Application of Markov Decision Processes to Mine Optimisation: A Real Option Approach

by

Louis Archambeault

Department of Mining, Metals and Materials Engineering

McGill University, Montreal

August 2006

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Engineering

Your file Votre référence
ISBN: 978-0-494-32578-0
Our file Notre référence
ISBN: 978-0-494-32578-0

Acknowledgements

I am greatly indebted to my thesis supervisor, Prof. Michel L. Bilodeau, for his guidance and continuous support throughout the course of this research; his comments and editing have been a great help.

I extend my gratitude to Prof. Roussos Dimitrakopoulos for introducing me to the concept of real options and for his expertise in geological simulation methods.

I would like to express recognition to Profs Joëlle Pineau and Doina Precup for initiating me to Markov decision processes and helping me in the early stages of development of the model.

Special thanks to Regis Sabbadin from the Institut National de la Recherche Agronomique for making modification to their toolbox in order to meet particular specifications required for applying MDPs to mining problems.

I dedicate this thesis to my parents who believe in me and give me their unconditional love and support in all of my endeavours.

English Abstract

This thesis describes preliminary research on the application of Markov Decision Processes (MDPs) to the optimisation of mine scheduling in an uncertain environment. The MDP framework is a novel approach to scheduling in a mining operation and option valuation. The task of scheduling in mining operations is dependent on the availability of models that permit the representation of some of the key stochastic properties of the environment, such as grade and price uncertainty. The tools used to model these processes are respectively sequential Gaussian simulation and Geometric Brownian motion. Three cases of increasing size are used to illustrate the results of the model and demonstrate its suitability to mine scheduling and option valuation. The computational efficiencies of solving an MDP formulation by Policy Iteration and Value Iteration are compared. The impact of the discount rate on the optimal policy is assessed. To determine the value of one or several options, an optimal policy without options is generated and valued. Then, the exercise is repeated with the relevant options to value (e.g., production rate, cut-off grade and time of mine closure). By comparing the values obtained in both cases, the financial benefit of having operational flexibility is determined, thus yielding the option value. A full size case study is conducted to validate the applicability of MDPs to real mining projects.

Sommaire

La présente thèse fait état d'un projet de recherche préliminaire visant l'application du Processus de Décision de Markov (PDM) à l'optimisation du calendrier d'extraction d'une mine dans un environnement incertain. L'approche préconisée est innovatrice relative au problème d'optimalisation des opérations minières et au calcul de la valeur d'une option selon la théorie des options. L'application du PDM se base sur des modèles permettant de simuler les caractéristiques aléatoires de certaines variables environnementales d'un projet minier, soit la teneur du minerai et le cours des prix. Les moyens utilisés pour simuler ces variables sont respectivement, la simulation séquentielle de Gauss et le mouvement géométrique de Brown. Trois cas de taille croissante sont utilisés pour illustrer les résultats du modèle et démontrer sa convenance dans l'optimisation du calendrier d'extraction d'une mine et le calcul de la valeur d'une option. L'efficacité de deux méthodes de résolution d'un PDM, soit l'itération de la politique et l'itération sur la valeur, est comparée. On analyse l'impact du taux l'actualisation sur la politique optimale. Pour déterminer la valeur d'une ou plusieurs options, on obtient la politique optimale de production sans options. Ensuite, on répète l'exercice en prenant en compte les options à évaluer (e.g., taux de production, teneur de coupure et date de fermeture de la mine.) En comparant les valeurs obtenues dans les deux cas, on démontre le bénéfice financier engendré par la flexibilité opérationnelle dans une opération minière, ce qui fait ainsi ressortir la valeur de l'option. Une étude est faite sur un cas réel afin de valider la mis an application du PDM aux projets miniers.

Table of Content

List of Figures

List of Tables

Chapter 1

Introduction

Problem Definition and Overview

Mining projects are characterized by high levels of uncertainty in ore grade estimation and by volatile fluctuations in commodity prices. Adjusting production parameters when working in such uncertain conditions is not a simple task. The aim of this thesis is to study the application of Markov Decision Processes (MDPs) to the task of mine production optimisation and Real Option Valuation (ROV). Using these tools, the value of having a robust mine design that can withstand estimation errors can be measured, thus providing a tool to justify the extra costs incurred in creating that robust design. It is believed that with simulations and MDPs, parameters such as production rate, cut-off grade and mine closure can be optimally adjusted in relation to commodity price fluctuations and uncertainties in reserve grade.

The development of ROV methods offers great insight on the potential value of operational flexibility and the benefits gained from generating robust mine designs. Yet, the mining industry is lagging on applying ROV concepts in project valuation practices. Traditional best practices prescribe that in order to maximise the value of a project, it is important to delay costs as much as possible and use "just in time" planning. However, using this approach is not very flexible and does not allow mine operators to react to market changes by increasing production when commodity prices are high. Many mines are currently faced with this problem. Due to the recent rapid increase in commodity prices, mining operations are attempting to increase production to meet market demands. Yet, because of the "just in time" approach, there is insufficient accessible ore to effectively increase production. Monkhouse (2005) discusses this problem in an example of an open pit mine, but the issue of ore accessibility applies to underground mines as well. Mine managers are starting to acknowledge the potential value gained by having accessible ore. This allows increased production when prices are high, thus resulting in selling a larger portion of the mineral products at a higher price. Clearly, pre-stripping to expose ore generates additional costs. Managers are thus faced with a non-trivial question: do the potential benefits gained from having exposed ore outweigh the additional cost? MPDs and ROV provide an effective tool to answer this question.

MDPs are reinforcement learning algorithms and, in essence, reinforcement learning works as follows. A virtual probabilistic environment is generated in which the learning agent, i.e. a modeled decision maker, is asked to make decisions. At the end of each iteration ("life"), the agent evaluates, in hindsight, the quality of its decisions. After a large number of iterations, the agent will have learned the decisions that will maximise his reward in all possible situations. The set of decisions selected to maximise the reward is called an optimal policy.

Using the MDP approach, we are accepting the fact that we cannot predict what our production parameters will be in the future. Instead, we provide an agent with a feasible range of production parameters and let it learn the optimal policy. Using this policy, the value of the project can easily be computed, thus providing a more representative distribution of project values. Furthermore, the optimal policy can serve as a guideline for managers once production has started. Incorporating option value during the valuation process could be misleading if there is insufficient information available for the planners to exploit the additional value during production. Therefore, an advantage of using MDPs to value real options is that not only does it provide the value of an option, but it also provides the policy that yields the maximum benefit from that option. This in turn provides the means of effectively exploiting the value of operating flexibility.

By incorporating the option to close down an operation, it is also possible to evaluate the financial risk associated with premature mine closure. This information is very valuable for long-term mine planning and leads to a substantial increase in project value.

In their book, Sutton and Barto demonstrate that an MDP is an implementation of Reinforcement Learning.

> We consider all of the work in optimal control also to be, in a sense, work in reinforcement learning. We define reinforcement learning as any effective way of solving reinforcement learning problems, and it is now clear that these problems are closely related to optimal control problems, particularly those formulated as MDPs. Accordingly, we must consider the solution methods of optimal control such as dynamic programming, also to be reinforcement learning methods. Of course, almost all of these methods require complete knowledge of the system to be controlled, and for this reason it feels a little unnatural to say that they are part of reinforcement *learning*. On the other hand, many dynamic programming methods are incremental and iterative. Like learning methods, they gradually reach the correct answer through successive approximation.
>
> \- Sutton and Barto (1998)

Thesis Organisation

In Chapter 2, previous attempts to incorporate the value of managerial flexibility in mine project valuation are reviewed. Furthermore, we survey cases of successful implementation of MPDs to optimisation problems. In Chapter 3 we illustrate the benefits of including managerial flexibility in project valuation and describe the real option approach. In Chapter 4, the Markov Decision Process is explained and the different algorithms available to solve MDPs are presented. Particularities associated with implementing MDPs in mining projects are mentioned as well. In Chapter 5, case studies are conducted on projects of different sizes. A small case study is used to explain how to read a policy and demonstrate the impact of discounting on the optimal policy. A medium case study is used to compare the computational requirements of different solving algorithms, and a large-scale study is conducted to measure the applicability of the MDP approach to real size mine scheduling optimisation. In Chapter 6, limitations are discussed, recommendations are offered and conclusions are drawn.

Chapter 2

Literature Review

Real Options

Interest in real options stems from the Nobel Prize winning work of Black and Scholes (1973), in which they demonstrated an analytical solution to valuing stock options and corporate liabilities. Brennan and Schwartz (1985) were the first to apply a real-options approach to mining investments. Using complex partial differential equations, the authors find the value of a project in which the option of a temporary shut-down of mine production is included. The decision to shut down with later reopening of the mine is based on market conditions associated with the commodity produced. Unfortunately, in order to maintain tractability, a few very limiting assumptions must be made. To a financial expert, these assumptions may seem reasonable. However, it is clear to any mining practitioner that these assumptions greatly affect the validity of the model.

The most detrimental assumption made by the authors relates to the ore reserves. It is assumed that reserves are homogenous and quantities are exactly known. As reserve block models are generated based on incomplete information, it follows that the uncertainty of the reserve estimate is quite high. In reality, the uncertainty associated with the reserves has an equal, if not greater, impact on the variability of the project's value than the volatility in commodity prices. The relevance of modeling fluctuations in commodity prices is questionable when larger sources of uncertainty such as grade variability are ignored.

Furthermore, the options that are valued are difficult to apply in the mining industry because there are substantial overhead costs related to closing and reopening a mine. Also, environmental regulations make it very complicated for mining companies to stop and then resume operations due to reclamation requirements. Therefore, the pertinence of valuing such options is questionable. Armstrong and Galli (1997) discuss the issues of temporary mine closure and the complications of reopening.

Although the techniques presented in this paper have important limitations, Brennan and Schwartz were the first to recognize the potential of option valuation in mining and their work lead the way to many other significant contributions in mine valuation.

Cortazar et al. (2000) attempt to resolve some of the shortcomings of the original work presented by Brennan Schwartz (1985) by incorporating the "volatility" of ore reserves in their valuation model. However, the reserve uncertainty is not modeled according to drill hole data and is not conditioned on the orebody variability. The authors arbitrarily select a Brownian motion to model the uncertainty, independently of measured geological uncertainty provided by drill hole data. Although their approach is a step forward, the fact that the geological particularities of the orebody are ignored leads to questionable results.

Samis (2000) examines options that are more pertinent to a mining framework than that of previous authors. As such, Samis proposes to value the option to close or expand sections of a mine rather than the option of closing/reopening the complete operation.

Samis (2000) improves on previous work by acknowledging the heterogeneous nature of the mining environment, thereby treating the reserves as zones of high grade and low grade. Also, an attempt to include geological uncertainty is made. Further reading in Samis (2000) provides more details concerning the modeling of geological uncertainty. The geological model used to consider the quality of the mineralization is a discrete probability distribution based on a geologist's intuition. Although this is an improvement on previous work, it will be shown in later sections of this report that more appropriate geostatistical methods are available to represent the variability of the reserves (Farrelly, 2002).

In his Ph.D. thesis, Samis (2000) states that geological risk is unsystematic, and therefore, is diversifiable. Yet he includes geological risk in his model. This presents some theoretical inconsistencies with other works. Monkhouse and Yeates (2005) state: "… (diversifiable) risks are unpriced (investors are indifferent about bearing them, e.g. geological uncertainty)." Monkhouse and Yeats (2005) conclude that geological uncertainty does not affect the mine plan. However, in his case study, Samis (2000) states:

> "Geological uncertainty is treated as a diversifiable (unsystematic) risk since it is uncorrelated with market uncertainty. Its impact on value is determined by calculating the value expectation with respect to the joint grade probability distribution…"

Therefore, Samis does accept the fact that geological risk is unsystematic, yet contrary to Monkhouse and Yeats (2005) he includes geological risk in his model.

I believe that geological risk is not entirely diversifiable and that it must be taken into consideration when valuing a mining venture.

Monkhouse and Yeates (2005) rightfully criticise current best practices in the mining industry and demonstrate the value of generating robust mine designs as opposed to deterministically optimised designs. They mention that current best practices advance revenues forward in time and delay costs as much as possible to maximise the Net Present Value (NPV) of the project, thus creating a very rigid design that does not behave well under conditions of uncertainty. In their Orebody Uncertainty section, the authors mention that taking into consideration orebody uncertainty in the mine design increases the NPV of the project. This is confirmed by Dimitrakopoulos, Farrelly and Godoy (2002), Godoy and Dimitrakopoulos (2004), and Ramazan and Dimitrakopoulos (2005). However, in their conclusion, Monkhouse and Yeates state: "... counter-intuitively, it was argued that the risk of geological uncertainty did not affect the mine plan...", and therefore, they did not include orebody uncertainty in their model.

This author believes that taking grade uncertainty into consideration is crucial to a successful robust mine design and thus, must be addressed in the design optimisation process for the following reasons:

- Monkhouse and Yeates (2005) introduce the financial concept of diversifiable and non-diversifiable risk in order to justify the disregard of geological risk. The distinction between diversifiable and non-diversifiable risk can be found in any introductory text to Corporate Finance (e.g. Ross et al., 1999). Diversifiable risk can be ignored during the valuation process if there are sufficient non-correlated assets to form a well diversified portfolio. However, it is unlikely that a mining company could create a diversified portfolio by combining various operating mines. Furthermore, it is wrong to assume that a well diversified portfolio of orebodies can be created using single deterministic kriging estimates, because it is impossible to measure the correlation between each orebody. Furthermore, If the correlation between orebody block models could be measured, some correlation would most likely be noticeable due to the fact that the same estimation method (kriging) is used, thus generating a systematic bias over all the orebodies.

- Kriging estimates are unbiased over the entire mineral deposit. Therefore, it is reasonable to assume that the estimation errors for all blocks cancel each other out as stated by Monkhouse and Yeates (2005). However, an orebody is selectively mined, i.e. the low grade blocks are left behind as waste, thus yielding a bias in the estimated average grade of the mined blocks. This bias is caused by the smoothing effect of kriging and the fact that only the low grade blocks are left behind. Furthermore, kriging does not yield a "truly central estimate" for cash revenue. The transformation from ore grade to cash revenue is non-linear, thus causing the average of the distribution of estimated values to differ from the theoretical mean value (Dimitrakopoulos et al., 2002). This systematic error amongst all of the estimates renders the risk non-diversifiable.

- Furthermore, Monkhouse and Yeates (2005) state: "... (diversifiable) risks are unpriced (investors are indifferent about bearing them, e.g. geological uncertainty)." This statement is debatable because exploration/mining companies invest substantial amounts of money in exploration and drilling in order to reduce geological uncertainties. At the extreme, it would be ludicrous to assume that a mining company would start mining a mineral deposit at a random location without investing in exploration to first delineate the deposit. Goria (2004) and Bilodeau and Mackenzie (1989) discuss the value of further delineation.

