THESE

pour obtenir le grade de

Docteur de l'Ecole des Mines de Paris

Spécialité « **Géostatistique** »

soutenue publiquement

par

Sarah Goria

le 30 mars 2004

Evaluation d'un projet minier : approche bayésienne et options réelles

Directeur de thése : Margaret Armstrong
Co-Directeur de thése : Christian Lajaunie

Jury

Mme	Margaret	ARMSTRONG	Examinateur
MM.	Wynand	KLEINGELD	Examinateur
	Christian	LAJAUNIE	Examinateur
	Christoph	LOCH	Rapporteur
	Eric	PARENT	Rapporteur
	Michel	SCHMITT	Président

Remerciements

Ce travail a été réalisé grâce à l'encadrement de Margaret Armstrong et Christian Lajaunie.

Mes remerciements vont à tous les membres du jury, à Christoph Loch et Eric Parent en tant que rapporteurs, ainsi qu'à Wynand Kleingeld et Michel Schmitt. Je remercie aussi Alain Galli pour ses precieuses remarques.

Je tiens à remercier aussi toutes les personnes du Centre de Géostatistique pour l'aide, les conseils et les encouragements qu'ils m´ont prodigués. Ce travail a également bénéficié de l'atmosphère agréable et conviviale du Centre.

Contents

Résumé

Le problème qui a motivé ce travail est l'évaluation d'un projet minier et en particulier la sélection de la meilleure option de développement incluant la possibilité d'obtenir des sondages d'exploration supplémentaires. Cette thèse porte sur la modélisation des réserves d'une mine d'or à ciel ouvert par des simulations conditionnelles et en particulier une approche bayésienne, et sur l'évaluation financière du projet en utilisant les options réelles.

Dans les projets concernant les ressources naturelles, tels que les mines, il y a au moins deux sources importantes d'incertitude : l'incertitude du marché représentée principalement par le prix du métal (or), et l'incertitude technique ou privée représentée par les réserves. Les méthodes traditionnelles d'évaluation de projets, la valeur actualisée nette (VAN) par exemple, sont basées sur des valeurs fixes pour tous les paramètres et un scénario de développement défini. Elles supposent que le management est passif aux changements de conditions. Les simulations de Monte Carlo peuvent être utilisées pour modéliser les incertitudes liées aux paramètres financiers et techniques, mais elles ont de grandes difficultés à intégrer des scénarios flexibles. De nos jours, il est largement reconnu, en finance et en management (Brealey et Myers, 1991), que les bons managers peuvent et doivent réagir aux changements de circonstances, et que ceci peut valoriser considérablement les projets en cours. Les décideurs ont la possibilité d'agir, car de nombreuses options réelles intègrent un projet d'investissement et ils peuvent choisir de les exercer dans l'intérêt de la compagnie.

Les options les plus fréquentes sont :

- attendre avant d'entreprendre le projet (renvoi à une date définie, ou à une date inconnue) et
- abandonner le projet (temporairement ou de façon permanente).

La méthode des options réelles a été développée à partir des techniques d'évaluation des options financières. Elle a été conçue pour intégrer la flexibilité managériale et les incertitudes sur les prix, mais peu de travaux ont été effectués pour prendre en compte les incertitudes des paramètres techniques. Les questions liées aux incertitudes techniques ne sont apparues que récemment dans la littérature traitant des options réelles, et elles concernent en particulier le pétrole (Chorn et Croft, 1998; Galli et al., 1999; Lund, 1999; Cortazar et al., 2001; Connell, 2002; Dias, 2002; McCarthy et Monkhouse, 2003). Au contraire, l'approche VAN est souvent combinée avec des simulations conditionnelles des réserves (Dowd, 1994 Sanguinetti et al., 1997; Thwaites, 1998), mais sans prise en compte de l'incertitude du prix et de la flexibilité. C'est pour cette raison que nous nous sommes intéressés à la combinaison des simulations conditionnelles géostatistiques avec les options réelles. Tout comme la volatilité dans les modèles traditionnels des options réelles, la sélectivité de la distribution des teneurs ajoute de la valeur au projet.

La première partie de la thèse est consacrée à l'introduction et l'analyse du cadre bayésien et à l'application

au cas d'étude. La deuxième partie de la thèse est consacrée aux options réelles.

Nous avons commencé par utiliser les simulations conditionnelles car elles sont bien acceptées dans l'industrie minière. Toutefois, des hypothèses fortes sur la fonction de covariance sont nécessaires. Ceci est habituellement obtenu en appréciant graphiquement « à la main » la covariance expérimentale et en l'introduisant dans les équations de prédiction comme s'il s'agissait de la vraie covariance. Cette approche est appelée plug- in. Pour tenir compte de cette incertitude, le cadre bayésien a été introduit. Les paramètres du modèle sont traités comme des variables aléatoires. En particulier le cadre bayésien a été considéré à cause de l'information disponible : le gisement est caractérisé par une partie supérieure échantillonnée de façon dense et une partie inférieure peu échantillonnée. Bien qu'on ait supposé la continuité entre les deux parties, l'approche bayésienne a permis de laisser de l'incertitude sur la structure de la covariance. Ceci etait particulièrement interessant dans notre cas pour la structure à grande porteé de la covariance qui décrit la correlation entre la partie superieure et inferieure du gisement.

Un point critique en statistique spatiale par les champs aléatoires gaussiens est l'identification de la structure de la covariance. La méthode de krigeage est largement utilisée pour la prédiction minière car elle permet des prédictions optimales quand la structure de covariance est connue. En réalité la covariance est inconnue et doit être estimée, de telle sorte que l'optimalité du krigeage est mise en question. En outre, la variance de krigeage risque d'être optimiste, puisqu'elle ignore cette source d'incertitude. Le paradigme bayésien fournit un cadre à l'intérieur duquel on peut analyser la performance prédictive du krigeage. La qualité de la mesure d'incertitude attachée au prédicteur est aussi importante que la qualité du prédicteur lui-même. La prédiction bayésienne est basée sur la distribution prédictive complète qui intègre la variabilité des paramètres du modèle.

