy I



(b) Investment policy II

Figure 12: Decision dependent scenario trees

We consider two investment policies. Investment policy I suggests that we install a WP at field A in t=1. If the size of A is found to be Low, then we install a WP at field B in t=2. On the other hand, if the size of field A is found to be High, we postpone the installation of the WP at field B to t=3. Investment policy II suggests that the WP at field A be installed in t=2 and the WP at field B be installed in t=3, irrespective of the size of field A.

Since the uncertainty in a field is resolved only when a WP is installed at that field, the scenario trees for the two investment policies are as shown in Fig. 12. Thus, it is clear that the shape of the scenario tree, and hence the non-anticipativity constraints to be applied, depend upon when we invest in the uncertain fields. Another important inference that can be drawn from this example is the fact that, depending on the timing of investments at the uncertain fields, the problem may (or may not) assume the form of a multistage stochastic programming problem.

C Constraints represented by (2)-(6)

Reservoir Constraints

$$q_{f,t}^{prod,s} \leq q_{f,t}^{deliv,s} \qquad \forall(f,t,s) \qquad (38)$$

$$\frac{q_{f,t}^{deliv,s}}{InitDeliv_f^s} + \frac{q_{f,t}^{cum,s}}{Size_f^s} = 1 \qquad \forall(f,t,s) \qquad (39)$$

$$q_{f,t}^{cum,s} = \sum_{\tau=1}^{t} q_{f,\tau}^{prod,s} \Delta_\tau \qquad \forall(f,t,s) \qquad (40)$$

$$q_{f,T}^{cum,s} \leq Size_f^s \qquad \forall(f,s) \qquad (41)$$

Surface Constraints

$$q_{wp,t}^{out,s} = q_{wp,t}^{prod,s} + \sum_{wp'} q_{wp',wp,t}^{out,s} \qquad \forall(wp,t,s) \qquad (42)$$

$$q_{wp,t}^{out,s} = \sum_{wp'} q_{wp,wp',t}^{out,s} + \sum_{pp} q_{wp,pp,t}^{out,s} \qquad \forall(wp,t,s) \qquad (43)$$

$$q_{pp,t}^{out,s} = \sum_{wp} q_{wp,pp,t}^{out,s} \qquad \forall(pp,t,s) \qquad (44)$$

$$q_t^{shr,s} = (1 - shrink) \sum_{pp} q_{pp,t}^{out,s} \qquad \forall(t,s) \qquad (45)$$

Capacity Constraints

$$q_{wp,t}^{out,s} \leq Cap_{wp,t}^s \qquad \forall(wp,t,s) \qquad (46)$$

$$q_{pp,t}^{out,s} \leq Cap_{pp,t}^s \qquad \forall(pp,t,s) \qquad (47)$$

$$Cap_{wp,t}^s = Cap_{wp,t-1}^s + Expan_{wp,t}^s \qquad \forall(wp,t,s) \qquad (48)$$

$$Cap_{pp,t}^s = Cap_{pp,t-1}^s + Expan_{pp,t}^s \qquad \forall(pp,t,s) \qquad (49)$$

Logic Constraints

$$\sum_t b_{(.),t}^s \leq 1 \qquad \forall((.),s) \qquad (50)$$

$$b^s_{wp,t} = \sum_{wp'} b^s_{wp,wp',t} + \sum_{pp} b^s_{wp,pp,t} \qquad \forall(wp,t,s) \tag{51}$$

$$b^s_{wp,wp',t} \leq \sum_{\tau=1}^{t} b^s_{wp',\tau} \qquad \forall(wp,wp',t,s) \tag{52}$$

$$b^s_{wp,pp,t} \leq \sum_{\tau=1}^{t} b^s_{pp,\tau} \qquad \forall(wp,pp,t,s) \tag{53}$$

$$b^s_{wp,wp',t} + b^s_{wp',wp,t} \leq 1 \qquad \forall(wp,wp',t,s) \tag{54}$$

Big-M Constraints

$$q^{out,s}_{(.),t} \leq M_{(.)} \sum_{\tau=1}^{t} b^s_{(.),\tau} \qquad \forall((.),t,s) \tag{55}$$

$$Expan^s_{wp,t} \leq M_{wp} b^s_{wp,t} \qquad \forall(wp,t,s) \tag{56}$$

$$Expan^s_{pp,t} \leq M_{pp} b^s_{pp,t} \qquad \forall(pp,t,s) \tag{57}$$

D Property of Constraint (8)