- It can be argued that a mining company would reduce systematic ore grade risk by creating a diversified portfolio of assets outside of the mining industry, thus changing the systematic risk of ore estimation into non-systematic risk. However, this is not common practice within the industry and it is hard to envision mining companies investing in pork belies to reduce the risk inherent in ore reserve estimation. Furthermore, even if a company opted to have a diversified set of assets outside of mining investments, the risk could not be completely diversified when the concept of changing cut-off grades is included in the design/optimisation process. Varying the cut-off grade will result in reserve recalculation. Therefore, through the interim of cut-off grade variations, the error in reserve estimation will follow the market, rendering one component of geological risk systematic.

It is now apparent that geological risk is systematic and should be incorporated in the optimisation process and in the calculation of the discount rate used to compute the NPV. Although there are some limitations to the techniques presented by Monkhouse and Yeates (2005), it is encouraging to see that some of the industry leaders are acknowledging the potential benefits of designing robust mine plans rather than using conventional optimisation techniques.

Goria (2004) presents the first real option valuation approach that uses conditional simulations to model geological risk. Although the main interest of Goria's work is the development of block model simulations using a bayesian approach to conditional simulations, an interesting contribution is to have introduced methods to value the option of conducting extra delineation prior to the start of site development.

One drawback to the methodology presented by Goria (2004) is that out of a total of 100 simulated block models, only three are selected for the valuation, i.e., the P90, P50 and P10, based on the distribution of average grades. It has been shown by Dimitrakopoulos et al. (2002) that a central grade estimate will not yield a central cash estimate, as the transformation from grade to cash is non-linear. Therefore, the block models selected will not necessarily generate the P90, P50 and P10 of the NPV distribution curve. Therefore, a distribution based only three values is only marginally better than a single point estimate, as it is impossible to determine where the three points are located within the distribution.

Markov Decision Processes

History

Sutton and Barto (1998) offer an overview of the history of Reinforcement Learning in their book entitled "Reinforcement Learning: An Introduction." Puterman (1994) also provides some historical background in his book "Markov Decision Processes, Discrete Stochastic Dynamic Programming." Much of the information presented in this section is based on these two publications.

Reinforcement learning as we know it today was born when its three major components came together in the late 1980s. These three components consist of: optimal control using dynamic programming; learning by trial and error; and temporal-difference methods. A brief overview of these threads is presented below with the exception of temporal-difference methods, as it is not relevant to the issues addressed in this thesis.

The Second World War delayed and complicated many of the publications of the 1940s, a time during which much work was being done on stochastic sequential decision-making. Twisted timelines make it difficult to truly assign credit to whom it is due. Sutton and Barto (1998) mention that much of the groundwork for "optimal control" was done by Bellman (1954). However, Lucien Le Cam (1990) mentions the following:

> "Massé had developed a lot of the mathematics about programming for the future. What had become known in the country (the United States) as dynamic programming invented by Richard Bellman, was very much alive in Massé's works, long before Bellman had a go at it."

An accurate timeline would be difficult to construct, as mentioned by Heyman and Sobel (1984):

> "The modern foundations were laid between 1949 and 1953 by the people who spent at least part of the period as staff members at the Rand Corporation in Santa Monica, California. Dates of actual publication are not reliable guides to the order in which ideas were discovered during this period."

As this is not an exercise in "optimal control" chronology, an accurate timeline is irrelevant. Rather, a list of major early contributors to the field, as presented by Puterman (1994), is as follows:

"... Arrow, Bellman, Blackwell, Dvoretsky, Girschik, Isaacs, Karlin, Kiefer, LaSalle, Robbins, Shapley and Wolfwitz. Their work on games (Bellman and Blackwell, 1949; Bellman and LaSalle, 1949; Shapley, 1953), stochastic inventory models (Arrow, Harris, and Marschak, 1951; Dvoretsky, Kiefer and Wolfowtiz, 1952), pursuit problems (Isaacs, 1955, 1965) and sequential statistical problems (Arrow, Blackwell, and Girshick, 1949; Robbins, 1952; Kiefer, 1953) laid the groundwork for subsequent developments."

Trial and error learning was first developed in the field of psychology. Thorndike (1911) was the first to express that an action followed by a good outcome will tend to be reselected in the future. Thorndike called this the "law of effect". The essence of trial and error learning, in which there is an element of *search* and an element of *memory*, is thus captured.

Minsky (1954) was the first to discuss the computational models of reinforcement learning. By the 1960s, the term reinforcement learning was widely used in engineering literature. In the mid-50s, Farely and Clark's interests shifted to supervised learning. However, the transition was not clear, and many still considered their more recent work to be reinforcement learning. In the 60s and

70s, actual reinforcement learning work became rare and supervised learning was more prominent, although the distinction between the two was not very clear. A system called STeLLA was developed in New Zealand by Andrea (1963). In his work, the use of trial and error learning was always maintained. A simple trial and error system called MENACE was developed by Michie (1963) for learning to play tic-tac-toe. With Chambers, Michie (1968) would go on to develop GLEE and BOXES. BOXES was applied to the famous pole-balancing problem. The pole-balancing problem was taken from Widrow and Smith's (1964) earlier work. This is one of the first reinforcement learning problems with incomplete information. Further work has been done in reinforcement learning with incomplete knowledge. However, the remainder of the development of trial and error learning is beyond the scope of this work as the Markov decision process (MDP) method implemented in this study is based on a complete model of the environment.

Successful applications of Markov decision processes

"Branching out from operations research roots in the 1950s, MDP models have gained recognition in such diverse fields as ecology, economics and communications engineering." (Puterman, 1994). In 1994, MDPs where already successfully used in a variety of fields. In his book, Puterman (1994) presents a few examples of the application of MDPs: inventory management, bus engine replacement, highway pavement maintenance, communication models and mate desertion in Cooper's hawks. Sutton and Barto (1998) present applications for elevator dispatching, dynamic channel allocation and job shop scheduling. Also, a series of more recent applications which appeared in various technical papers are described below.

Inventory Management (Puterman, 1994)

MDPs have been successfully applied to inventory management problems in which an optimal reordering policy is required to maintain customer satisfaction while minimizing ordering and storage costs. The uncertainty in this model is inherent in the stochastic purchasing behaviour of the consumer. Puterman describes an inventory management problem using local Canadian Tire dealerships. A total of 21 dealerships are supplied by a main warehouse. There are fixed costs associated with placing an order and variable costs associated with storing merchandise at the dealerships. Management requires that at least 97.5 % of the demand be satisfied from the supply on hand. Assuming that inventory is measured once a week, we can now describe the problem as an MDP. The current state is determined by the inventory level at the time it is measured. The actions correspond to the ordering of new parts following the inventory review. The transition probabilities are dependent on the number of parts ordered and the random consumer demand.

Bus Engine Replacement (Puterman, 1994)

A mechanic performing maintenance inspections on busses must decide whether their engines need to be replaced. The replacement cost of the engine is a large fixed cost that consists of the purchase price of the new engine and the labour associated with its installation. The probabilistic component is the possibility of engine failure, in which case significant costs would be incurred for towing, repairs and loss of goodwill. Changing the engine to soon results in unnecessary costs while engine failure is detrimental to the transporter's reputation. An optimal policy is required to minimise the long-run cost of maintaining a bus fleet. The current state is represented by bus mileage at the time of inspection, and inspections are conducted on a monthly basis. The optimal policy would indicate the mileage at which an engine needs to be replaced. A multitude of algorithms are suitable to solve the bus engine replacement problem.

Highway Pavement Maintenance (Puterman, 1994)

The Arizona Department of Transportation (ADOT), in collaboration with Woodward-Clyde Consultants from San Francisco, developed an MDP model for its pavement management system to allocate limited funds to ensure that the quality of Arizona roadways was preserved. An estimated $101 million was saved by using the system, while no decline in road quality was observed. The Arizona roadway system was divided into 7400 one-mile sections, each having 120 possible states based on roughness, cracking, the change in cracking conditions from the previous year, the time since the last maintenance and the nature of the maintenance operation. Extensive data was required for this model to accurately characterize the state of the roads, and statistical regressions were used to determine the transition probabilities. A total of 17 different maintenance actions were available, ranging in price from $0 to $6.30 per square yard. The model is a constrained average reward MDP and can be solved using different unichain algorithms as presented in Puterman (1994).

Elevator Dispatching (Sutton and Barto, 1998)

The elevator problem presented by Sutton and Barto addresses a ten floor building serviced by 4 elevators. The objective is to find a policy to reduce the average squared waiting time of passengers. A particularity of this problem is the number of possible states estimated at 10^{22}. A problem with such a large number of states is intractable using standard dynamic programming techniques. The model was tested on a detailed elevator simulator, where the arrival of passengers is a discrete stochastic event. The authors applied a modified one-step Q-learning algorithm to solve the model.

MDP Model for Capacity Expansion and Allocation (Bhatnagar et al., 1999)

The authors present a model to provide decision support for operational decisions concerning additional capacity and production type conversion. A finite-horizon model is used. The state is represented by the current production capacity and the type of units being produced. The actions available relate to increasing production or changing the type of units being produced. The stochastic components of the model are the operating costs and the probability of successful production increase. Dynamic programming software called SYSCODE is used to solve the model.

MDP Model for Airline Meal Catering (Goto, 2000)

At first glance, the problem of meal catering seems relatively trivial: simply prepare the amount of meals for the number of passengers on the flight. However, the meals must be prepared well before the final boarding. Therefore, the ultimate seating on the plane in unknown at the time the meals are prepared. As boarding time approaches, the caterer has more information concerning the final boarding and adjustments can be made to the order. In the case presented by the author, the caterer has five opportunities to modify the size of the order. The caterer incurs losses when food is over supplied, and losses in goodwill are incurred when passengers do not receive their meals. A policy is required to establish modifications to the order size as departure time approaches. The author reports that the application of the optimal policy resulted in a 17% reduction in average catering cost per flight and a 33% reduction in short-catered flights.

Using MDPs to Solve a Portfolio Allocation Problem (Bookstaber, 2005)

Portfolio allocation is a problem that can be formulated easily using an MDP. As the movement of stock prices is well documented, transition probabilities can be computed. In the author's example, a fixed amount of funds is allocated to the portfolio. Therefore, the objective is to determine the weights of the portfolio assets in a way to maximise expected return. For small state-space problems, the author used Value Iteration to find the optimal policy. For larger state-space problems, off-policy Q-learning and Sarsa (λ) where used as approximation methods.

This overview on the application of MDPs is in no way exhaustive. MDPs can be applied to a very wide range of problems. Many less relevant models have intentionally been left out. The examples discussed above are sufficient to hypothesise that, based on past successful applications of MDPs and their similarities to mine scheduling problems, they can be applied successfully to real option valuation and mine schedule optimisation.

Chapter 3

Managerial Flexibility and Real Option Value

Deterministic Valuation versus Option Valuation

An increasing body of work documents the shortcomings of standard deterministic Net Present Value (NPV) analysis as a tool for valuing investment projects. The main critique against this approach is its inability to incorporate managerial flexibility in the valuation process, which ultimately undervalues the project. Therefore, an alternative method is desirable. In his book "Real Options", Trigeorgis (1996) states:

> "A *call option* is an option [...] that gives the right, with no obligation, to acquire the underlying asset by paying a pre-specified price on or before a given maturity. Similarly, a *put option* gives the right to sell the underlying asset and receive the exercise price. The asymmetry deriving from having the right but not the obligation to exercise the option lies at the heart of the option's value."

From the definition above it can be argued that real assets are similar to options. Investing in a project provides the right to produce a certain product but with no obligation to do so. By considering projects as portfolios of real options, we are incorporating the value of managerial flexibility, thereby attempting to minimise the downside risk and maximise the upside potential of a project. Mayers (1984) demonstrates the similarities between financial assets and real assets, and argues that if the traditional NPV approach fails for financial options, it should not be applied to real options, i.e. those associated with real assets. Therefore, if we accept that projects can be considered as a portfolio of real options, option valuation techniques should be used to determine their value. In this chapter, we will demonstrate that managerial flexibility does in fact increase project value, using an example from Dixit and Pindick (1994). Once it is demonstrated that options add value to projects, we look at some of the options available in mining projects. Then, the real option valuation approach is introduced, followed by the methods available to solve the valuation problem.

Value in Options

Dixit and Pindyck (1994) present a simple example demonstrating how managerial flexibility adds value to a project. In their example, which has been modified for clarity, the option to postpone a risky investment is evaluated. Management is faced with an investment decision in which the revenue (that is a function of price) is uncertain.

They assume an irreversible capital investment of $1600, with no operating costs. In the first year, the revenue is $200. In the following years, the revenue will either be $100 or $300 with probabilities of 0.5 each, and will remain constant thereafter. A graphical representation of the possible project cash flows is shown in Figure 1.

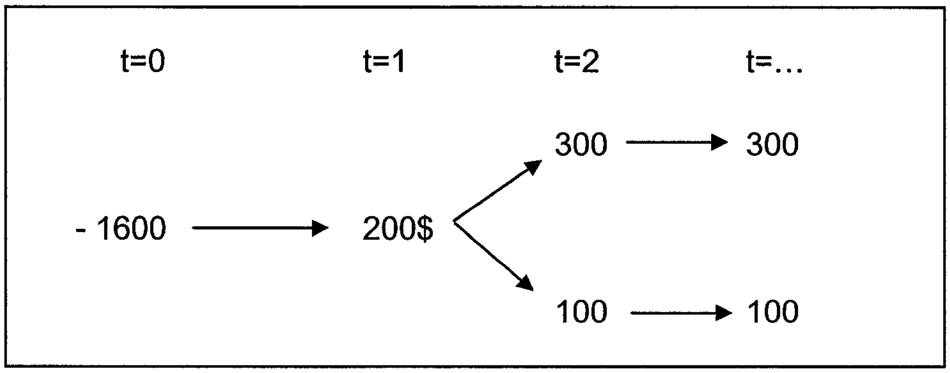


Figure 1: Possible Project Cash Flows

Management has the option to invest today or to wait until the uncertainty in revenue is resolved in one year's time and only invest if conditions are favourable. The objective of this exercise is to determine whether the option to postpone the investment actually has value.

If management decides to invest today, the investment is done under uncertainty and the expected revenue is the weighted-average of the two possible scenarios, i.e. $200. However, if management decides to delay the investment, the investment would only be made at the beginning of the second year if the revenue is favourable. The expected NPVs today of the two scenarios described above using a discount rate of 10% are:

Invest today

$$\text{NPV} = -1600 + \sum_{t=1}^{\infty} \frac{200}{(1.1)^t} = -1600 + 2000 = \$400 \tag{3.1}$$

Delay investment

$$\text{NPV} = (0.5)\left[\frac{-1600}{1.1} + \sum_{t=2}^{\infty} \frac{300}{(1.1)^t}\right] = \frac{700}{1.1} = \$636 \tag{3.2}$$

A comparison of (3.1) and (3.2) shows that the option to delay the investment has a value of \$636 minus \$400, i.e. \$236. In this example, the issue of discounting and risk adjustment is ignored but will be treated later in the chapter.