Le point le plus délicat de l'approche bayésienne est la définition de la distribution a priori des paramètres qui a un rôle important en particulier pour les jeux de données de petite taille. Plusieurs distributions a priori ont été définies, non- informatives ou informatives. Les distributions non- informatives sont souvent impropres : il faut alors vérifier que la distribution a posteriori est bien une distribution propre. La sensibilité des distributions a posteriori des paramètres et des distributions prédictives à ces a priori a été étudiée empiriquement.

Selon la distribution a priori choisie, les distributions prédictives bayésiennes peuvent ne pas être calculables analytiquement, ni simulables directement. Les algorithmes de Monte Carlo par chaînes de Markov doivent alors être utilisés.

Nous avons étudié le modèle gaussien développé par Kitanidis (1986) et Handcock et Stein (1993). L'importance de la prise en compte de l'incertitude sur les paramètres du modèle a été mise en évidence dans plusieurs exemples et dans le cas d'étude. Dans le cas des champs aléatoires gaussiens, l'incertitude sur les paramètres de corrélation influence fortement la distribution prédictive, contrairement à l'incertitude sur le paramètre de variance. Le point le plus critique semble être l'effet de pépite, c'est-à-dire le comportement à petites distances pour lequel il n'y a pas d'information disponible. Bien que l'on attende une variance plus importante dans le cas où tous les paramètres sont inconnus, ce n'est pas toujours le cas.

Il faut noter que la taille des jeux de données en sciences de la terre est généralement importante voir très importante. Il en résulte des temps de calcul longs. Par exemple, les simulations conditionnelles du gisement ont dû être effectuées pendant la nuit, par lots de 20. L'approche bayésienne est même plus lourde encore, car elle nécessite en plus la simulation de paramètres inconnus. De plus l'inversion de la matrice de

covariance est nécessaire pour chaque valeur des paramètres de corrélation.

Comme les données dans le cas étudié ne sont pas gaussiennes, nous avons introduit le modèle gaussien transformé proposé par De Oliveira et al. (1997). La famille Box- Cox pour les transformations de puissance est combinée avec le modèle gaussien. Le modèle gaussien est défini pour la variable aléatoire transformée. Le paramètre de transformation est aussi traité comme un paramètre inconnu. Ce modèle permet de prendre en compte facilement l'incertitude sur la transformation. Mais il est assez difficile à utiliser en pratique car pour chaque valeur du paramètre de transformation, le domaine des données transformées est différent. L'interprétation physique des paramètres est problématique. Box et Cox (1964) et De Oliveira et al. (1997) soulignent la nécessité que les paramètres soient a priori dépendants. Mais il est difficile de définir une telle distribution. De plus, les résultats semblent sensibles au choix de la distribution a priori. Nous remarquons qu'il s'agit de la distribution a priori : elle est mise à jour par les données.

Néanmoins, nous recommandons le modèle gaussien transformé comme premier outil pour analyser la non-normalité des données. Ensuite on pourrait utiliser, soit une simple transformation (la transformation logarithmique, qui peut être utilisée pour des données de distribution statistique manifestement asymétrique, bien que cette hypothèse doive être soigneusement vérifiée), soit, comme l'indiquent Christensen et al., (2001) quand la méthode d'échantillonnage suggère une distribution non gaussienne spécifique, une incorporation dans le cadre du modèle linéaire mixte généralisé. Le modèle gaussien transformé peut aider à décider si, par exemple, la transformation logarithmique est appropriée ou non à la modélisation des données étudiées. Et plus généralement, ceci permet d'éviter la sélection d'une transformation incorrecte due, par exemple, à l'effet d'influence de quelques observations. Nous soulignons que le cadre bayésien fournit le prédicteur optimal en modèle log- normal, alors que ceci n'est pas le cas pour le krigeage sous le même modèle quand la moyenne est inconnue.

Nous avons défini une approche bayésienne plus générale pour évaluer les réserves du cas d'étude. Selon la pratique habituelle en géostatistique, nous avons utilisé la fonction d'anamorphose gaussienne, au lieu de la transformation de Box- Cox, pour transformer les données en valeurs gaussiennes. Les résultats ont été comparés aux résultats obtenus avec l'approche plug- in. Nous rappelons qu'en général les hypothèses pour le cadre bayésien sont plus fortes que pour le krigeage. La distribution spatiale de la variable d'intérêt doit être spécifiée. Par contre, tout comme dans le cas de simulations géostatistiques habituelles, la normalité n'est demandée qu'à une anamorphose près. A la fois une moyenne constante et une dérive quadratique verticale ont été introduites dans le modèle. Une distribution a priori non- informative a été considérée pour les paramètres de la moyenne. Nous remarquons que la distribution a posteriori des paramètres permet de vérifier la présence d'une dérive. Plusieurs covariances ont été fixées a priori et considérées comme ayant la même probabilité. Ceci nous a permis d'inverser la matrice de covariance seulement une fois dans l'algorithme d'échantillonnage. L'anisotropie géométrique a été facilement prise en compte dans le modèle. Pour chaque échantillon de Monte Carlo de la moyenne et de la covariance, une simulation conditionnelle géostatistique a été effectuée. L'importance de la prise en compte de l'incertitude des paramètres de covariance a été mise en évidence. Les distributions bayésiennes et plug-in des réserves diffèrent principalement dans les queues.

En conclusion, nous recommandons l'approche bayésienne, mais elle doit être appliquée avec certaines précautions car elle est dépendante du cas considéré comme, par exemple, pour la définition des distributions a priori. Nous pensons que c'est un point fort de cette approche, car elle force à un examen critique du problème à étudier.

Une partie importante de ce travail est consacrée au choix du modèle de prédiction. Le modèle aurait été plus général si on avait utilisé une famille de fonctions de covariance, au lieu de choisir un modèle spécifique. La classe de fonctions de covariance de Matérn qui inclut le modèle exponentiel aurait pu être envisagée. Ceci aurait permis de prendre en compte des comportements différents à l'origine. Nous n'avons considéré que deux modèles : exponentiel et sphérique.