Property:

$$\bigwedge_{uf\,\in\,\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b^s_{uf,\tau}\right)\right] \Leftrightarrow \bigwedge_{uf\,\in\,\mathcal{D}(s',s)} \left[\wedge_{\tau=1}^{t}\left(\neg b^{s'}_{uf,\tau}\right)\right] \qquad \forall(s,s',t)$$

Proof:

We first prove that $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b^s_{uf,\tau}\right)\right] \Rightarrow \bigwedge_{uf\in\mathcal{D}(s',s)} \left[\wedge_{\tau=1}^{t}\left(\neg b^{s'}_{uf,\tau}\right)\right]$

Suppose $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b^s_{uf,\tau}\right)\right] = TRUE$

We have,

$$\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t}\left(\neg b^s_{uf,\tau}\right)\right] \Rightarrow \bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t-1}\left(\neg b^s_{uf,\tau}\right)\right]$$

Similarly,

$$\begin{aligned}
\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t-1}\left(\neg b^s_{uf,\tau}\right)\right] &\Rightarrow \bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t-2}\left(\neg b^s_{uf,\tau}\right)\right]\\
\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t-2}\left(\neg b^s_{uf,\tau}\right)\right] &\Rightarrow \bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t-3}\left(\neg b^s_{uf,\tau}\right)\right]\\
\vdots \qquad &\ \vdots \qquad \vdots\\
\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{2}\left(\neg b^s_{uf,\tau}\right)\right] &\Rightarrow \bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{1}\left(\neg b^s_{uf,\tau}\right)\right]
\end{aligned}$$

Since, $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b^s_{uf,\tau}\right)\right] = TRUE$

$$\Rightarrow \bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b^{s}_{uf,\tau}\right)\right] = \bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t-1}\left(\neg b^{s}_{uf,\tau}\right)\right] = \ldots = \bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{1}\left(\neg b^{s}_{uf,\tau}\right)\right] = TRUE$$

From (8), it follows that

$$Z_t^{s,s'} = Z_{t-1}^{s,s'} = \ldots = Z_1^{s,s'} = TRUE \tag{58}$$

From (7) we have,

$$\begin{array}{llllll} Z_{t-1}^{s,s'} & \Rightarrow & b^{s}_{uf,t} & = & b^{s'}_{uf,t} & \forall uf \\ Z_{t-2}^{s,s'} & \Rightarrow & b^{s}_{uf,t-1} & = & b^{s'}_{uf,t-1} & \forall uf \\ Z_{t-3}^{s,s'} & \Rightarrow & b^{s}_{uf,t-2} & = & b^{s'}_{uf,t-2} & \forall uf \\ \vdots & & \vdots & & \vdots & \vdots \\ Z_{1}^{s,s'} & \Rightarrow & b^{s}_{uf,2} & = & b^{s'}_{uf,2} & \forall uf \\ & & & & & \\ \text{Also,} & & b^{s}_{uf,1} & = & b^{s'}_{uf,1} & \forall uf \qquad \text{(from (9))} \end{array}$$

$$\Rightarrow \left[\bigwedge_{\tau=1}^{t}\left(\neg b^{s}_{uf,\tau}\right)\right] = \left[\bigwedge_{\tau=1}^{t}\left(\neg b^{s'}_{uf,\tau}\right)\right] \qquad \forall uf$$

$$\Rightarrow \bigwedge_{uf} \left[\bigwedge_{\tau=1}^{t}\left(\neg b^{s}_{uf,\tau}\right)\right] = \bigwedge_{uf} \left[\bigwedge_{\tau=1}^{t}\left(\neg b^{s'}_{uf,\tau}\right)\right]$$

In particular, $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t}\left(\neg b^{s}_{uf,\tau}\right)\right] = \bigwedge_{uf\in\mathcal{D}(s',s)} \left[\bigwedge_{\tau=1}^{t}\left(\neg b^{s'}_{uf,\tau}\right)\right] \qquad (\mathcal{D}(s,s') = \mathcal{D}(s',s))$

Since $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t}\left(\neg b^{s}_{uf,\tau}\right)\right] = TRUE$,

We get $\bigwedge_{uf\in\mathcal{D}(s',s)} \left[\bigwedge_{\tau=1}^{t}\left(\neg b^{s'}_{uf,\tau}\right)\right] = TRUE$

Thus $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b^{s}_{uf,\tau}\right)\right] \Rightarrow \bigwedge_{uf\in\mathcal{D}(s',s)} \left[\wedge_{\tau=1}^{t}\left(\neg b^{s'}_{uf,\tau}\right)\right]$

Converse: $\bigwedge_{uf\in\mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b^{s}_{uf,\tau}\right)\right] \Leftarrow \bigwedge_{uf\in\mathcal{D}(s',s)} \left[\wedge_{\tau=1}^{t}\left(\neg b^{s'}_{uf,\tau}\right)\right]$
Follows by the same logic as above using the symmetry of the condition and model (SPM).