Figure 2 is a graphical representation of option value relative to commodity price uncertainty. The straight diagonal line represents the value of a hypothetical project without the option to delay the investment. The break-even commodity price is located where the project value line crosses the x-axis. The curved line, above the project value line, represents the option value. The difference between the two lines is therefore the option premium. The option premium is at a maximum when the project NPV is equal to 0 (referred to as *at the money*).

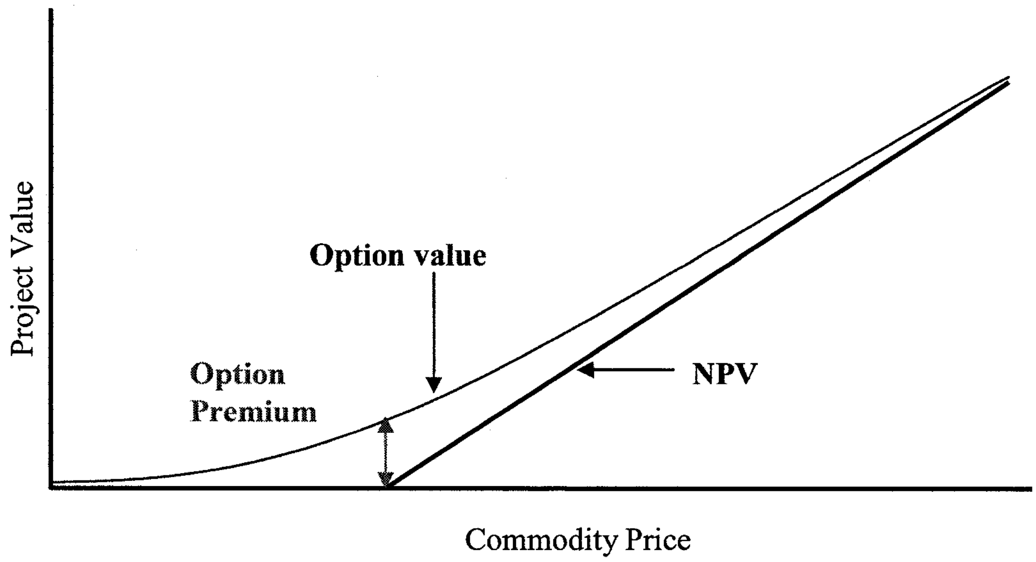


Figure 2: Option Value

As the price of the commodity increases and the project becomes more and more lucrative, the option to delay project start-up loses value because starting the project earlier generates earlier revenues. At the other end of the spectrum, when the value of the project is very negative (*out of the money*), the option of keeping the project has little value because the probability of it becoming lucrative is very low.

In the example above, we have seen that the option to delay the start of production increased the value of the project. This example is a very simplified case in which only one decision must be made. In reality, there are many options that can increase project value. Trigeorgis (1996) lists a series of common real options and Armstrong and Gali (1997) provide a list of options specific to mining projects. These are shown in Table 1.

Table 1: Common Project Options

Trigoergis (1996)	**Armstrong and Gali (1997)**
Option to defer investment	Option to expand reserves
Option to default during construction	Production rate options
Option to expand	Cut-off grades options
Option to contract	
Option to stop/start	
Option to abandon	
Option to switch use	
Corporate growth options	

Real Options in Mining

Contrary to Brennan and Schwartz's (1985) assumption, it is not generally accepted that a mining operation has the option to stop and start production based on price conditions. Due to labour relations and environmental regulation requirements, there are substantial overhead costs related to mothballing a mining operation. The options that are generally available in mining projects are: the option to delay the start of production, exploration and delineation options, mining sequence options, resource allocation options, capacity utilization options, cut-off grade options, stockpiling options, pit expansion options and permanent closure options. A description of these options is provided below.

- During periods of low commodity prices, marginal exploration projects are often temporarily shelved to be reconsidered when prices improve. The feasibility of these projects can be re-evaluated at that time.

- Goria (2004) and Bilodeau and Mackenzie (1979) demonstrate the value gained from performing delineation. Delineation drilling is costly but will lead to more informed investment decisions and better mine design resulting in a more efficient operation. Also, exploration drilling can lead to the discovery of new mineralized zones and increased project value.

- It is nearly impossible to establish a detailed mining sequence on a long-term basis. Modifications to the mining sequence are made on a weekly and sometimes daily bases. Adjustments are required to optimise mill recovery through effective blending and feed control.

- Mining operation flexibility is limited by the resources available on site. Managers must allocate these resources effectively. Trucking capacity must be matched to shovel capacities. Shovels must be assigned to production or development areas of the pit to ensure there is sufficient exposed ore to meet mill requirements.

- When commodity prices are high, mining operations increase production to achieve higher profits. Higher production rates during periods of high prices will result in the sale of more mineral products at a higher price, thus increasing the overall value of the project.

- Mine operators often adjust the cut-off grade according to commodity price fluctuations. When prices decrease, the mine operator increases the cut-off grade. This practice is commonly known as high-grading. High-grading insures that the revenue generated from mining will remain above further processing costs.

- Stockpiles are often used for blending purposes and for cut-off grade optimisation. When commodity prices decrease, the low grade material is stockpiled and saved. When commodity prices increase, the low grade stockpiles can be blended into the mill feed. Stockpiling therefore permits the processing of a higher proportion of the mineral deposit, thus extending the life of the mine.

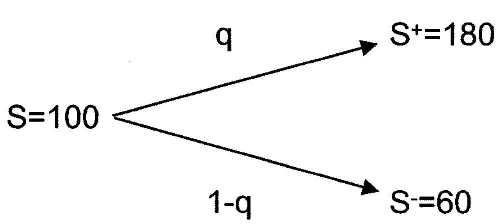
- In the event that ore reserves were initially underestimated and substantial mineralization would be left behind given the original mine design, the size of the ultimate pit can be re-evaluated during the life of the mine, thereby increasing its life.

- When it is no longer beneficial to continue production, the mine can be permanently closed. Early mine closure could be exercised in the event of unanticipated lower ore reserves or if commodity prices decrease drastically over an extended period of time.

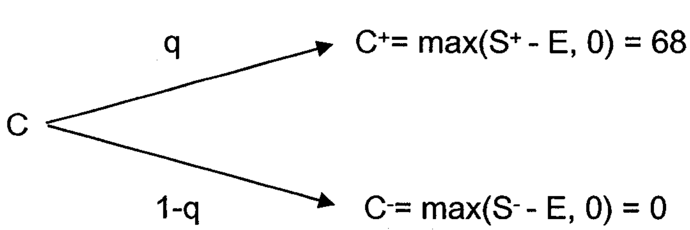
Option Pricing Theory

In order to evaluate an option, a risk free portfolio that replicates the payoff of the option must be created. This portfolio is constructed by purchasing N shares of the underlying asset at a price S, and by borrowing an amount B at the risk-free rate. The following example taken from Trigeorgis (1996) demonstrates a simple example of option valuation.

Assume an option to purchase an asset with a current (time t) market price S of \$100. In one time period from now (time t+1), the price may increase to \$180 (S^+) with a probability of q, or decrease to \$60 (S^-) with a probability of (1-q) as depicted below.



Assume that the exercise price E of the option is \$112. The value C of the option at time t+1 could either be \$68 (C^+) or \$0 (C^-), as shown below.



An effective replicating portfolio should have the same value as the option, whether the price of the stock goes up or down. The value of the option can be expressed as $C \approx N \bullet S - (1+r)\, B$, in which r is the risk-free rate. Therefore, the value of the portfolio is:

$$\begin{tikzcd} & N \bullet S^{+} - (1+r)\, B \\ N \bullet S - B \arrow[ur, "q"] \arrow[dr, "1-q"'] & \\ & N \bullet S^{-} - (1+r)\, B \end{tikzcd}$$

The value of the portfolio must be identical to the option value for each state.

$$N \bullet S^{+} - (1+r)\, B = C^{+} \tag{3.3}$$

and

$$N \bullet S^{-} - (1+r)\, B = C^{-} \tag{3.4}$$

Using the two equations above as well as a risk-free rate of 8%, it is possible to find the values of N and B.

$$N = \frac{C^+ - C^-}{S^+ - S^-} = \frac{68-0}{180-60} = 0.56 \tag{3.5}$$

and

$$B = \frac{S^- C^+ - S^+ C^-}{\left(S^+ - S^-\right)\left(1+r\right)} = \frac{N \bullet S^- - C^-}{1+r} = \frac{0.56 \times 60 - 0}{1.08} = \$31 \tag{3.6}$$

In order to compute the value of the option, the risk neutral transition probabilities are required. These can be obtained with the following equation:

$$p = \frac{\left(1+r\right)S - S^-}{S^+ - S^-} = \frac{1.08\,(100) - 60}{180 - 60} = 0.4 \tag{3.7}$$

Therefore, the value of the option is:

$$C = \frac{p \bullet C^+ + \left(1-p\right)C^-}{1+r} = \frac{0.4\,(68) + 0.6\,(0)}{1.08} = \$25 \tag{3.8}$$

There are numerous ways of determining the value of real options. The simplest is to use the closed analytical solution to a set of differential equations as presented by Black and Scholes (1973). However, this method does not lend itself well to real options. An alternative is the finite difference method. This method is the discrete analog of the derivative and can be solved implicitly or explicitly. For more details on finite differences, refer to Duffy (2006). Options can also be valued using binomial trees as presented by Trigeorgis (1991). In recent years, Monte-Carlo simulations and the least-squares method has been gaining popularity. This method is presented by Longstaff and Schwartz (2001). They consider the trade-off between the immediate value of exercising the option and the expected value of continuation. The expected value of continuation is determined by regression on a constant, X and X^2, in which X is the current simulated market value of the underlying asset.

In this dissertation, a novel method of optimisation in open pit mine scheduling is presented that permits a real option approach to mine valuation. The optimisation method implemented is called a Markov Decision Process.

Note on discounting

The risk-neutral transition probabilities are an important component in option valuation. Samis et al. (2006) present a series of examples demonstrating the importance of risk adjustment at the source and how future prices can be used to calculate the risk neutral transition probabilities.

When options are included in project valuation, risk is greatly reduced and in some cases completely eliminated. In these cases, it is reasonable to use a risk-free discount rate. However, in the case of real options in the mining industry, the risk inherent in the project is often not completely eliminated by the inclusion of options and risk adjustment of prices at the source. Therefore, project valuation may still require a risk component in the discount rate. As this work is not an exercise in capital budgeting, the issues related to the use of an appropriate discount rate are not addressed. However, it is important to note that an MDP can easily be adapted for risk adjustment at the source or for a risk-adjusted discount rate.

In **Error! Reference source not found.**, Laughton (2003) provides a graphical representation of the different valuation methods available, ranked according to the level of uncertainty/flexibility available in the model and the method used to account for risk. From top to bottom, there is an increase in modeling of uncertainty and accountability for risk. On the left, risk is accounted for by using a risk adjusted discount rate while on the right, risk adjustment is applied at the source. Depending on the risk adjustment method used, an MDP can be located on either side of the top area of Laughton's chart.

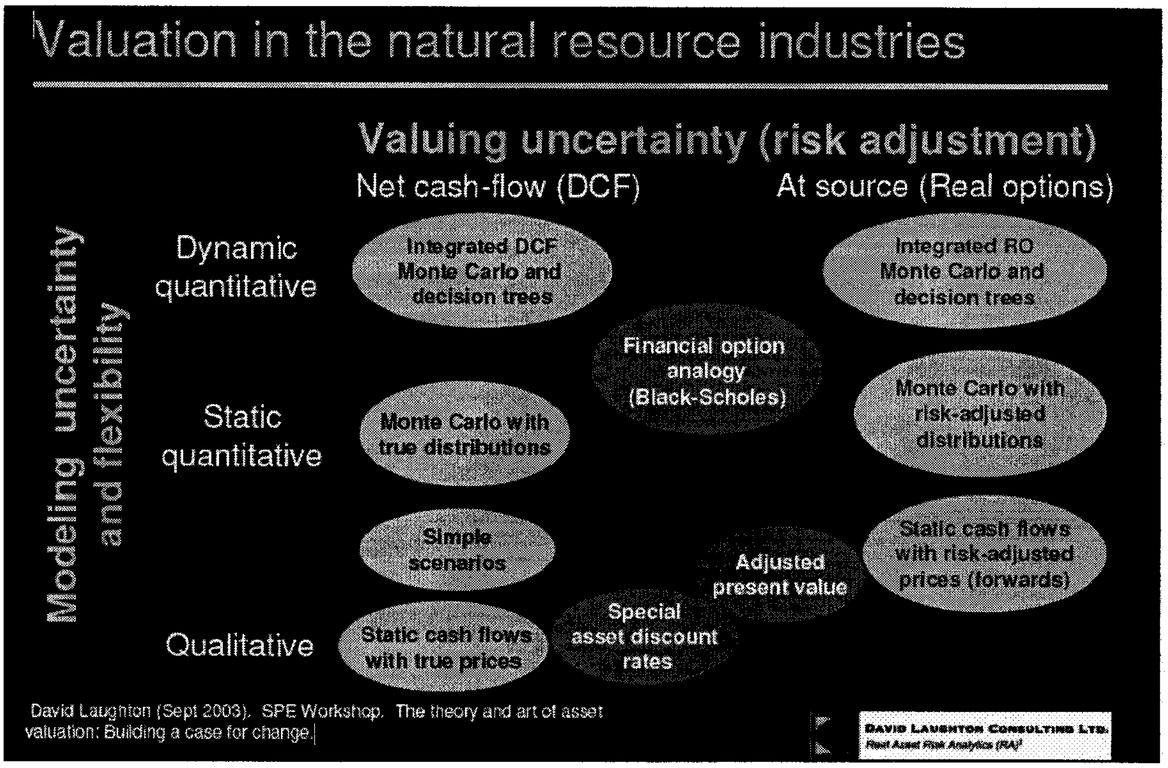


Figure 3: Valuing in the Natural Resource Industry

Chapter 4

Markov Decision Process Framework

General Framework

When making sequential decisions, it is important to consider both the short-term and long-term rewards. An action that seems bad in the current state may generate many positive rewards in the future. A simple example is the decision to invest in a project. In the short term, this seems like a bad idea because the instantaneous reward is negative. However, because the project will generate benefits in the future, the overall reward associated with the investment decision could be positive. Markov Decision Processes (MDPs) offer a tool to assist in sequential decision-making under uncertainty (Puterman, 1994). The MDP framework can effectively manage the instantaneous and future rewards in a way that will maximise the overall reward.