L'approche bayésienne qui a été utilisée repose fortement sur l'hypothèse gaussienne. Pour les données qui ne semblent pas suivre une distribution gaussienne, mais qui conservent une certaine similarité avec elle, la distribution normale asymétrique (Azzalini et Dalla Valle, 1996) pourrait être choisie à sa place. C'est une distribution permettant une grande flexibilité d'asymétrie et de kurtosis. Lorsque l'histogramme des données présente une distribution continue et légèrement asymétrique, une transformation gaussienne est appropriée. Il est vrai que toute variable continue est « anamorphosable », mais si l'histogramme est trop asymétrique et montre de longues queues, une transformation gaussienne limitera l'amplitude des valeurs fortes et amplifiera les différences peu importantes entre les valeurs faibles. Ceci peut être intéressant si, par exemple, la proportion de zéro est élevé.

La méthode des options réelles a été considérée pour évaluer le projet minier sujet à la fois aux risques du marché et aux risques techniques. Son point fort est qu'elle peut être utilisée pour évaluer la flexibilité inhérente à chaque projet. Négliger la flexibilité peut conduire à une sous-évaluation des projets et par voie de conséquence au refus de projets qui pourraient être profitables.

L'approche risque- neutre a été considérée. On a fait l'hypothèse que le prix de l'or suivait un mouvement brownien géométrique (GBM). Il s'agit d'un modèle non stationnaire. Il est vrai que pour des durées courtes, relatives au temps de retour à la moyenne, il est difficile de différencier un GBM d'un modèle avec retour à la moyenne. Nous avons utilisé la programmation dynamique stochastique pour réaliser les calculs car ceci nous permet d'évaluer les options à exercice anticipé (options américaines). Le modèle binomial (Cox, Ross et Rubinstein, 1979) a été utilisé pour discrétiser le GBM.

De nombreux types de projets, relatifs tant aux ressources naturelles qu'à la R & D dans les industries de hautes technologies, sont menés par étapes séquentielles, et les informations obtenues à chaque étape sont cruciales pour les prises de décision concernant les étapes suivantes de développement du projet. Une option permettant d'obtenir des informations supplémentaires peut s'avérer une source majeure de valeur pour le projet.

Cette thèse aborde la question de l'évaluation de cette option. Dans le cas d'étude, une partie du gisement a été peu échantillonnée et des sondages supplémentaires peuvent être obtenus afin de mieux connaître cette zone. Le projet serait alors mieux défini : la variance des réserves réduite ou le modèle initial corrigé. Le management doit décider si oui ou non l'information supplémentaire justifie le coût. La question à laquelle nous aimerions répondre est quel sera le coût pour nous dans le futur si nous ne disposons pas de cette information maintenant ? Ou comment chiffrer le gain apporté par les nouveaux sondages ? L'investissement nécessaire pour obtenir de nouvelles informations est une alternative importante, tant pour décider d'un développement immédiat, que pour attendre de meilleures conditions de marché. Il comprend des défis pratiques complexes, dans un cadre dynamique prenant en compte l'expiration de l'option de démarrage du développement, le temps d'obtenir l'information et l'interaction avec les incertitudes du marché.

Martzoukos et Trigeorgis (2001) discutent le paradoxe des options réelles concernant l'acquisition de connaissances et d'informations : si les options sont en général des fonctions croissantes de la volatilité tan-

dis que l'acquisition de connaissances diminue les incertitudes, pourquoi alors voudrions-nous avoir plus d'information ? L'acquisition d'information et le moment optimal de cette acquisition entraîne de meilleures prises de décisions et augmente la valeur de l'option réelle. Les deux auteurs s'intéressent au moment optimal d'acquisition d'information.

Alors que pour Marzoukos et Trigeorgis l'apprentissage n'est relié qu'au prix, Dias (2002) prend en considération les incertitudes techniques relatives à la dimension et à la qualité des réserves. Il se focalise sur la sélection de la meilleure alternative pour l'investissement en information. Mais il suppose que des informations supplémentaires ne peuvent que réduire l'incertitude sur les réserves. Dans le présent travail, l'intérêt est mis sur la valeur de l'information supplémentaire, dans le cas où le management décide de l'obtenir; cela valait-il la peine de l'obtenir et a-t-elle aidé au choix du meilleur projet ? Le bénéfice résultant des informations supplémentaires pourrait être une augmentation du retour sur l'investissement prévu lié à une décision plus informée, par rapport à une décision prise sans informations additionnelles. Pour cela, les données supplémentaires ont été simulées de façon à doubler l'information de la partie peu connue, et trois scénarios ont été choisis pour représenter un cas riche, un cas moyen et un cas pauvre.

Les décisions initiales qui ont été considérées sont :

1. démarrer le développement immédiatement,
2. attendre un an sans rien faire,
3. attendre un an pour avoir des sondages supplémentaires.

Lorsque le développement a commencé, les choix disponibles (flexibilité opérationnelle) sont :

1. continuer l'exploitation,
2. abandonner le projet.

Les choix de développement considérés sont :

1. développer une petite fosse,
2. développer une grande fosse,
3. démarrer le développement d'une grande fosse mais revenir à une petite fosse si les résultats ne sont pas prometteurs.

Les résultats ont mis en évidence la valeur de la flexibilité managériale, qui peut être importante.

Le modèle a permis de prendre en compte la flexibilité d'initiation, la flexibilité de l'information et la flexibilité opérationnelle. Dans ce cas la flexibilité d'initiation est importante, tandis que la flexibilité opérationnelle a peu de valeur. On a pu constater, comme l'on pouvait s'y attendre, que plus les réserves sont incertaines, plus la flexibilité est importante. L'importance de l'incertitude des réserves a été aussi mise en évidence par la comparaison des résultats des deux approches Bayésienne et plug-in. Elles ont fournis les mêmes résultats, c'est-à-dire le même projet optimal, tandis que la valeur du projet et la valeur de la flexibilité d'information sont très différentes. Notre principal intérêt s'est mis sur la flexibilité d'obtenir plus d'information. Sur la base de l'information initiale, la décision optimale consistait à retarder le début du projet d'un an, puis à développer une grande fosse. Si l'information additionnelle est riche, elle permet de bien différencier les deux projets : le développement de la grande fosse s'avère bien plus intéressant

que celui de la petite fosse. Si l'information additionnelle est moyenne ou pauvre, la réduction du risque est moins importante car les deux projets sont encore très voisins. L'information additionnelle permet de réduire le risque de 40%, montrant ainsi la valeur de cette information. Une information a de la valeur non seulement si elle entraîne une modification de la décision initiale, mais aussi si elle permet de mieux définir le projet.