Hence,

$$\bigwedge_{uf \in \mathcal{D}(s,s')} \left[\wedge_{\tau=1}^{t}\left(\neg b_{uf,\tau}^{s}\right)\right] \Leftrightarrow \bigwedge_{uf \in \mathcal{D}(s',s)} \left[\wedge_{\tau=1}^{t}\left(\neg b_{uf,\tau}^{s'}\right)\right]$$

E Application of Non-anticipativity constraints: Illustrative Example

We use the example with two uncertain fields A and B, that was introduced in Appendix B. Here again, we assume that only the sizes of these fields are uncertain. The scenarios in the problem are represented by $s = (\xi_{size,A}, \xi_{size,B})$, where $\xi_{size,A}$ and $\xi_{size,B}$ represent the sizes of fields A and B respectively. Two realizations H(High), L(Low) for the size of each uncertain field are considered. The four scenarios in the problem are represented by s_1, s_2, s_3 and s_4 where, $s_1 = (L, L); s_2 = (L, H); s_3 = (H, L); s_4 = (H, H)$. The second column of Table 16 lists $\mathcal{D}(.,.)$ for each pair of scenarios.

For scenario s_1, there are four possible cases regarding whether or not WPs have been installed at fields A and B in or before time period t. We derive and explain the application of non-anticipativity constraints for each of these cases.

Case I: WP installed at field B but not at field A

$$\begin{array}{lrcl} \text{Thus,} & \left[\wedge_{\tau=1}^{t}\left(\neg b_{A,\tau}^{s_1}\right)\right] & = & True \\ \text{whereas} & \left[\wedge_{\tau=1}^{t}\left(\neg b_{B,\tau}^{s_1}\right)\right] & = & False \end{array}$$

Since, $\mathcal{D}(s_1, s_3) = \{A\}$,

$$\Rightarrow Z_t^{s_1,s_3} = \bigwedge_{uf \in \mathcal{D}(s_1,s_3)} \left[\wedge_{\tau=1}^{t}\left(\neg b_{uf,\tau}^{s_1}\right)\right] = True.$$

Thus from (7), non-anticipativity requires that, for scenarios s_1 and s_3, operation decisions in t be the same and investment decisions in $t+1$ be the same. This is obvious since s_1 and s_3 differ only in size of field A and since a WP has not been installed at field A, the uncertainty in A has not been resolved by t. Thus, s_1 and s_3 are indistinguishable after investments have been made in period t. On the other hand, since $\left[\wedge_{\tau=1}^{t}\left(\neg b_{B,\tau}^{s_1}\right)\right] = False$, and $B \in \mathcal{D}(s_1, s_2), \mathcal{D}(s_1, s_4)$,

$$\Rightarrow Z_t^{s_1,s_2} = Z_t^{s_1,s_4} = False$$

(s,s')	$\mathcal{D}(s,s')$	$Z_t^{s,s'}$ if in s,t			
		$\neg A \wedge B$	$A \wedge \neg B$	$\neg A \wedge \neg B$	$A \wedge B$
(s_1,s_2)	B	$False$	$True$	$True$	$False$
(s_1,s_3)	A	$True$	$False$	$True$	$False$
(s_1,s_4)	A,B	$False$	$False$	$True$	$False$
(s_2,s_3)	A,B	$False$	$False$	$True$	$False$
(s_2,s_4)	A	$True$	$False$	$True$	$False$
(s_3,s_4)	B	$False$	$True$	$True$	$False$

Table 16: Values for set $\mathcal{D}(s,s')$ and Boolean variables for the two-uncertain field example

Case II: WP installed at field A but not at field B

Similar to the case above, now $Z_t^{s_1,s_2} = True$ and hence non-anticipativity constraints link operation decisions for scenarios s_1 and s_2 in period t and investment decisions of the scenarios in period $t+1$. Also, $Z_t^{s_1,s_3} = Z_t^{s_1,s_4} = False$

Case III: WPs not installed at either of fields A and B

$\left[\wedge_{\tau=1}^{t}\left(\neg b_{A,\tau}^{s_1}\right)\right] = \left[\wedge_{\tau=1}^{t}\left(\neg b_{B,\tau}^{s_1}\right)\right] = True$

Thus,

$Z_t^{s_1,s_2} = Z_t^{s_1,s_3} = Z_t^{s_1,s_4} = True$

Hence, by (7), non-anticipativity will link decisions for all scenarios. This is obvious because no uncertainty has been resolved yet, and hence all scenarios are mutually indistinguishable.