An MDP is a reinforcement learning type algorithm in which, through iterations, the performance of an agent is improved until it plateaus (Kaelbling et al., 1996). An MDP is a representation of an agent interacting in a stochastic environment, described as a five-tuple (S, A, T, R, γ), in which S is the state space, A is the action set, T is the state-transition probability distribution set, R is the instantaneous reward set and γ is the discount rate.

The state space S is a set of all the possible configurations of the environment in which the agent interacts. Therefore, s is simply the representation of the environment at time t. The action set is a list of all the actions available to the agent. Actions permit the agent to modify its current state s in order to attain a new state s' at time t+1. When an action is performed, the agent receives an instantaneous reward r. The reward set can take the form R(s,a) or R(s,a,s'). In the first case, the reward is solely dependent on the action taken in the current state $s_{(t)}$. In the second case, the reward is also dependent on the new state $s'_{(t+1)}$ reached as a result of performing the action.

Each action has its own state-transition function that can be expressed in the form of a transition probability matrix. The state-transition function is defined as follows, for all i, j $\in$ S, and k $\in$ A:

$$p_{ij}^{k} = \Pr(s_t = j \mid s_{t-1} = i, a_t = k) \tag{4.1}$$

A policy π is a mapping from states to actions. Thus, it determines which actions should be taken in each state.

An infinite-horizon MDP is used in this study because as stated by Bellman (1957), it is established that for the infinite-horizon case, an optimal deterministic stationary policy exists. The optimal policy is denoted as π^*.

Several techniques are available to solve an MDP. In this work, we discuss three methods: Policy Iteration, Value Iteration and linear programming.

Policy iteration

Kealbling, Littman and Moore (1996) present a concise overview of infinite horizon policy iteration. They explain how the optimal value of a state is the expected infinite discounted sum of rewards that the agent will gain if it starts in that state and executes the optimal policy

$$V^*(s) = \max_{\pi} E(\sum_{t=0}^{\infty} \gamma^t r_t) \tag{4.2}$$

This optimal value function is unique and can be defined as the solution to the simultaneous equations:

$$V^*(s) = \max_a \left(R(s,a) + \gamma \sum_{s' \in S} T(s,a,s') V^*(s') \right), \forall s \in S \tag{4.3}$$

The above equation demonstrates that the value of a state *s* is the instantaneous reward plus the discounted value of the next state multiplied by the probability of reaching that state, using the best available action. Given the optimal value function, we can specify the optimal policy as:

$$\pi^*(s) = \operatorname{argmax}_a \left(R(s,a) + \gamma \sum_{s' \in S} T(s,a,s') V^*(s') \right) \tag{4.4}$$

The policy iteration algorithm modifies the policy directly. The agent starts with a random policy. The value of each state is computed following the current policy as described above in equation (4.3). When the value of each state is determined, the agent makes modifications to the policy in order to improve the value of each state as prescribed in equation (4.4). The agent attempts to improve the value of the policy by making modifications to the first actions. In this fashion, strictly improvements will be brought to the policy. When improvements can no longer be made to the value of the state space, the optimal policy has been found and the algorithm stops. The algorithm is presented below.

Choose an arbitrary policy π'

Loop

$\pi := \pi'$

Compute the value function of policy π;

Solve the linear equations:

$$V_{\pi}(s) = \max_{a}\left(R(s,\pi(s)) + \gamma \sum_{s' \in S} T(s,\pi(s),s') V_{\pi}{}^{*}(s') \right)$$

Improve the policy at each state:

$$\pi'(s) = \operatorname{argmax}_{a}\left(R(s,a) + \gamma \sum_{s' \in S} T(s,a,s') V_{\pi}{}^{*}(s') \right)$$

Until $\pi = \pi'$

Kealbling, Littman and Moore (1996) present a concise note on the complexity of policy iteration:

> Note: Since there are at most $|A|^{|s|}$ distinct policies, and the sequence of policies improves at each step, this algorithm terminates in at most an exponential number of iterations (Puterman, 1994). However, it is an important open question as to how many iterations policy iteration takes in the worst case. It is known that the running time is pseudo-polynomial and that for a fixed discount factor, there is a polynomial bound in the total size of the MDP (Littman et al., 1995)

Value iteration

Value iteration is quite similar to policy iteration. The main difference between the two methods is in the way in which the policy evaluation is performed. Policy evaluation is done iteratively and may require many iterations before converging to an exact value for V^{π}. As stated by Sutton and Barto (1998), value iteration truncates the policy evaluation iterations down to a single sweep of the state space, without losing the convergence guaranties of policy iteration. The value of each state can be expressed as follows:

$$V_{k+1}(s) = \max_a E\{r_{t+1} + \gamma V_k(s_{t+1}) \mid s_t = s, a_t = a\} \tag{4.5}$$

$$\max_a \sum_{s'} \mathcal{P}^a_{ss'} \left[\mathcal{R}^a_{ss'} + \gamma V_k(s')\right] \tag{4.6}$$

Value iteration will take less time than policy iteration to perform an iteration, but will require more iterations in order to converge on the optimal policy. The difference in computation time between the two methods will depend on the nature of the MDP being solved. An evaluation of the performance of the two algorithms in solving mining problems is presented in Chapter 5.

Linear programming

An infinite-horizon MDP can be formulated as a linear program (D'Epenoux, 1963). The primal linear program involves maximizing the value of the state space. This is achieved by maximizing the sum of the values of all the states,

$$\sum_{j \in S} v_j \tag{4.7}$$

subject to the following constraints:

$$v_i \leq r_i^k + \gamma \sum_{j \in S} p_{ij}^k v_j \tag{4.8}$$

in which

v_i = state value $\qquad$ r_i^k = instantaneous reward

p_{ij}^k = transition probabilities $\qquad$ γ = discount rate

The dual of the above linear program can also be considered for minimisation:

$$\sum_{i \in S} \sum_{k \in A} x_i^k c_i^k \tag{4.9}$$

subject to the following constraint:

$$\sum_{k \in A} x_j^k = 1 + \gamma \sum_{i \in S} \sum_{k \in A} p_{ij}^k x_i^k \tag{4.10}$$

in which

x_j^k = policy flow

p_{ij}^k = transition probabilities

c_i^k = instantaneous cost / reward

γ = discount rate

The MDP Applied to Mining

This study focuses on the problem related to the optimisation of the mine extraction rate and cut-off grade of an open pit mine with respect to fluctuations in commodity prices, uncertainty in grades, and depleting ore reserves. During periods when commodity prices are high, it is advantageous to increase production in order to sell the most mineral product possible over these lucrative periods. When prices are low however, mine operators should reduce production in anticipation that prices will rise again in the future. In the case of cut-off grade, it should be raised when prices are low because the revenue generated will not be sufficient to justify further processing. Therefore, we must assess the benefits gained from having a robust mine plan that will permit an adjustment of production rate and cut-off grade according to commodity price and mineralisation grade.

To simplify the problem, it is assumed that the sequence in which the blocks are mined is predetermined, thus avoiding complications related to slope constraints. This predetermined sequence also reduces the computational complexity of the problem by limiting the size of the action set and the state space. Without this constraint, the agent would have to determine which block to mine in each period as opposed to whether or not it should mine the next predetermined block. In the first case, the number of possible actions would be at the least equal to the number of blocks in the pit, while in the second case, the actions are limited to a discretisation of feasible production rates and cut-off grades.

At each decision stage, the agent must determine the optimal production rate and cut-off grade. The option to terminate production is also available. However, when a decision is made to terminate production, operations cannot be resumed. Therefore, the action to terminate production leads to an absorbing end state. Figure 4 is a graphical representation of the agent's decision process. At each mining location, the agent must decide to either operate at a high production rate, low production rate or to terminate production and proceed to mine reclamation. If the agent elects to continue the extraction process, a decision must be made whether to send the extracted material to the mill or to a waste pile. The current model does not permit ore/waste selectivity within a decision step. Multi-agent MDPs show some potential to resolve this issue but is beyond the scope of the current work.

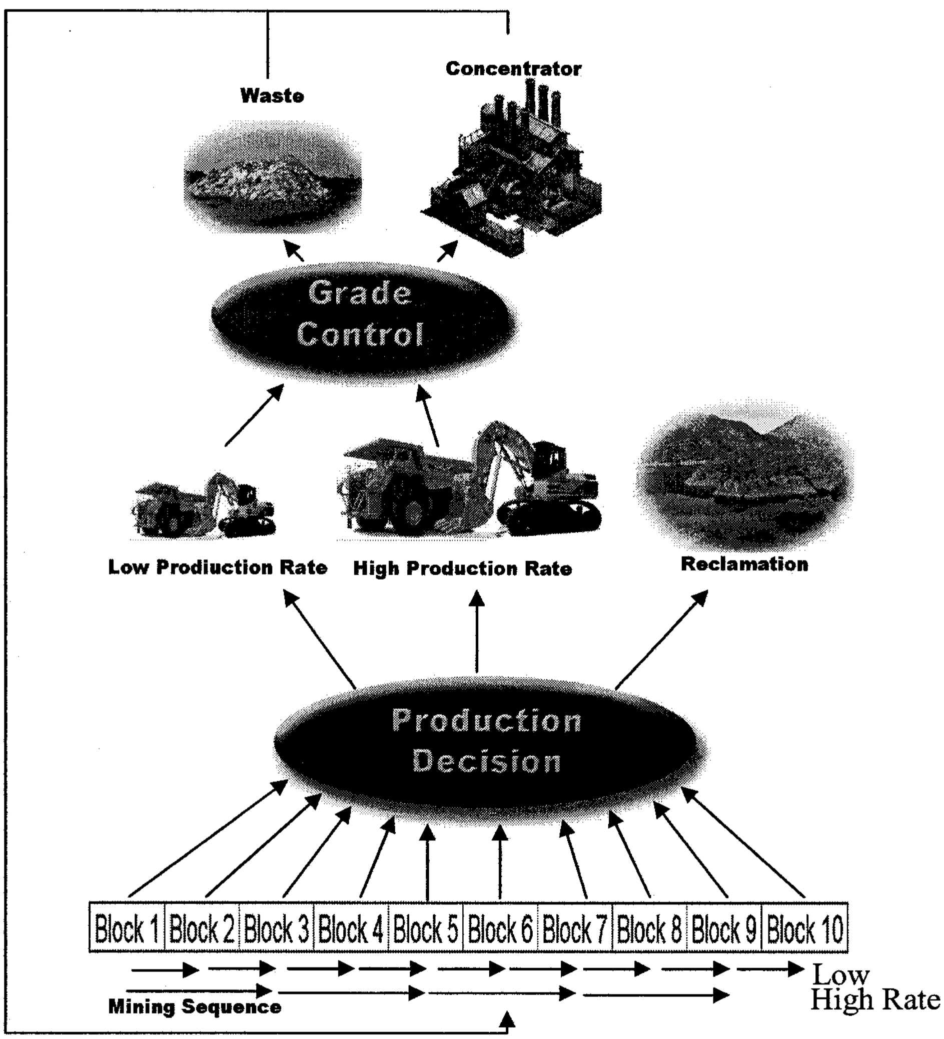


Figure 4: MDP Flow Chart

A low production rate results in a slower progression through the mining sequence as the incremental jumps are smaller. This is depicted in Figure 4 by the horizontal arrows below the blocks.

Optimisation Model

In this section we describe the models used to create the state space, transition probability matrices and the reward function required to implement an MDP.

In order to generate the state space and the transition probability matrices, the two sources of uncertainty considered (i.e. grade and commodity price) are modeled individually and then combined.

Modeling block grades

To portray the uncertainty associated with the mineral content (grade) of the mining blocks, geological block models are generated using Sequential Gaussian Simulation (SGS).

This method is based on the decomposition of the multivariate Gaussian distribution into the product of univariate conditional distributions. As stated by Isaaks (1990), these univariate distributions are normal with a conditional mean equal to the simple kriging estimate and a conditional variance equal to the simple kriging variance. The simulation is done following (4.11).

$$z\delta + \mu = x \qquad (4.11)$$

in which

z = Normal random number with mean of 0 and standard deviation of 1

δ = Simple kriging standard deviation

μ = Simple kriging estimate of node value

x = Simulated node value

The first step of SGS is to create a random path that contains all of the nodes in the block model to be simulated. Once the path is established, the value of the first node is simulated by: 1) determining the conditional cumulative distribution function (ccdf) by performing simple kriging using all sample data, and 2) generating a value at the node by sampling the ccdf. The new node value is then added into the data set and the algorithm moves on to the following node.

Simulated block models where generated as described in Farrelly (2002). These block models are equally probable representations of the actual mineral deposit

(Isaaks, 1990). The block models are used to represent the geological uncertainty of the deposit.

Modeling commodity prices

There are many time series models that may be used in price simulation. The most commonly used are the random walk model, the *Orntein-Uhlenbeck* mean reverting model and the geometric Brownian motion model. It is important to carefully select the diffusion process that best represents the variability of the underlying process being modeled.

The simplest form of the mean reverting process is known as an *Orntein-Uhlenbeck* process:

Error! Objects cannot be created from editing field codes. (4.12)

in which

dx = Change in price over one period

δ = Reversion speed, i.e. force with which the price is pulled back towards the mean

$\bar{x}$ = Mean price (the historical mean is often used)

x = Spot price, i.e. current price of the commodity

dt = Change in time

σ = Volatility, i.e. the price standard deviation over one unit of time

dz = Normal random number

The use of the *Orntein-Uhlenbeck* model requires some caution as it may generate negative values. The solution to this problem is to use Geometric Brownian motion (GBM) to model the commodity prices. This generates a logarithmic random number following a Weiner process. A diffusion process is said to follow GBM if it satisfies the following stochastic differential equation

$$dS_t = u\, S_t\, dt + v\, S_t\, dW_t \tag{4.13}$$

This equation has an analytical solution:

$$S_t = S_0 \exp\left((u - v^2/2)t + vW_t\right) \tag{4.14}$$

In order to model prices efficiently for an MDP, a discretised version of the diffusion process is required to reduce the size of the transition probability matrix.

The discretisation is done following the bin method presented by Barraquand and Martineau (1995).

Once the transition probability matrices for the block grades and prices are created individually, the general transition probability matrix is generated. This combination is simplified by the fact that the transition probabilities of the grades and prices are independent. Thus, the combined probabilities are simply the product of the probabilities of the two parameters.