La sensibilité de la flexibilité à la valeur initiale du prix de l'or a été vérifiée. La flexibilité d'initiation, c'est-à-dire l'option d'attendre, est intéressante lorsque les prix de l'or sont élevés. La flexibilité d'information a peu de valeur aussi bien dans le cas de prix très bas que de prix très élevés. Ceci indique qu'il n'est pas intéressant de doubler l'information concernant la partie inférieure du gisement, si le prix de l'or est bas, car dans ce cas le développement du gisement n'est pas intéressant, ou s'il est très élevé, car le risque d'un développement non optimal n'est alors pas important.

L'option de réorientation vers la petite fosse si la partie inférieure du gisement n'est pas intéressante augmente de façon importante la valeur du projet.

L'hypothèse risque- neutre est douteuse lorsque les options réelles sont évaluées car un projet minier (ou pétrolier) présente des risques spécifiques qui ne sont pas cotés sur le marché et un portefeuille répliquant le projet est impossible à trouver. Nous avons aussi considéré l'approche décisionnelle en fixant un taux de discount représentant le risque du projet. Les résultats sont naturellement très différents des résultats obtenus avec le taux risque- neutre : la stratégie de développement est différente. Il est donc très important de définir correctement le taux de discount. Une prime de risque qui couvre le risque privé devrait venir s'ajouter au taux sans risque. Les approches décrites par Smith et Nau (1995) et Slade (2000), bien que différentes, utilisent les taux de risque ajustés.

Le modèle a permis de prendre en compte l'incertitude sur les réserves et sur le prix de l'or. Des coûts inconnus ont été considérés pour évaluer la partie inférieure du gisement. D'autres sources d'incertitude auraient aussi pu être inclues dans le modèle, comme par exemple le taux de production et la teneur de coupure. De plus, il pourrait être intéressant d'introduire dans le modèle la corrélation entre la production et le prix de l'or, ou de façon analogue, entre la teneur de coupure et le prix de l'or, ou entre la teneur de coupure et les coûts. Des prix élevés entraînent des teneurs de coupure plus faibles. A des prix élevés, la société minière pourrait vendre plus et diminuer la durée de vie de la mine (Lane, 1991).

Nous avons considéré seulement trois jeux de donnees supplementaires. Les probabilités associées ont été definies en supposant que ces jeux de données representaient 25 %, 50 % et 25 % des jeux de donnees supplementaires possibles.

Nous avons fait l'hypothèse que le calcul des réserves est mis à jour seulement si des sondages supplémentaires étaient réalisés. Il pourrait être intéressant de considérer une mise à jour continuelle des réserves sur la base d'informations provenant de la production. Pour ceci, Connell (2002) introduit un GMB pour modéliser l'évolution des réserves dans le temps. Toutefois, ceci suppose que la variance des réserves augmente avec le temps.

L'approche considérée pourrait prendre en compte les opportunités du marché pour couvrir les risques du projet. Il pourrait être intéressant d'évaluer le cas étudié en considérant que, par exemple, 50 % de la production annuelle sont couverts. Le marché des produits dérivés de l'or a augmenté rapidement au cours de la dernière décennie (Neuberger, 2001). Les compagnies minières couvrent souvent une certaine pro-

portion de leur production prévue, de telle sorte que les institutions financières soient sûres que les dettes seront payées. C'est à dire que les compagnies minières utilisent le marché des dérivés pour couvrir leur production. La principale raison de cette politique est la réduction et le contrôle du risque. En pratique, seulement quelques compagnies vendent plus qu'une faible fraction de leurs réserves prévues. Le management sait qu'une stratégie de vente à terme (contrat forward) qui apparaît comme une gestion prudente du risque lorsque le prix de l'or chute, serait inadaptée en période de forte hausse du prix de l'or. Le producteur ayant une couverture totale ne tire aucun bénéfice de l'amélioration du prix. Une autre possibilité pour la compagnie minière est d'acquérir un put. Les produits dérivés augmentent la flexibilité de la direction des compagnies minières.

Plus d'information concernant la compagnie qui exploite le gisement pourrait aider à mieux définir la règle de décision pour le projet optimal. Nous avons considéré que le décideur avait un comportement neutre vis-à-vis du risque et choisissait donc le projet avec le profit estimé le plus élevé, même si ce projet est plus incertain. Une petite compagnie qui ne veut pas prendre de risques ou qui est conservatrice préfèrera un projet plus sûr à un projet à plus forte valeur mais avec une plus grande incertitude. Au contraire, une grande compagnie qui est impliquée dans un grand nombre de projets n'hésitera pas à choisir le projet avec le meilleur retour sur investissement même s'il est plus risqué. La théorie de l'utilité devrait être utilisée pour l'analyse de la décision dans le cas d'une attitude conservatrice vis-à-vis du risque. La règle de décision est alors de choisir l'alternative qui a la plus grande utilité attendue.

Le caractère novateur de cette thèse tient à ce qu'elle combine des méthodes géostatistiques et des outils de la finance pour évaluer un projet minier. Nous pensons qu'il est important d'utiliser un modèle qui prend en considération à la fois les incertitudes techniques et celles du marché pour l'évaluation de projets relatifs à des ressources naturelles. Le modèle bayésien finalement utilisé est plus général que le modèle gaussien transformée proposé par de De Oliveira et al. (1997). De plus, le domaine de la variable gaussienne est maintenant bien défini et les difficultés liées à la distribution a priori ne sont plus présentes.