Case IV: WPs installed at fields A and B

$\left[\wedge_{\tau=1}^{t}\left(\neg b_{A,\tau}^{s_1}\right)\right] = \left[\wedge_{\tau=1}^{t}\left(\neg b_{B,\tau}^{s_1}\right)\right] = False$

Thus,

$Z_t^{s_1,s_2} = Z_t^{s_1,s_3} = Z_t^{s_1,s_4} = False$

Thus no non-anticipativity constraints apply, which is obvious because WPs at both uncertain fields have been installed and hence all uncertainty has been resolved.

Logic variables $Z_t^{s,s'}$ for the above four cases are shown in Table 16. By $\neg A \wedge B$, we represent the case that a WP has been installed at field B but not at A in or before period t. Note that because non-anticipativity constraints are applied for scenario pair (s, s') only if $s < s'$, logic variables for only six of the sixteen $(= 4 \times 4)$ possible scenario ordered pairs are listed in the table. In this example the scenarios have been assumed to be ordered as $s_1 < s_2 < s_3 < s_4$.

F Models

F.1 Model $(CSPM_K)$

$$(CSPM_K) \qquad \text{Max} \quad \sum_s p^s \left[\sum_t (c_{1t}^T q_t^s + c_{2t}^T d_t^s + c_{3t}^T y_t^s) \right] + \sum_t \sum_{uf} c_{4t,uf}^T b_{uf,t} \tag{59}$$

s.t. (14), (15), (16), (17), (18), (19), (20), (21), (22)

$$\sum_{(uf,t)\in B_k} b_{uf,t} - \sum_{(uf,t)\in N_k} b_{uf,t} \leq |B_k| - 1 \qquad \forall k = 1, 2, \ldots, K-1 \tag{60}$$

$$q_t^s,\ d_t^s \geq 0, \qquad b_{uf,t} \in \{0,1\},\ y_t^s \in \{0,1\}^{dim(y)}, \qquad Z_t^{s,s'} \in \{True, False\}$$

F.2 Model $(RCSPM_K)$

$$(RCSPM_K) \qquad \text{Max} \quad \sum_s p^s \left[\sum_t (c_{1t}^T q_t^s + c_{2t}^T d_t^s + c_{3t}^T y_t^s + \sum_{uf} c_{4t,uf}^T b_{uf,t}^s) \right] \tag{61}$$

s.t. (24), (25), (26), (27), (28)

$$\sum_{(uf,t)\in B_k} b_{uf,t}^s - \sum_{(uf,t)\in N_k} b_{uf,t}^s \leq |B_k| - 1 \qquad \forall s; k = 1, 2, \ldots, K-1 \tag{62}$$

$$q_t^s,\ d_t^s \geq 0, \qquad b_{uf,t}^s \in \{0,1\},\ y_t^s \in \{0,1\}^{dim(y)}$$

F.3 Model (EVP_K)

$$(EVP_K) \qquad \text{Max} \quad \left[\sum_t (c_{1t}^T q_t + c_{2t}^T d_t + c_{3t}^T y_t)\right] + \sum_t \sum_{uf} c_{4t,uf}^T b_{uf,t} \tag{63}$$

s.t. (30), (31), (32), (33), (34)

$$\sum_{(uf,t)\in B_k} b_{uf,t} - \sum_{(uf,t)\in N_k} b_{uf,t} \leq |B_k| - 1 \quad \forall k = 1, 2, \ldots, K-1 \tag{64}$$

$$q_t,\ d_t \geq 0, \qquad b_{uf,t} \in \{0,1\},\ y_t \in \{0,1\}^{dim(y)}$$

G Property of algorithm if problem involves one uncertain field

Property

$|UF| = 1 \Rightarrow (SPM) \equiv (CSPM)$

Proof:

Suppose $UF = \{uf^*\}$.

We show that for this case, (2)-(11) $\Rightarrow b_{uf^*,t}^{s} = b_{uf^*,t}^{s'} \qquad \forall (t, s, s')$.