Reward function

It is assumed that there is no delay between the time at which a block is mined and processed and the time at which the recovered mineral product is sold. This assumption simplifies the calculation of the reward generated from extracting a particular block. The reward is the revenue generated from selling the mineral product recovered from the block less the cost incurred. The reward function is:

$$R = Grade \bullet Price - Cost \tag{4.15}$$

This function can be expanded to include other parameters such as mine dilution, mill recovery, Net Smelter Return, depth dependant mining costs and variable processing costs.

MDP requirements

The MDP framework requires that the current state be fully observable. Accordingly, it is assumed that the sample cuttings from production drilling conducted prior to mining provide enough information to claim that the grade of the current block to be mined is fully observable. The commodity price is also fully observable in the current state as indicated by the current market price. The block to be mined in the current state is also known. Therefore, the current state is fully observable. Another requirement is that the environment must be Markovian, which means that all the information from the past is contained in the current state. Path dependency complications are avoided by having a fixed mining sequence as input to the MDP. Otherwise, the Markov property would not be respected, and to do so would require an enormous state space. Also, due to the discrete nature of the decision steps, each action must require the same completion time. This requirement is respected because the production decisions involve different production rates. Thus, the amount of material extracted changes, but the time unit remains the same.

The MDP framework lends itself quite well to the task of schedule optimisation and valuing options. The contribution of options to project value can be measured by modifying the options available to the agent. Once the best policy is found, Monte-Carlo simulations can be conducted on the project while implementing that policy, thus generating a probability distribution curve of the project value. By comparing the distribution curves of policies in which different options are included or excluded, it is possible to asses the effect of these options on project value. The most important advantage of using MDPs to value options is that the optimal policy, i.e. that required to achieve the maximum value of the state space, is found concurrently. This policy can then be used as a guide in the optimisation of the production schedule.

Chapter 5

Experimentation

Introduction

In this section, we look at some examples of applying an MDP to mine scheduling optimisation. Actual size case studies take approximately 8 days to converge on an optimal policy, using a 64 bit inter(R) Xeon 3.6 GHz processor with 2 gigs of RAM operating on Linux SUSE V10. Consequently, for clarity and time considerations, most of the testing and demonstrations are performed on hypothetical examples. A small hypothetical case study is used to demonstrate how to read a policy and to assess the value of incorporating options in the valuation process. The policy's sensitivity to discount rates is also tested using the small case study. A medium-size case study, based on real geological data, is used to measure the computational time required by the two algorithms used to solve MDPs. A large-scale case study based on actual values of a mineralized deposit is used to assess the applicability of MDPs to real size projects. The description of the mineralized deposit is not disclosed due to a confidentiality agreement. The MPD is implemented using the Matlab MDP toolbox V2.0 provided by the Institut National de la Recherche Agronomique (INRA).

Reading the Policy

The graphical representation of the policy is somewhat complicated by its three dimensional nature. To simplify policy interpretation, we first consider an example in which the grade dimension is omitted. This results in a two-dimensional policy, as illustrated in Figure 5, strictly dependent on prices and remaining ore reserves.

The commodity prices are given on the y axis while the mining locations (i.e. blocks to be mined) are given on the x axis. The colours indicate the actions prescribed by the policy. To determine which action should be taken when in a particular state, the decision maker must find the intersection of the parameters describing the current state. Consider that a decision must be made concerning block 3. If the current market price of the commodity is $4/unit, the policy prescribes that the block should be mined. The X in the ***Mine*** rectangle indicates the current state in Figure 5.

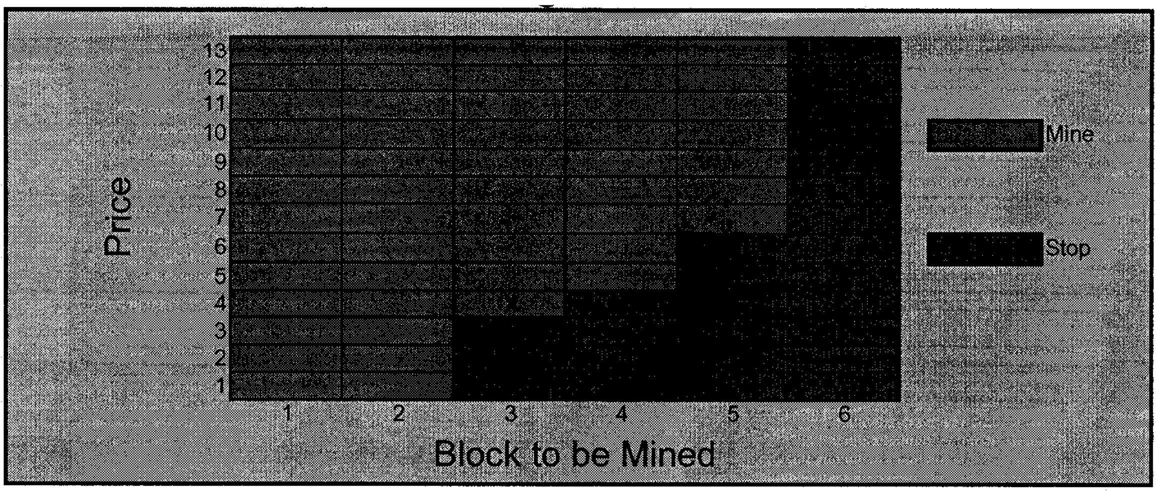


Figure 5: Sample Policy

Now, considering the same block (3) but a commodity price of $3/unit, the operation should be stopped, as indicated by the *Stop* rectangle on the policy chart.

Adding a third dimension representing grade makes the graphical representation of the policy cumbersome and difficult to read. Therefore, policy charts are created for each grade possibility, as illustrated in Figure 6 below. In this example, there are only 3 possible discrete grade transitions. In a real case study however, more grade state transitions would be used to generate a better representation of the grade distribution. Consider again the decision concerning block 3. If the grade of the block is high, the policy prescribes to mine the block regardless of the price. However, if block 3 is low grade, the policy prescribes to mine it only if the price is above $4/unit.

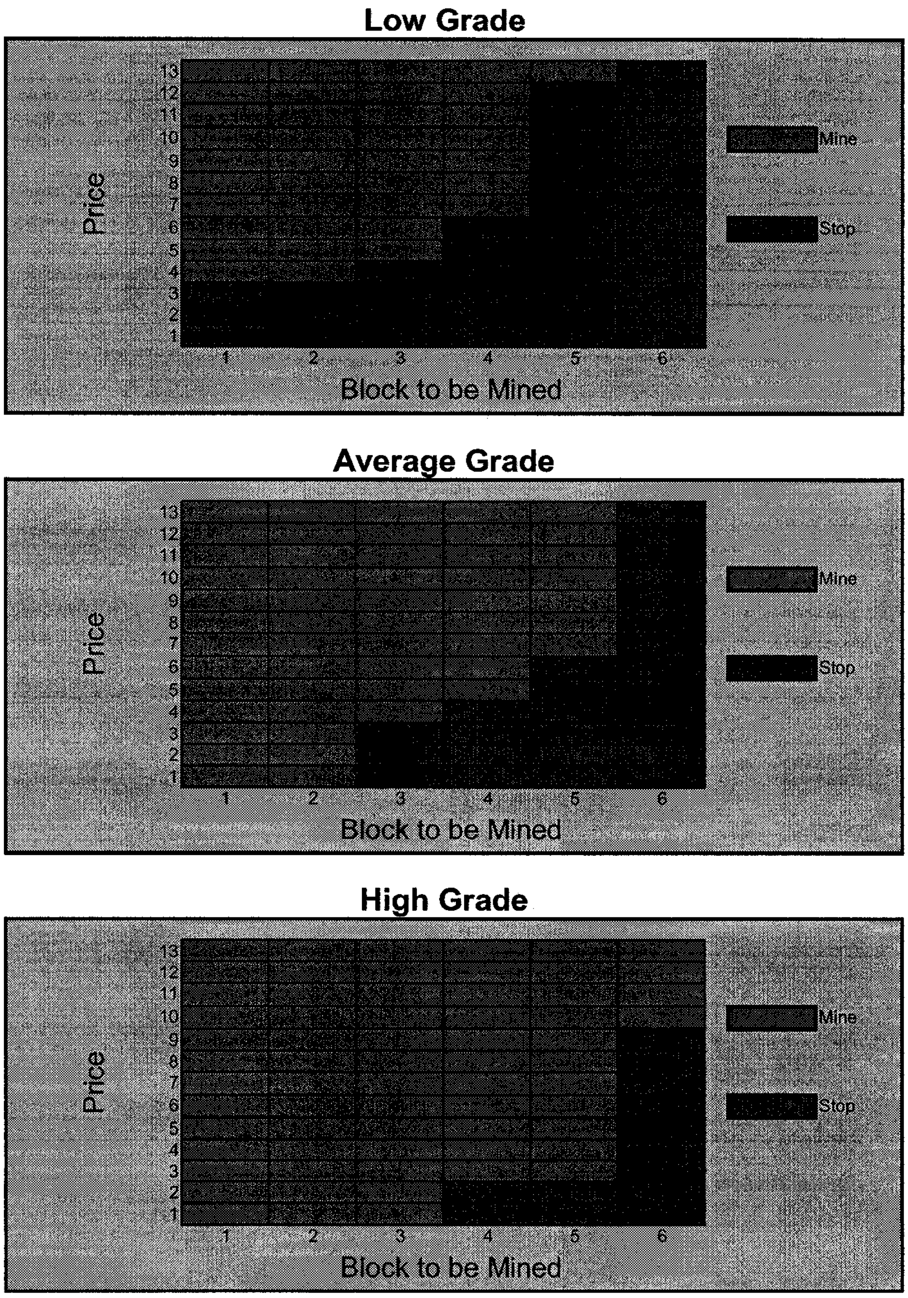


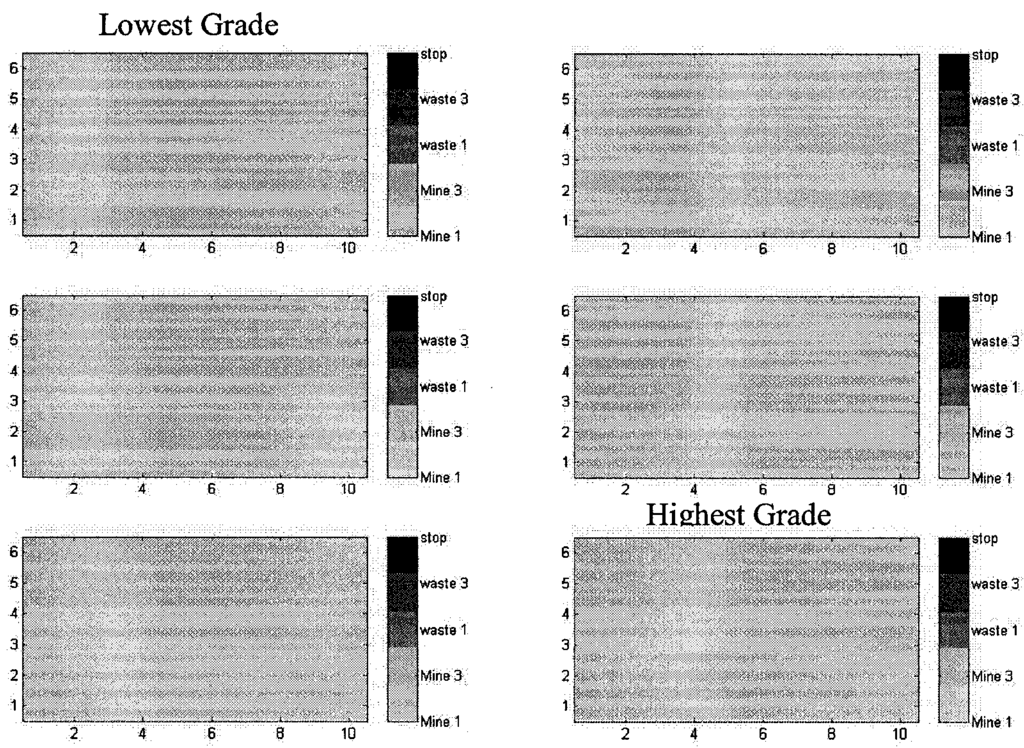
Figure 6: Sample Policy with Multiple Grades

Policy Interpretation

The optimisation model is applied to a hypothetical small-scale mine. In this example, there are 10 blocks to be mined, each of which has six possible grade values. The number of possible prices is limited to 6. Therefore, the total size of the state space is 361, including the end absorbing state.

Three cases are considered to demonstrate the financial benefits of implementing a robust mine plan. In the first case, the original mine plan is executed without any modification, i.e. the entire pit is extracted regardless of changes in market conditions and geological variability. In the second case, the agent is allowed to react to information; the actions available to the agent are either to mine the next block or terminate mining. In the third case, a robust mine plan permits a second production rate and the option to make a waste pile. Thus, there are five actions to choose from at each decision step: mine the next block as ore, mine the next three blocks as ore, mine the next block as waste, mine the next three blocks as waste or terminate the process. Each action has the same execution time. It is assumed that the price does not change during the execution of the action and is only updated at the next decision step.

The policy for case 1 shown in Figure 7 is not very interesting because the only action available to the agent is to mine at a low production rate. The policy to mine at a low rate is indicated by the ***Mine 1*** covering the entire policy charts.



Y axis = Commodity Price $/unit   X axis = Block to be mined

Figure 7: Policy without Options

Case 2 however, is much more interesting. By including the option to terminate the mining process, it is now possible to get an appreciation of the pit design's exposure to early mine closure risk. In Figure 8, ***Mine 1*** represents the action to mine at a low rate, while the ***Stop*** represent the action to close down the operation. The chart on the top left is the policy for the lowest grade and the chart on the bottom right is the policy for the highest grade. It can clearly be seen that the lower grade cases have a higher probability of early mine closure than the higher grade cases. Also, as the mine matures, higher prices or grades are required to maintain the mine open.