Chapter 1

Introduction

In natural resource projects such as mines and oil fields, there are at least two important sources of uncertainties, market uncertainty represented mainly by the price of the commodity, and technical or private uncertainty represented by the reserves. Traditional methods of evaluating projects such as discounted cash flow analysis (DCF) are based on fixed values of all the parameters and a fixed development scenario. They assume that firms are passive to changing conditions. Monte Carlo simulations can be used to model the uncertainty on financial and technical parameters but have great difficulty in incorporating flexible scenarios. Nowadays it is widely recognised in finance and management (Brealey and Myers, 1991) that good managers can and do react to changing circumstances and this can add considerable value to projects. Management has the opportunity to act because many investment opportunities have real options embedded in them, options which management can exercise when it is in the firm's interest to do so. The most common and important real options are

- wait before starting the project (deferral/ postponement),
- abandon the project (temporarily or permanently).

Real options is an approach to evaluating projects that has been developed from techniques used in finance for pricing options. It has been designed to incorporate managerial flexibility and uncertainty on commodity prices, but little work has been done on incorporating uncertainty on technical parameters. Technical uncertainties have only recently become issues in the literature of real options, and concern in particular petroleum (Chorn and Croft, 1998; Galli *et al.*, 1999; Lund, 1999; Cortazar *et al.*, 2001; Connell, 2002; Dias, 2002; McCarthy and Monkhouse, 2003). On the contrary, the DCF approach is often combined with conditional simulations of the reserves (Dowd, 1994; Sanguinetti *et al.* 1997; Thwaites, 1998) but without accounting for price uncertainty. For this we thought of combining geostatistical conditional simulations with real options. Like volatility in traditional real option models, variability of the reserves adds value to the project.

Many types of projects such as natural resources as well as R & D in high tech industries, are carried out sequentially in stages and information obtained is vital in making decisions about the subsequent development of the project. Managers can react to capitalize an upside potential (if the news is good) or to mitigate downside risk (if not). The option to obtain additional information can be a major source of value to the project. Most of the real options literature (Dixit and Pindyck, 1994; Trigeorgis, 1996) has examined the value of flexibility in investment and operating decisions, but little has been written about management's ability to change strategy or acquire information (that is learn).

This thesis addresses the question of how to evaluate this option, using an open cut gold mine as a test case. In this case, additional information can be obtained by drilling extra drillholes. The project would then be better defined: the reserves variance reduced or the initial model corrected. Management has to decide whether the additional information justifies the cost. The question we would like to answer is what will it cost us later not to have this information now? Or **how to value the additional information**? The investment in additional information is an important alternative for both the early development and the waiting for better market conditions. It presents complex practical challenges in a dynamic framework considering the expiration of the option to start the development, the time to learn and interaction with market uncertainties.

Martzoukos and Trigeorgis (2001) discuss the real options paradox of learning and information acquisition: since options are in general increasing functions of volatility whereas learning reduces uncertainty, why would we want to learn? Learning and optimal timing of learning leads to superior decision- making and enhances real option value. Their focus is the timing of learning. While in Martzoukos and Trigeorgis learning is related only with the underlying asset value, Dias (2002) considers technical uncertainty about the size and the quality of the reserve. He focuses on the selection of the best alternative for the investment in information. In this work the primary interest is the value of additional information, that is if management decides to obtain it: was it worthwhile to obtain it and does it help in choosing the best project? The benefit of additional information could be the increased expected payoff from a more informed decision, compared to a decision without additional information.

The first part of the thesis focuses on evaluating the reserves: firstly by using multiple conditional simulations, which, as it was said earlier, is standard practice in mining geostatistics nowadays, and secondly by a Bayesian approach. In the second part of the thesis, real options are used to evaluate the project.

The next section of this chapter introduces the case study and the objective of this thesis. The motivations for a Bayesian approach to spatial prediction and real options to value flexibility are also given. Chapter 2 details the case study. The additional data are simulated and three scenarios are chosen to represent a rich, an average and a poor case. Kriging is carried out as a first approximation of the reserves of both the large and small pits and then conditional simulations are considered. For this the covariance model is specified and the covariance parameters once defined are supposed known. Chapter 3 describes in detail the Bayesian approach to spatial prediction for both gaussian and transformed gaussian random fields. A few examples are discussed and the lower data are analysed to highlight the importance of taking into account the uncertainty on the parameters. Chapter 4 presents the application to the case study to obtain the reserves. The same anamorphosis as for conditional simulations is used to transform the data in gaussian values. This is done to directly compare the results and to underline the importance of the uncertainty on the covariance parameters. For this the covariance model is specified and the covariance parameters are treated as unknown random variables. Chapter 5 introduces real options. It describes and compares contingent claims analysis based on the risk- neutral valuation and stochastic dynamic programming based on an adjusted discount rate as two methods to value real options. The test case is then evaluated using the risk- neutral valuation and highlighting the value of initiation, operating and information flexibility.

1.1 The problem

The Eldorado gold mine is an open pit gold mine. Initially the mining company was planning a small pit about 400m deep, and designed the drilling and sampling program with this in mind. So the drill-holes go from the surface at a height 400m above sea level to sea level. Subsequent drilling suggested that high-grade

mineralisation extends downward to about 200m below sea level. Drilling gave reliable predictions for upper levels but below that the reserves are highly uncertain. More than eighty inclined drill-holes intersect the top part of the deposit; only six were drilled into the lower part. So management has several options for developing the deposit including

1. open up a small pit now, with a sure profit;
2. open up a large pit now, with the chance of a larger profit;
3. first carry out additional drilling which would reduce the uncertainty on the grades in the lower part of the orebody, but would be costly and would delay the start of the project; then choose between a large or small pit;
4. open up a large pit now, with the possibility to revert to a small pit.

Figure 1.1 presents schematically the four options.