The equivalence of (SPM) and (CSPM) follows as a result.

We proceed using mathematical induction.

From (9), we have that

$$b_{uf^*,1}^{s} = b_{uf^*,1}^{s'} \qquad \forall (s, s') \tag{65}$$

Suppose that the following holds for scenario pair (s_1, s_2)

$$\begin{array}{rcl} b_{uf^*,1}^{s_1} & = & b_{uf^*,1}^{s_2} \\ b_{uf^*,2}^{s_1} & = & b_{uf^*,2}^{s_2} \\ \vdots & = & \vdots \\ b_{uf^*,t^*}^{s_1} & = & b_{uf^*,t^*}^{s_2} \end{array}$$

$\Rightarrow \left[\wedge_{\tau=1}^{t^*}\left(\neg b_{uf^*,\tau}^{s_1}\right)\right] = \left[\wedge_{\tau=1}^{t^*}\left(\neg b_{uf^*,\tau}^{s_2}\right)\right]$

Since uf^* is the only uncertain field, therefore $\mathcal{D}(s_1, s_2) = \{uf^*\}$

$\Rightarrow \bigwedge_{uf\in\mathcal{D}(s_1,s_2)} \left[\wedge_{\tau=1}^{t^*}\left(\neg b_{uf,\tau}^{s_1}\right)\right] = \bigwedge_{uf\in\mathcal{D}(s_1,s_2)} \left[\wedge_{\tau=1}^{t^*}\left(\neg b_{uf,\tau}^{s_2}\right)\right]$

Case I: $\left[\wedge_{\tau=1}^{t^*}\left(\neg b_{uf^*,\tau}^{s_1}\right)\right] = TRUE$

Since $\mathcal{D}(s_1, s_2) = \{uf^*\}$, $\Rightarrow \bigwedge_{uf \in \mathcal{D}(s_1,s_2)} \left[\wedge_{\tau=1}^{t^*}\left(\neg b_{uf,\tau}^{s_1}\right)\right] = TRUE$
(8) $\Rightarrow Z_{t^*}^{s_1,s_2} = TRUE$
Thus (7) $\Rightarrow b_{uf^*,t^*+1}^{s_1} = b_{uf^*,t^*+1}^{s_2}$
Hence, using our induction hypothesis, we get

$$b_{uf^*,t}^{s_1} = b_{uf^*,t}^{s_2} \qquad \forall t$$

In general

$$b_{uf^*,t}^{s} = b_{uf^*,t}^{s'} \qquad \forall (t, s, s')$$

Case II: $\left[\wedge_{\tau=1}^{t^*}\left(\neg b_{uf^*,\tau}^{s_1}\right)\right] = FALSE$

$\Rightarrow \bigwedge_{uf \in \mathcal{D}(s_1,s_2)} \left[\wedge_{\tau=1}^{t^*}\left(\neg b_{uf,\tau}^{s_1}\right)\right] = FALSE$

$\Rightarrow \exists \tau^*$ such that $1 \leq \tau^* \leq t^*$ and $b_{uf^*,\tau^*}^{s_1} = 1$

Since $1 \leq \tau^* \leq t^*$, our induction hypothesis implies that

$$b_{uf^*,\tau^*}^{s_1} = b_{uf^*,\tau^*}^{s_2}$$

$\Rightarrow b_{uf^*,\tau^*}^{s_2} = 1$
But from (5), we have

$$\sum_{\tau} b_{uf^*,\tau}^{s_1} \leq 1$$

and

$$\sum_{\tau} b_{uf^*,\tau}^{s_2} \leq 1$$

Since $b_{uf^*,\tau^*}^{s_1} = b_{uf^*,\tau^*}^{s_2} = 1$
$\Rightarrow b_{uf^*,t^*+1}^{s_1} = b_{uf^*,t^*+1}^{s_2} = 0$
Hence, by induction, we get

$$b_{uf^*,t}^{s_1} = b_{uf^*,t}^{s_2} \qquad \forall t$$

Again, in general

$$b^s_{uf^*,t} = b^{s'}_{uf^*,t} \qquad \forall (t, s, s')$$

Thus, if $|UF| = 1$, then (2)-(11) $\Rightarrow b^s_{uf^*,t} = b^{s'}_{uf^*,t} \qquad \forall (t, s, s')$.

Therefore $|UF| = 1 \Rightarrow$ (SPM) $\equiv$ (CSPM)