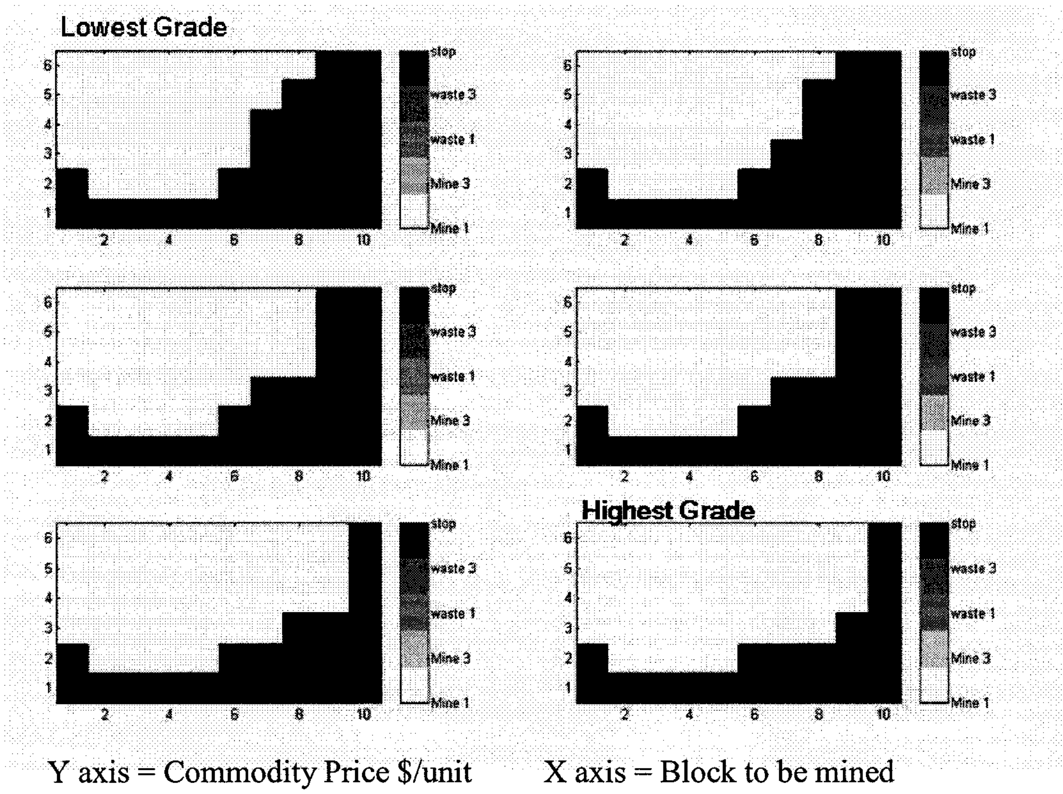


Figure 8: Policy with Option to Stop

Case 3 demonstrates the policy obtained when all five actions mentioned above are available to the agent. The policy is illustrated in Figure 9. ***Waste1*** represents the action to mine blocks as waste at a low rate, ***Waste 3***, to mine blocks as waste at a high rate, the ***Mine 3***, to mine blocks as ore at a high rate, *Mine 1*, to mine blocks as ore at a low rate and *Stop*, to close down the operation.

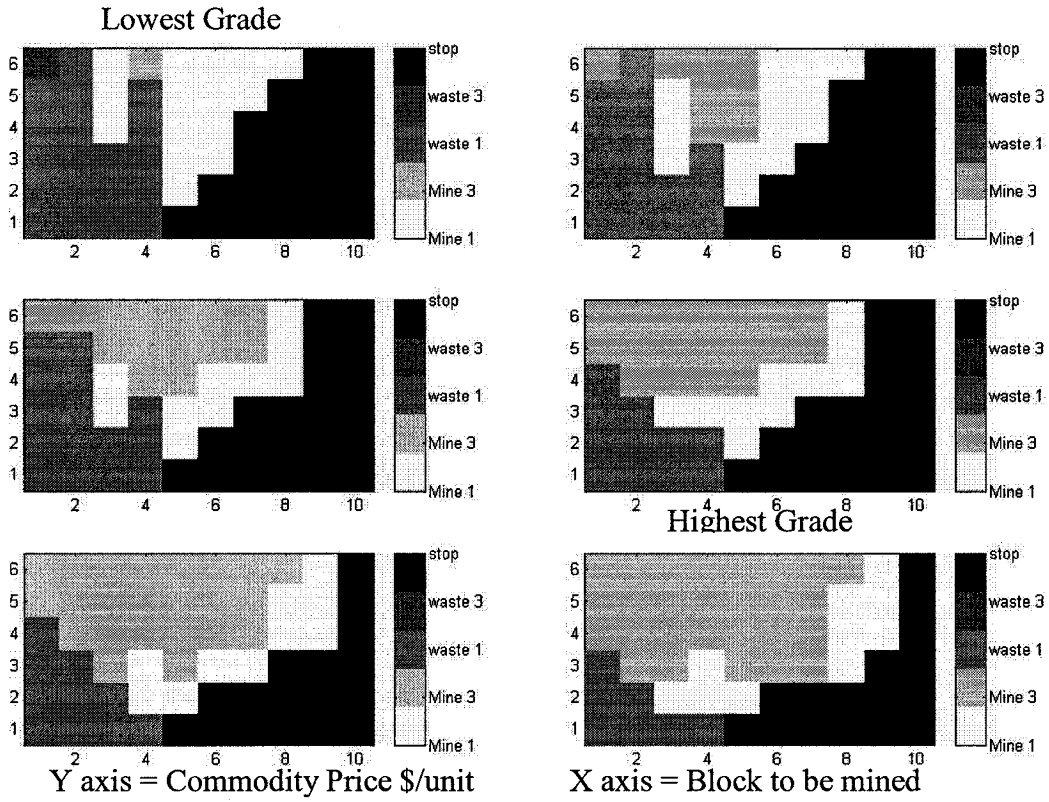


Figure 9: Policy with all Options Available

The agent's behaviour is in accordance to a manager's intuition. When prices are high, the production rate is increased. By doing so, a larger portion of the mineral product contained in the reserves is sold at high prices, thus increasing the overall revenue. When prices are low, the production rate is reduced with the anticipation they will rise in the future. Deterministic NPV optimisation prescribes that revenues should be generated as early as possible during the project life. This suggests using the highest feasible production rate during at least the early years of project life. When using an MDP, the agent recognises the growth potential of commodity prices. By selecting a lower production rate at low prices, the agent demonstrates it can effectively measure the trade-off between a reduction in present value of a future cash flow (due to more discounting) and a potentially increased cash flow value due to a higher commodity price.

From Figure 9, we can also get an appreciation of the cut-off grade, or stated otherwise, the cut-off price of a block. As the price increases, some blocks make the transition from being mined as waste to being mined as ore. When the price and metal content of a block is high enough to justify further processing, the block is considered ore and sent to the mill. Otherwise, it is stockpiled as waste.

The Effect of Options on Project Value

Once the optimal policy is found, it is possible to determine the value of this policy using Monte-Carlo simulations. Using the same stochastic environment created for the learning process, the optimal policy is implemented on many project iterations. The present value of project cash flows is computed for each iteration, thus providing a distribution curve of project value under the optimal policy. Comparing the distribution of present values of the three cases mentioned above yields the additional value of having a robust mine plan over a "just in time" approach.

Figure 10 (CASE 1) illustrates the distribution associated with case 1. The distribution is symmetric and resembles a Gaussian distribution with a mean of 0. As it would be difficult to justify an investment with an expected present value of benefits of zero and a probability of loss of about 50 percent, this project would likely be rejected.

In case 2, the option of early mine closure is added to the valuation. Early mine closure is only exercised when conditions are unfavourable. Therefore, it is an option that reduces the downside risk. The reduction in downside risk can be observed in

Figure 10 (CASE 2) in which the probability of financial loss is reduced and the left side of the distribution is compressed towards 0. Because we have not yet included any options that permit us to take advantage of the upside potential associated with the uncertain variables of the project, the right side of the distribution remains unaffected.

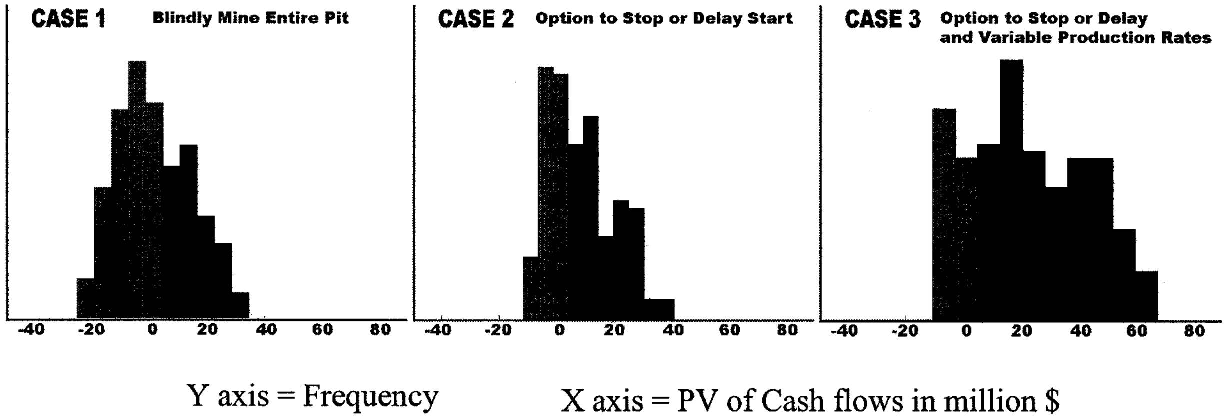


Figure 10: Distribution of Projects Values

By adding the option of variable production rates and ore/waste selectivity, it becomes possible to not only reduce the downside risk, but to benefit from the upside potential as well.

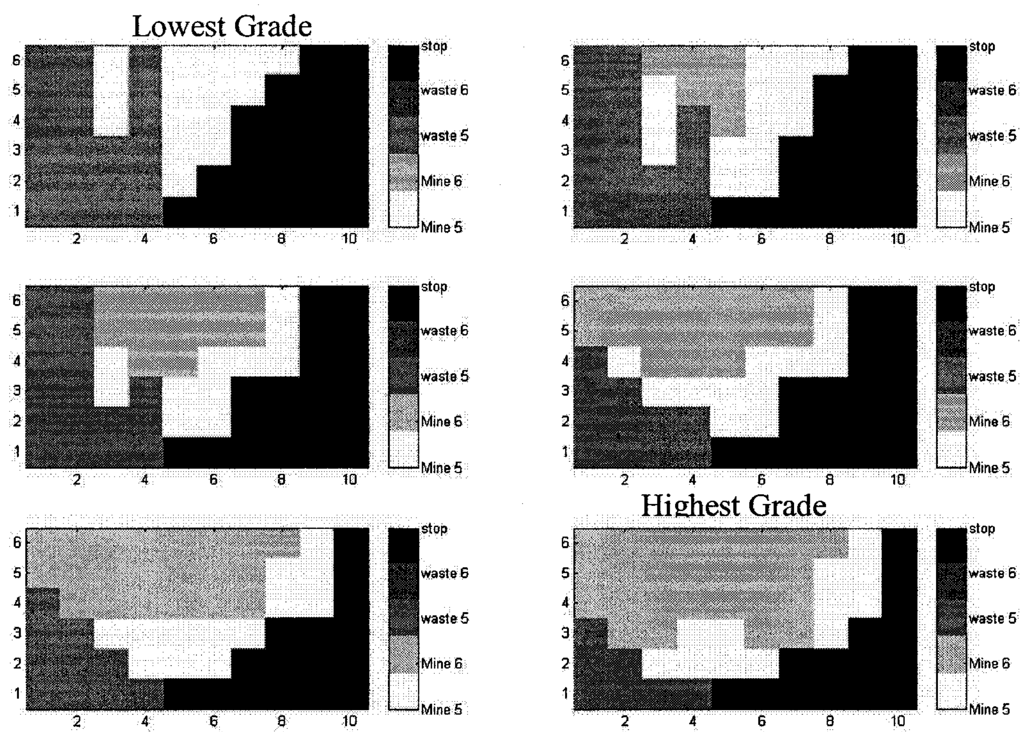
Figure 10 (CASE 3) illustrates the value gained from including both options in the valuation process. As can be seen, the probability of financial loss remains about the same as in case 2, but the positive side of the distribution is stretched to the right, thus increasing the project's value by exploiting the upside potential resulting from grade and price uncertainties.

Effect of Discount Rate on the Policy

To study the effect of discounting, optimal policies were obtained using different discount rates. Optimal policies were found for discount rates ranging from no discounting to 40%. Results are shown in Figure 11. As the discount rate increases, its impact on the optimal policy is threefold. The first effect is to force extraction at the higher production rate. The second is to increase the situations in which the option of early mine closure is exercised. The third results from the reduced selectivity at the higher production rate, thus causing the dilution of ore with waste. These three effects are explained below.

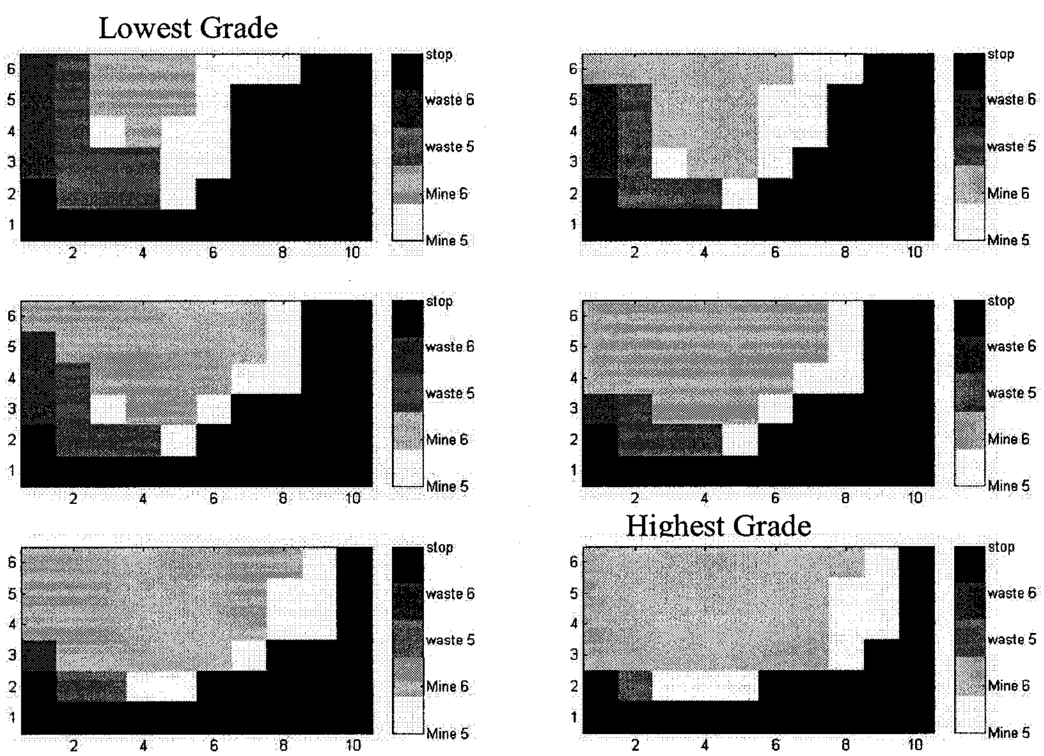
As the discount rate increases, it is natural to think that the policy would favour higher production rates. When the time value of money is high, it is essential to produce faster to reduce losses due to discounting. This phenomenon can be observed in Figure 11. As the discount rate increases, the area ***Mine 6*** representing the higher production rate increases.