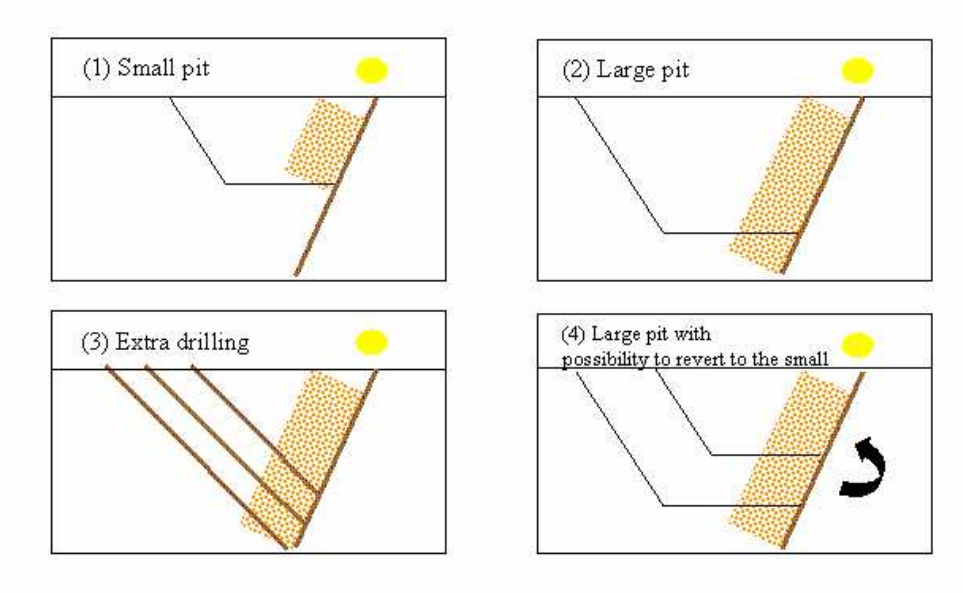


Figure 1.1: *Schematic representation of the development options available to the mine management*

The question is which of these options is the best and in particular, whether it is worthwhile to obtain additional information on the lower part of the large pit or not. Additional information is obtained to reduce the uncertainty on the project value and to optimally delimit the pit to be developed.

Before attempting to answer these questions, we review the standard procedure for defining a mining project. The key steps involved are

1. set up a block model of the deposit by predicting the average grade of each block, for example by kriging. The block size is determined by the mining method used and the size of the equipment, with small blocks being more selective.

2. Set up a family of nested pits corresponding to different cutoff grades, typically using the Lerchs-Grossman algorithm[1]. The innermost pits correspond to the highest revenue area.

3. Choose the economically optimal cutoff grade, for example using the procedure developed by Lane (1991). This presupposes known costs and a fixed metal price.

4. Select the blocks which are above the cutoff grade for mining. To be more precise, their grades are predicted based on the blast hole information[2].

As the purpose of this thesis is to evaluate the impact of infill drilling on a mining project, we are going to make a certain number of simplifying assumptions

- the cutoff grade has been set, once and for all, at 1g/t (this corresponds to a breakeven at $ 250 per ounce of gold).

- Because of the geometry of the deposit, there are only two economically viable nested pits: a large one and a small one (as shown in Figure 1.1).

- The infill drilling will modify the grades of block predictions but will not alter the pit contours.

- As the blast hole data are very closely spaced, there is very little difference between the final kriged block grades and the true grades. That is, we are going to ignore the effect of information on the selection process, and consider a selection on the true block grades.

1.1.1 Evaluation formula

To resolve this decision problem an objective function that depends on both the price and the reserves has to be specified. The term reserves (or recoverable reserves) is used for tonnages of mineralised material which could be ore under certain circumstances.

The objective function specifies the value of alternative management actions and usually accounts for both benefits and costs. For this, once the deposit is divided in blocks of a given size, we consider a selection based on two levels

- the first level defines the project in terms of the blocks that will be exploited. Each project U (U_1 or U_2, that is large or small pit) corresponds to a set of indices, $\mathcal{D}_U$. Let $Z(v_i)$ denote the average grade of block v_i. A project is defined as $\{Z(v_i); i \in \mathcal{D}_U\}$.

- The second level is a selection within each project based on a cutoff. Any block with a mean grade below cutoff is waste and any block with grade above cutoff is ore. For the moment we suppose that free selection applies, that is all blocks above cutoff can be mined irrespective of where they are located[3]. The choice of the cutoff grade between ore and waste is critical. A low cutoff implies that most of the material being developed is treated as ore. The recovery of mineral is high because very little is classified as waste but the average grade of the ore is low. On the contrary, a high cutoff implies that the recovery of mineral is low because the lower grades are classified as waste but the average grade of the ore is high. Intermediate cutoff grades give rise to intermediate positions (Lane, 1991).

[1] Algorithm used in «Whittle 4D»

[2] Blast holes (8- 10 inches in diameter) are drilled into blocks just before mining to receive the explosive charge; drill holes (2 inches in diameter) are drilled from surface to explore the deposit

[3] In reality an open pit mine has to move all blocks of ore and waste: waste blocks are moved and dumped outside the deposit area

If the cutoff grade is z_c then the value of the project U is

$$V_U = \sum_{i \in \mathcal{D}_U} (Z(v_i) - z_c) 1_{Z(v_i) \geq z_c} - C_U = B_U - C_U \tag{1.1}$$

where B_U is the conventional income and C_U is the fixed cost of project U. This expression assumes that z_c is well chosen, that is the selection of blocks above z_c is free and the quantity of metal recovered from a block of grade z_c pays for its marginal mining and processing costs. This hypothesis is rather simplistic but has the advantage of allowing simple and rapid computations.

To obtain V_U the decision to accept or reject block v_i is based on a predicted value $Z^*(v_i)$. This is called the *information effect*. The recoverable reserves should take it into account[4]. As it was said, we will assume that the information effect is negligible.

Linear geostatistics, that is simple or ordinary kriging, cannot be used here. It provides predictions of the average grade of $Z(v)$ or more generally of any linear function of the grade Z, but we need the distribution of $Z(v)$ (or $Z^*(v)$). In non-linear geostatistics, disjunctive kriging, for example, has been developed to estimate functions as $E[1_{Z^*(v_i) \geq z_c}]$. Another approach is that of conditional simulations, which has the additional advantage of providing multiple realisations of the grade distribution and in particular the variability of possible values each block can take.

... the four options

Now, the first option (Figure 1.1) is easy to evaluate. As the upper part of the deposit is densely drilled, the prediction of the reserves can be obtained through kriging (for example, if no cutoff is considered), and the value of this option will depend principally on the gold price. The second and thus the fourth options are inherently riskier, and conditional simulations would effectively add value by providing the histogram of possible reserves, allowing management to decide whether the upside potential justifies the downside risk.

It is much more difficult to evaluate the third option. We have to quantify the effect of additional data before they are obtained. Supplementary observations are obtained to reduce the risk and better identify and delimit the project. The first step in predicting the impact of additional drilling is the definition of the number and location of the extra holes. The next is to generate possible sets of the additional data. We simulated 100 such sets and chose 3 of these to represent an average case, a rich case and a low grade case, as is done in the oil industry, for the lower part of the deposit. We call these artificial sets of data *fictive data* to distinguish them from the real data.

Conditional simulations to each set of additional data are then carried out to evaluate the variability of the reserves of the deposit after additional drilling.

Therefore in order to evaluate the third option we take into consideration the following 2 sets of data

- the data that are available, that we will call the initial data; and

[4]The predicted grades, $Z^*(v_i)$, are used to select blocks above cutoff and thus

$$B_U = \sum_{i \in \mathcal{D}_U} (Z(v_i) - z_c) 1_{Z^*(v_i) \geq z_c}.$$

- the data that could be collected to get a better knowledge of the lower part of the deposit, that we will call the additional (or supplementary) data.

This second data set is defined by the strategy for extra information that is adopted (and could be optimised). Different strategies could be considered differing in the number and location of additional data. In this work we decided to compare two strategies: no additional information (I_0) and doubling the number of samples in the lower zone (I_1). Let $Z(I) = Z_I$ be the additional observations if I is chosen. We want to evaluate the gain due to the additional information and define a strategy to choose between the two projects that depends on the additional information. To summarise, we have two levels of decision: the strategy I and the project U and at least two levels of uncertainty: $Z(I)$ and the distribution of $Z(v)$ (and of $Z^*(v)$ if it was used for selection). The problem under study can be described using a decision tree as shown in Figure 1.2. For simplicity, the random vector $Z(I_1)$ is supposed to take two values, high and low. Analogously, the grades distribution is limited to two values.

The tree is solved backward. Given the formula (1.1) we need to compute for each block v_i

$$\nu_i(z_c) = E[(Z(v_i) - z_c)1_{Z(v_i)\geq z_c}]$$

and

$$\nu_i(Z_I; z_c) = E[(Z(v_i) - z_c)1_{Z(v_i)\geq z_c}|Z_I]$$

which depend on the distribution of $Z(v_i)$ given the initial data or given the initial + additional data.

Let $u^*(Z_I)$ be the optimal project if the result from extra drilling is Z_I,

$$u^*(Z_I) = \arg\max_U E[V_U|Z_I] = \arg\max_U \{\sum_{i\in D_U} \nu_i(Z_I; z_c) - C_U\}$$

and let $b(Z_I) = E[V_{u^*}|Z_I]$ be the associated profit. Given the initial data, that is if I_0 is considered, the best project is

$$u_0 = \arg\max_U E[V_U]$$

and $E[V_{u_0}]$ is its profit. All the distributions we will consider are conditional on the initial data even if it is not always specified. The value of the extra information Z_I is usually considered given by

$$g(Z_I) = E[V_{u^*} - V_{u_0}|Z_I]. \tag{1.2}$$

Information that does not imply a change of the optimal decision is then worthless (if $E[V_{u_0}|Z_I] = E[V_{u_0}]$). This implies that the expected reserves for project u_0 are the same whether they are computed conditional to additional information or not. Observe that only the expected value of V_U, that is the conventional income, is taken into account and not its dispersion. Thus a decision that gives a higher expected profit will be chosen even if the profit is more uncertain. The strategy for additional data is evaluated by

$$l(I) = E_{Z_I}[V_{u^*(Z_I)}] - C_I - E[V_{u_0}] = E_{Z_I}[g(Z_I)] - C_I$$

where E_{Z_I} indicates that the expected value is taken with respect to Z_I and C_I is the cost for additional data. It must be chosen to maximise the expected value of the total profit

$$\begin{aligned} I^* &= \arg\max_I E_{Z_I}[b(Z_I)] - C_I \\ &= \arg\max_I \{E_{Z_I}[\max_U E[V_U|Z_I]] - C_I\}. \end{aligned}$$