The present value of future cash flows is inversely proportional to the discount rate. Therefore, higher discount rates result in reduced project value. This reduction in value may result in a project being delayed or abandoned due to the negative present value of remaining cash flows. The latter is confirmed by the increased probability of early mine closure represented by the ***Stop*** areas in the policy chart. As the discount rate increases, the ***Stop*** area expands towards the lower left corner of the policy chart. When the time value of money is ignored, the optimal policy indicates that production should begin for the entire range of prices (leftmost column of the policy chart). However, at a discount rate of 40%, the project is delayed for commodity prices of 1 and 2 $/unit.



Y axis = Commodity Price $/unit X axis = Block to be mined

No discounting



40% discount rate

Figure 11: Effect of Discounting on Optimal Policy

As explained above, higher discount rates force extraction at the higher production rate. This in turn reduces mine selectivity and causes dilution. Blocks that would be mined as waste at the low production rate are now extracted as ore at the higher production rate, along with higher grade blocks. This effect can be observed by the reduction in size of the ***Waste*** area in the policy grid as the discount rate increases.

Policy Iteration versus Value Iteration

In a previous section, we discussed the distinction between Policy Iteration and Value Iteration. In this section we discuss the impact of these algorithms on the number of iterations and computing time required to solve the MDP. The test is conducted on a medium-size MDP. The block model is based on an actual deposit, but the support is quite large. The smallest mining unit in this example is 50 m x 50 m x 50 m. Therefore, the pit contains 406 blocks. The state space consists of 5 prices and 6 simulated block models, thus yielding 8120 states.

Figure 12 illustrates the learning curves associated with the two algorithms. Computing times are also reported.

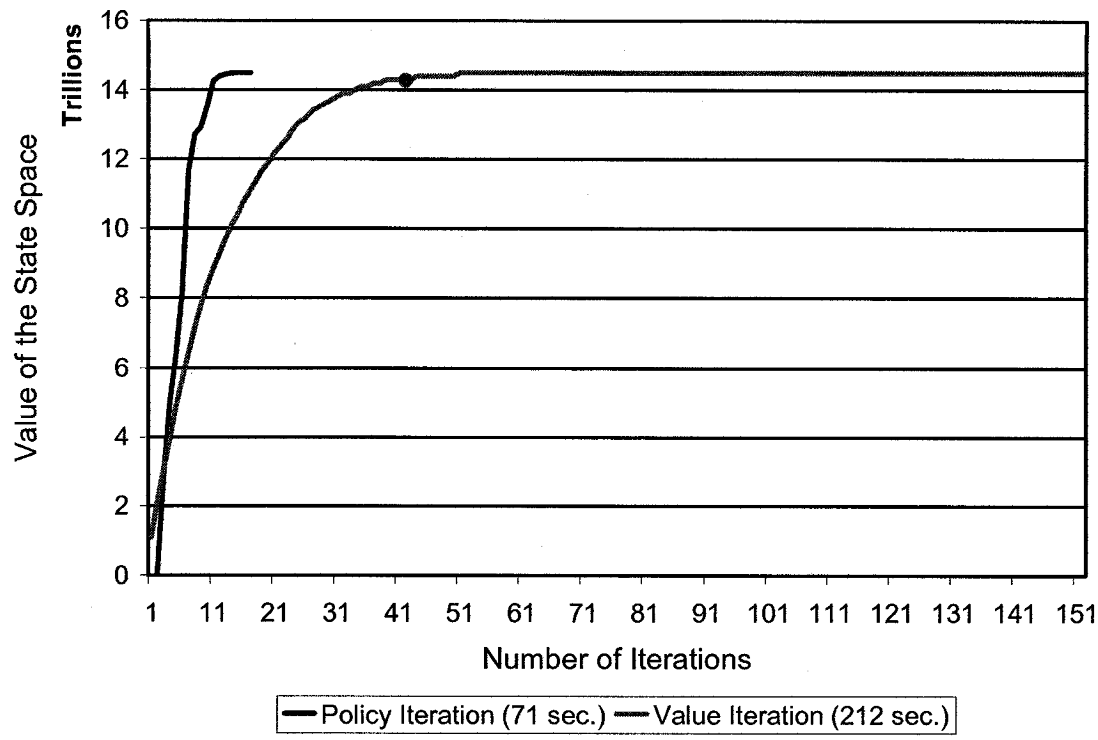


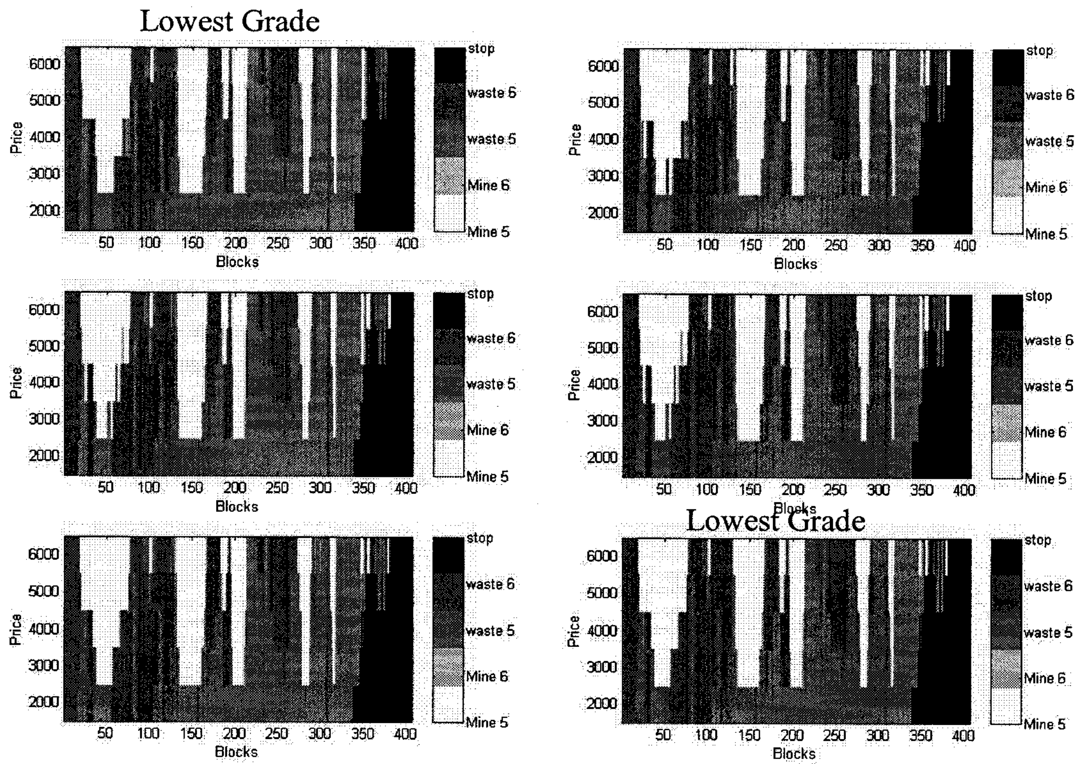
Figure 12: Evolution of State Space Value with Respect to Number of Iterations[1]

The policy iteration algorithm found the optimal policy after 17 iterations while value iteration required 152 iterations. In terms of computation time, Policy Iteration is much faster than Value Iteration, taking 71 seconds as opposed to Value Iteration which required 212 seconds. It can be observed in the figure above that the longer computation time associated with the value iteration algorithm results from the diminishing marginal improvements to the policy after 50 iterations. Value Iteration can find a near optimal policy within a shorter number of iterations, but requires many more iterations to reach the optimal

[1] The value of the state space is the sum of the values of all of the states. In absolute terms, it does not give much information on the actual project value but is an effective tool to monitor the learning process.

policy. If the exact optimal policy were not required, Value Iteration could be used as a good approximation method. The computation time required to reach the dot on the value iteration curve is only 33 seconds. Therefore, when an exact optimal policy is required in mine optimisation problems, Policy Iteration is faster than Value Iteration. However, for larger scale cases in which computation time is significant, Value Iteration can be used as a faster estimation alternative.

Both Value Iteration and Policy Iteration generated the same policy. This policy can be seen in Figure 13. The first thing to notice in this policy is that the increased ore production rate is never used within the simulated price range. However, the increased waste production rate is used. The second is that there is a clear cut-off price at $2000, as all the blocks are considered waste when the price is below $3000. Finally, the delay option is never exercised but the early mine closure option has a high probability of being exercised.



Y axis = Commodity Price $/unit X axis = Block to be mined

Figure 13: Optimal Policy for Medium-size Case Study

Note on input mining sequence

The Whittle software was used to generate the mining sequence. Whittle is not generally used for short-term planning. The results obtained here testify as to why the software is not an adequate tool for short-term scheduling. The reason is that Whittle does not distribute the input constraints uniformly within a time period. For example, if the mill limit is 10 000 tonnes per year, Whittle does not distinguish between milling these 10 000 tonnes within the first month or uniformly over the entire year. Therefore, during the first portion of the year, ore will be mined and sent to the mill until the maximum units are reached for the year. Then, the milling will be stopped and waste will be mined for the remainder of the year. This phenomenon can be observed in the alternation between waste and ore in the policy chart.

Case Study on Copper Deposit

A large scale study is conducted on a copper deposit. In-situ reserves are estimated at 737 878 tonnes of copper contained in 106 millions tonnes of ore. Assuming a mining rate of 8 355 500 tonnes per year, the expected mine life is approximately 28 years.

Whittle

The Whittle Four-X software is used to determine the mining sequence that will be used as input to the MDP. The Whittle optimisation is performed on the kriged block model. The parameters used to generate the mining sequence are given in Table 2.

Table 2: Cost Parameters

Mining Cost = 2 $/t of rock	Milling Cost = 4.50 $/t of ore
Selling Price = 5500 $/t of Cu	Selling Cost = 1500 $/t of concentrate
Mining limit = 8 355 500 t/year	Milling limit = 4 000 000 t/year

Markov decision process

Once the mining sequence is generated, it is overlapped with the simulated block models to generate the sequences of metal content that will be used as input to the MDP.

In this case study, two production rates are considered: the base production rate is 22 000 tonnes per day and the higher production rate is 26 400 tonnes per day, representing a 20% increase in production[2]. The base case mining cost and milling cost are the same as the ones used in Whittle, i.e. $2 per tonne and $4.50 per tonne. There is a 10% premium on both the mining and milling costs to operate at the higher production rate. When a block is mined as waste, only the mining cost applies. Mill recovery is estimated at 80%. The mean selling price is $5500 per tonne of copper and the selling cost is estimated at $1500 per tonne of concentrate. A mean reverting diffusion process is used to simulate copper prices and compute the price transition probabilities.

The block model contains 12 979 blocks. There are 6 simulated grades for each block and 5 simulated prices. The state space therefore contains 389 370 states.

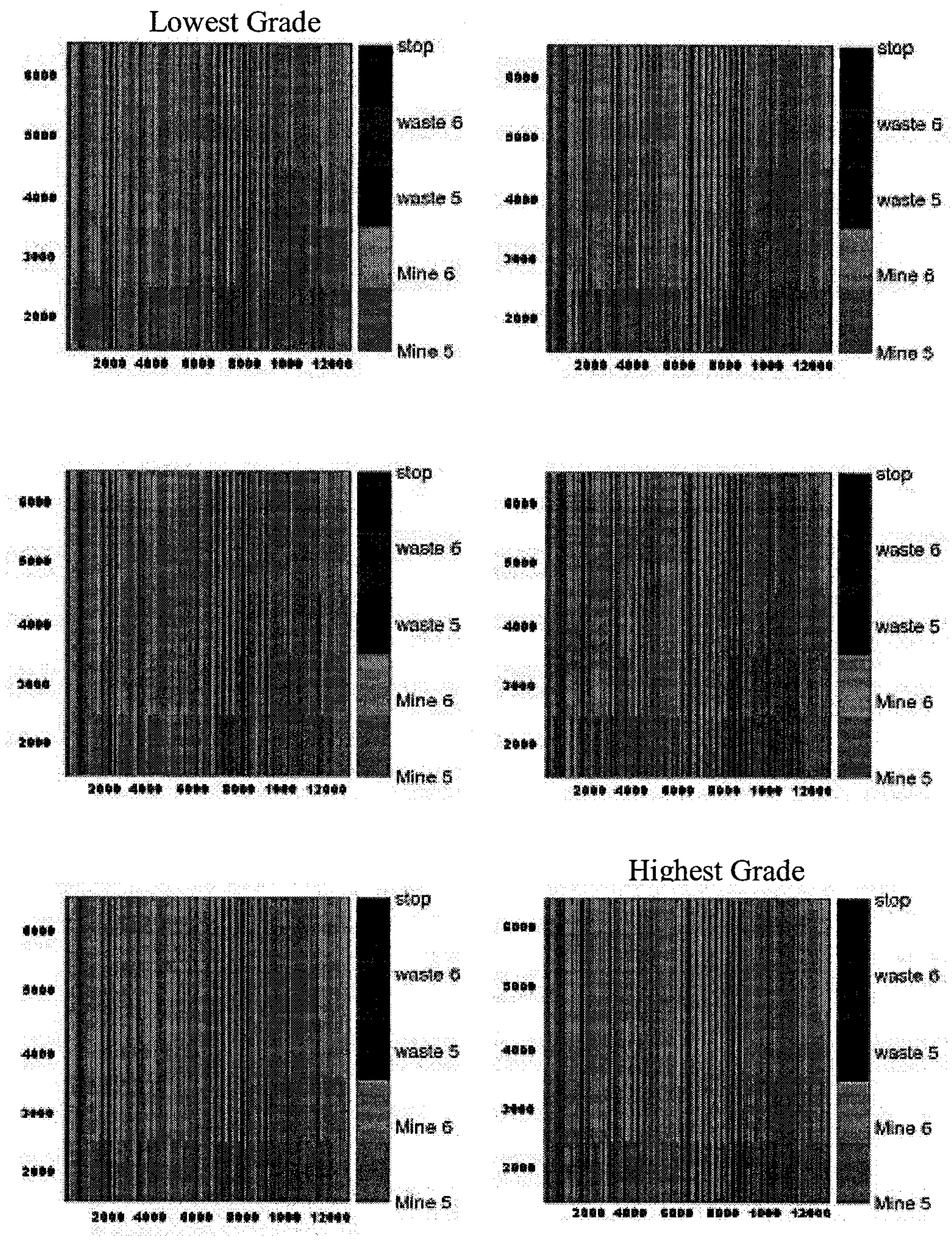
[2] Neither of the selected production rates or combination of the two will produce the maximum present value of the project. In order to find the true maximum present value many production rates over then entire interval of technical possibilities should be studied. However, considering such a large number of options is not computationally feasible

It took approximately 50 hours to solve the MDP on a Xeon computer and required just over 2000 iterations.

The block model used for the case study is associated with a very lucrative deposit and ore limits are well defined. The size of the ultimate pit is not very sensitive to commodity prices. In these conditions, the Lerch-Grossam algorithm used by Whittle is quite effective in determining the size of the ultimate pit. Therefore, the option of early mine closure is very unlikely to be exercised as shown in Figure 14. Also, there aren't any simulated price scenarios that warrant a delay in production start-up. Knowing that these two options are not exercised provides valuable information to the mine planners. In this case, since the early mine closure option has no value, managers have no incentive in keeping this option available. Therefore, selling contracts, such as futures, can be negotiated.

From the policy chart shown in Figure 14, it can be seen that the variable production rate option is exercised when mining ore. During periods of high prices, the production rate is increased, while over periods of low prices, it is reduced. It is interesting to note that in periods during which waste is mined, the production rate is rarely changed. The higher rate is almost always selected to accelerate the revenue inflow by exposing ore at a faster rate.

For this case study, the policy demonstrates that the cut-off grade is not very sensitive to prices. There are few blocks that change from being considered as waste to being considered as ore at higher prices. Most of the blocks that do make a transition from waste to ore do so when the price changes from $2000 to $3000.



Y axis = Commodity Price $/unit X axis = Block to be mined

Figure 14: Optimum Policy for Large-scale Case Study

Chapter 6

Conclusions and Recommendations

Conclusions

A brief literature review was conducted on the use of real options in the mining industry. Some of the shortcomings of the techniques proposed were presented, demonstrating the need for improvement in real option valuation techniques for the mining industry. Successful application of Markov Decision Processes to various optimisation problems were presented and it was hypothesised that MDPs could be successfully applied to mine scheduling optimisation, thus providing an alternative method of valuing real options in the mining industry.

It was demonstrated that managerial flexibility, expressed in the form of options, does in fact contribute to project value. A list of mining specific options was presented. The real option approach was explained, followed by a comment on the issue of discounting. A series of option valuation algorithms were described.

The MDP framework was presented along with three algorithms used to solve MDPs, namely Policy Iteration, Value Iteration and linear programming. The particularities of applying MDPs to the mine scheduling optimisation problem were discussed, followed by a detailed explanation of the information required as input to the MDP. It was established that the model developed respected the MDP requirements.

Three case studies of increasing scale were conducted. A small hypothetical case was used to demonstrate how to interpret a policy and to assess the value of incorporating options in the valuation process. As well, the sensitivity of a policy to discount rates was tested using the small hypothetical case. A medium-size case was used to compare the computational requirements of Value Iteration versus Policy Iteration. A large scale study was conducted on an actual deposit in order to measure the applicability of MDPs to real-size project valuation.

It is shown that optimisation using MDPs generates substantial financial savings due to improved managerial decision-making. The application of MDPs to mining operations presents great potential and warrants further development.

Recommendations

Using MDPs for project valuation and optimisation provides more information to the decision maker than the traditional NPV approach. However, this additional information comes at a cost. The traditional NPV approach is simple and easy to apply. Generating a distribution of project values using an MDP is very time consuming and computationally expensive. Such an extensive analysis should only be conducted on projects where managerial flexibility is anticipated to contribute substantially to the value of the project.

The application of MDPs to mining projects generates very large state spaces and requires large amounts of RAM. Due to the fact that the current version of Matlab is unable to use swap memory on a Windows operating system, large MDPs cannot be implemented on this platform. As a large proportion of mining companies use the Windows platform, new hardware would be required in order to implement MDPs.

The options considered in the case studies where the change in production rate and cut-off grade, as well as early mine closure. As mentioned previously, there are a number of other options inherent in mining projects that could contribute to increased project value, namely further exploration, mining sequence and stockpiling. The inclusion of these options would require an enormous state space and would render the MDP virtually intractable. Particular interest should be given to the mining sequence option as MPDs could potentially outperform the Lerch-Grossman algorithm currently used by Whittle to generate a mining sequence. Further work on multi-agent MDPs could reduce the state space size and enable the inclusion of more options in the valuation process.

The cost structure of the mining operation, i.e. the relationship between fixed and variable costs on the one hand, and production rate and cut-off grade on the other, was ignored. Nevertheless, the potential benefit of additional flexibility was well demonstrated. Given adequate information on the cost structure of the company, this aspect could easily be included in the MDP model.

In the case studies used in Chapter 5, a somewhat ad-hock conditional cumulative distribution function (ccdf) was used to calculate the grade transition probabilities. At this time, the software used to generate the Sequential Gaussian simulations does not provide the ccdf. Modifications to the source code would be required to do so. Given the availability of the ccdf, a more accurate transition probability matrix could be generated.

The selection of an action in a particular state is based on the expected value generated from choosing that action in that state. The expected future value is easy to compute but by using and expected value, much of the information from the distribution of possible future values is lost. Rather than making decisions based on expected value, it would be more realistic to base decisions on a utility function that would incorporate the complete distribution of possible outcomes. By doing so, the probabilities of financial loss and maximum potential gains would be included in the optimisation process.

The model developed here does not allow selectivity within a decision period. Thus, when a decision is made to mine blocks as waste or ore during a particular period, all of the relevant blocks are mined as such. Therefore, caution must be taken in that the number of blocks mined during a decision period must remain small to reduce the effect of dilution. However, as the production rate is determined by the number of blocks mined in a period, a large number of blocks per period are required to replicate small variations in the production rate. Therefore, a trade-off must be made between mine selectivity and the smallest production rate increment. Multi-agent learning could potentially eliminate this limitation.

As computational costs are reduced and larger amounts of memory become available, smaller increments could be used in the discretisation of the commodity price and more grade simulations could be used. Also, the number of actions made available to the agent is limited by the RAM capacity of the computer used to implement the algorithm. Developments in computer technology will facilitate the implementation of MDPs in the mining sector.

References

Andreae, J. H., 1963: STELLA: A Scheme for a Learning Machine. In Proceedings of the 2nd IFAC Congress, Basle, pp. 497-502. Butterworths, London.

Andreae, J. H., 1969a: A Learning Machine with a Monologue. *International Journal of Machine Studies*, 1:1-20.

Andreae, J. H., 1969b: Learning Machines – a Unified View. In A. R. Meetham and R. A. Hudson (eds.), *Encyclopaedia of information, Linguistics and Control*, pp. 261-270, Pergamon, Oxford.

Andreae, J. H., 1977: *Thinking with a Teachable Machine.* Academic press, London.

Armstrong, M. and Galli, A., 1997: Option Pricing: A new Approach to Valuing Mining Projects, CIM bulletin, vol. 90, pp. 37-43.

Arrow, K.J., Blackwell, D., Girshick M. A., 1949: Bayes and Minimax Solution of Sequential Decision Problems, Econometrica 17, pp. 312-244.

Arrow, K.J., Harris, T., Marschak, 1951: Optimal Inventory Policy, Econometrica 19, pp. 250-272.

Barraquand, J., Martineau, D., 1995: "Numerical Valuation of High Dimensional Multivariate American Securities", Journal of Financial and Quantitative Analysis, vol. 30, no.3, pp. 383-405.

Bellman, R., 1957: *Dynamic Programming.* Princeton University Press, Princeton, NJ.

Bellman, R. and Blackwell, D., 1949: On a Particular Non-zero Sum Game, RM-250, Rand Corp., Santa Monica, USA.

Bellman, R. and Lasalle, J.P., 1949: On Non-zero Sum Game and Stochastic Processes, RM-212, Rand Corp., Santa Monica, USA.

Bhatnagar, S. et al., 1999: A Markov Decision Process Model for Capacity Expansion and Allocation, Proceedings of the 38th Conference on Decision & Control, Phoenix, Arizona, USA.

Black, F and Scholes, M, 1973: The Pricing of Options and Corporate Liabilities. Journal of Political Economics 81, pp 637-659.

Brennan, M. J., Schwartz, E. S., 1985: Evaluating Natural Resource Investments, The Journal of Business, Vol. 58, issue 2, pp. 135-157.

Bilodeau, M.L. and Mackenzie, B.W., 1979: Delineation Drilling Investment Decisions, in Computer Methods for the 80's in the Mineral Industry, A. Weiss, Ed, SME of AIME, pp. 190-201.

Bookstaber, D., 2005: Using Markov Decision Processes to Solve a Portfolio, Allocation Problem. Undergraduate honours thesis, Brown University.

Cortazar, G., Schwartz, E. and Casassus, J., 2001: Optimal Exploration Investments under Price and Geological-Technical: a Real Options Model, R&D management, Vol. 31, issue 2, 181-189.

D'Epenoux, F., 1963: A Probabilistic Production and Inventory Problem. Management Science, 10, pp. 98-108.

Dimitrakopoulos, R., Farrely, C. T., Godoy, M., 2002: Moving Forward from Traditional Optimisation: Grade Uncertainty and Risk Effects in Open Pit Mine Design. *IMM transaction,* the institution of mining and metallurgy, Section A, pp. A82-A88.

Dixit, A.K. and Pindyck, R.S., 1994: *Investment Under Uncertainty.* Princeton University Press, Princeton, USA.

Duffy, D.J., 2006: Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach, Wiley.

Dvoretsky, A., Kiefer, J. and Wolfowitz, J., 1952: The Inventory Problem: I. Case of Known Distributions of Demand, Econometrica 20, 187-222.

Farrelly, C., 2002*: Risk Quantification in Ore Reserve Estimation and Open-Pit Mine Planning.* M.Sc. thesis, University of Queensland.

Godoy, M. C. and Dimitrakopoulos, R., 2004: Managing Risk and Waste Mining in Long-Term Production Scheduling, SME transactions, 316: 43-50.

Goovaerts, P., 1997: *Geostatistics for Natural Resources Evaluation.* Oxford University Press.

Goria, S., 2004: Évaluation d'un projet minier: approche bayésienne et options réelles, Ph.D. dissertation, École des Mines de Paris, Paris, France.

Goto, J., 1999: A Markov Decision Process Model for Airline Meal Provision. CORS- SCRO 2000 Annual conference. Edmonton, Alberta, Canada.

Heyman, D.P. and Sobel, M.J., 1984: *Stochastic Models in Operations Research* McGraw-Hill, New York, Vol. II.

Isaacs, R., 1955: Games of Pursuit, P-257, Rand Corp., Santa Monica, USA

Isaacs, R. 1965: Differential Games, Wiley, New York.

Isaaks, E.H. and Srivastava, R. M., 1989: *An Introduction to Applied Geostatistics*, Oxford University Press, New York.

Isaaks, E.H., 1990: *The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data.* Ph.D. thesis, Stanford University.

Kaelbling, L., Littman, M. and Moore, A. 1996: *Reinforcement Learning: A Survey.* Journal of Artificial Intelligence Research. vol.4. pp. 237-285.

Kiefer, J., 1953: Sequential minimax search for a maximum, *Proc. Amer. Math. Soc.* 4, pp. 502-506.

Laughton, D., 2003: Society of Petroleum Engineers. "The theory and art of asset valuation: Building a case for change." SPE Professional Development Workshop, Banff, Canada.

Le Cam, L., 1990: *More Mathematical People*, edited by Albens, D.J., Alexanderson, G.L. and Reed, C. Harcourt Brace Jovanovich, Boston.

Littman, M.L., Dean, T.L. and Kaelbling, L.P., 1995: *On the Complexity of Solving Markov Decision Problems*. Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence.

Longstaff, F.A., Schwartz, E.S., 2001: "Valuing American Options by Simulation: a simple least-squares approach," Review of Financial Studies, 14, pp. 113-147.

Mayers, S. C., 1984: "Finance Theory and Financial Strategy." Interface 14, pp. 126-137.

Monkhouse, P.H.L., Yeates, G., 2005: *Beyond Naïve Optimisation*. AUSIMM Spectrum Series, Vol.14. pp. 3-8.

Michie, D., 1961: Trial and error. In S. A. Barnett and A. McLaren (eds.) Science Survey. Part 2, Penguin, Harmondsworth, pp. 129-145.

Michie, D., 1963: Experiments on the Mechanisation of Game Learning. 1. Characterization of the model and its parameters. Computer journal, 1, pp. 232-263.

Michie, D., 1974: *On machine intelligence*, Edinburgh University press, Edinburgh.

Minsky, M. L., 1954: Theory of Neural-Analog Reinforcement Systems and its Application to the Brain-Model Problem. Ph.D. thesis, Princeton University.

Puterman, M., 1994: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons.

Ramazan, S. and Dimitrakopoulos, R., 2005: Stochastic Optimisation of Long-term Production Scheduling for Open Pit Mines with a New Integer Programming Formulation in Orebody Modeling and Strategic Planning, Spectrum Series, Number 14, (the Australian Institute of Mining and Metallurgy, Melbourne).

Robbins, H., 1952: Some Aspects of the Sequential Design of Experiments, Bull. Amer. Math. Soc 58, pp. 527-536.

Ross, S., Westerfield, R., Jordan, B., and Roberts, G., 1999: *Fundamentals of Corporate Finance*, third Canadian edition, McGraw-Hill Ryerson.

Samis, M., 2000: Multi-Zone Mine Valuation Using Modern Asset Pricing (Real Options) Techniques, Ph.D. dissertation, Department of Mining and Mineral Processing Engineering, university of British Columbia, Vancouver, BC, 161 p.

Samis, M. R., Laughton, D. and Poulin, R., 2001: "Valuing a Multi-Zone Mine as a Real Asset Portfolio - A Modern Asset Pricing (Real Options) Approach", SSRN electronic paper collection.

Samis, M. R., Davis, G. A., Laughton, D. and Poulin, R., 2006: Valuing Uncertain Asset Cash Flows when there are no Options: A Real Options Approach, Resources Policy, 30: 285-298.

Shapley, L.S., 1953: Stochastic Games, *Proc. Nat. Acad. Sci. U.S.A.* 39: 1095-1100.

Sutton, R.S. & Barto, A.G., 1998. *Reinforcement Learning: An Introduction.* Cambridge, MA: MIT Press.

Thorndike EL (1911) Animal intelligence. New York: Macmillan

Trigeorgis, L., 1991: "A Log-transformed Binomial Numerical Analysis Method for Valuing Complex Multi-Option Investments." Journal of Financial and Quantitative Analysis 26, no. 3, pp. 309-326.

Trigeorgis, L., 1996: Real Options - Managerial Flexibility and Strategy in Resource Allocation, Cambridge, MA, The MIT Press: 427 p.

Widrow, B., and Smith, F. W., 1964: Pattern-recognizing Control Systems. In J. T. Tou and R. H. Wilcox (eds.), *Computer and information science*, Spartan, Washington, D.C., pp. 288-317.

Markov Decision Process (MDP) Toolbox v2.0 for MATLAB,

http://www.inra.fr/bia/T/MDPtoolbox.

Whittle Four X Strategic Planning, Gemcom Software International.