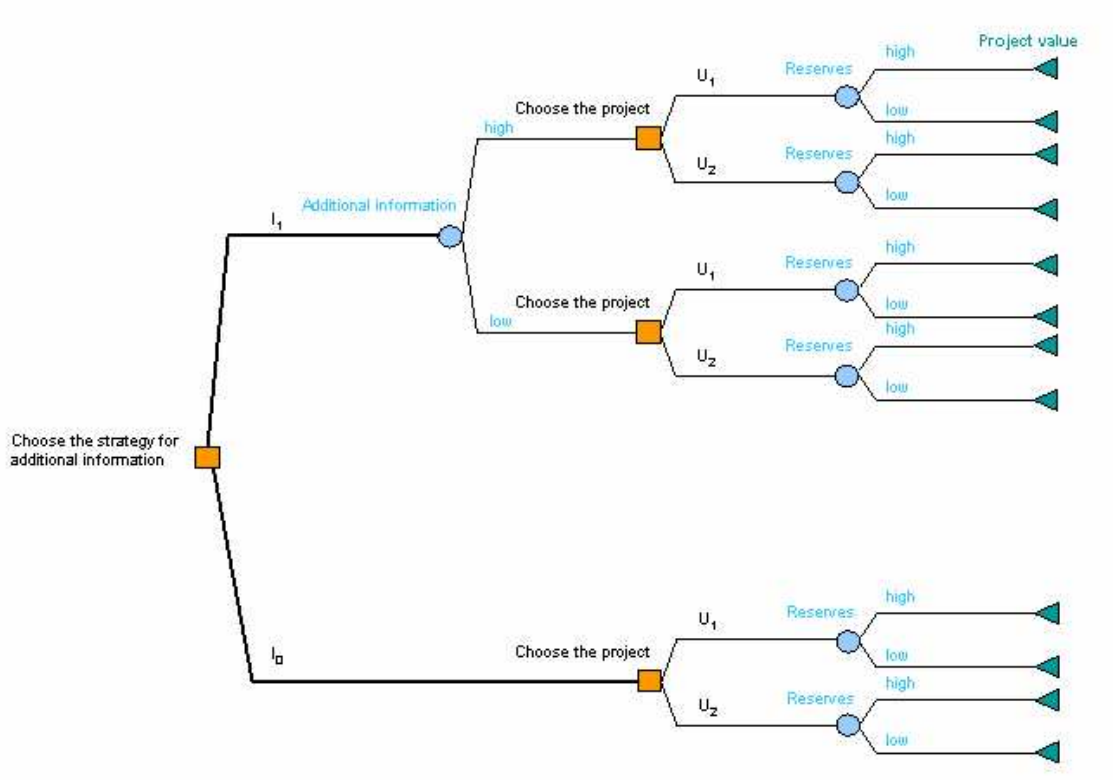


Figure 1.2: *Decision tree: the optimal strategy for additional information (I_1 or I_0) and the optimal project (U_1 or U_2) must be chosen. The squares represent a decision node, the circles represent a random node (realisation of a random variable). For simplicity, only two possibilities at each random node are represented: high and low for the additional information and high and low for the reserves of project U_1 or U_2. The two project values are computed conditional on the initial data or on the initial data + additional information*

Note that only the univariate conditional distributions of blocks are necessary to obtain the expected profit. However, if we were interested in the dispersion of this valorisation function, we would need the joint distribution of blocks.

The objective function given in (1.1) ignores the fluctuations in the commodity price, the sequence of exploitation and the discounting of the revenues which must be considered as well. However, as we will see, there will be no difficulty in introducing other random variables independent of the grades distribution. The revenue from project U is a function of V_U and the commodity price. We will suppose that the reserves and the price are independent. Rigorously, they are not independent as, for example, higher prices could permit to extend the mine's life exploiting with a lower cutoff.

The importance of technical uncertainty in evaluating a mining project is thus clear. Usually in real options papers the revenue from project U is a function of only the commodity price. This can be realistic for projects with no technical uncertainty or with technical uncertainty small compared to price uncertainty.

1.1.2 Valuing the reserves: towards a Bayesian approach

Let $\{Z(x), x \in \mathcal{D}\}$, $\mathcal{D} \subseteq \mathbb{R}^3$, be the random field of interest and $\mathbf{Z} = (Z(x_1), \ldots, Z(x_n))^T$ a set of n observations from a single realisation of this random field, where $x_1, \ldots, x_n$ are known distinct locations in $\mathcal{D}$. Let

$$E[Z(x)] = \boldsymbol{\beta}^T f(x)$$

be the mean of $Z(x)$, where $\boldsymbol{\beta}$ is a vector of unknown regression coefficients and $f(x)$ is a vector of known location-dependent covariates ($f_0(x) = 1$). Let

$$\text{Cov}[Z(x), Z(x')] = \alpha^{-1} K(x, x') \quad \text{for } x, x' \in \mathcal{D}$$

be its covariance function, where $\alpha > 0$ is a scale parameter. We focus on the prediction of the unobserved random variable $Z(x_0) = Z_0$ at a known location $x_0 \in \mathcal{D}$, where Z_0 comes from the same realization as the data vector $\mathbf{Z}$.

If we are only interested in a predictor for Z_0 given $\mathbf{Z}$, kriging can be considered and we do not need to specify the distribution of Z, just its mean and covariance. That is, we focus on the class of predictors that are linear combinations of the data of the form

$$\sum_{i=1}^{n} \lambda_i Z(x_i).$$

The best linear unbiased predictor, $\hat{Z}_0$, is the unbiased linear predictor that minimizes the variance of the prediction error, $\hat{Z}_0 - Z_0$. It is given by

$$\hat{Z}_0 = k_0^T K^{-1} \mathbf{Z} + b_0 (F^T K^{-1} F)^{-1} F^T K^{-1} \mathbf{Z}$$

where $F = \{f_j(x_i)\}_{n \times p}$, $k_0 = \{K(x_0, x_i)\}_{n \times 1}$ and $b_0 = f(x_0) - F^T K^{-1} k_0$. This predictor is known as the universal kriging predictor or the ordinary kriging predictor if $\boldsymbol{\beta} = \beta_0$. It is the optimal predictor under the gaussian assumption. Note that it is obtained supposing the covariance function and its parameters known. These are frequently obtained by fitting the experimental covariance (in the stationary case) "by eye" and the parameters are plugged- in into the prediction equations as if they were the truth.

Kriging provides the best linear predictor of $Z(x_0)$ but to obtain ν_i and be able to choose between the projects we need the entire predictive distribution. For this conditional simulations, which combine kriging and non conditional simulations, are used. They imply carrying out conditional simulations of Z[5]. For each simulation, ν_i is computed and we finally obtain a set of simulations of ν_i. Analogously to kriging to carry out conditional simulations the covariance function and its parameters are fixed. This is why it is called the *plug-in* approach. However, it requires the gaussian assumption.

A Bayesian approach makes it possible to take into account the uncertainty on the drift and covariance parameters. The model parameters are treated as random variables. This is interesting for the case study as the lower part of the deposit is sparsely sampled and the model will thus be defined by the upper data. In addition the uncertainty on the range of the covariance model allows to take into account different hypotheses concerning the correlation between the upper and lower part of the orebody. We start by studying

[5] In particular we need the predictive distribution of $Z(v_i)$. The change of support must then be considered. Support denotes the volume upon which average values may be computed or measured. The available data $\mathbf{Z}$ have a punctual support or their support is small compared to that of blocks which is 10m $\times$ 25m $\times$ 10m. The influence of the support on the distribution of grades is called the *support effect*

the gaussian random field model following Handcock and Stein (1993). As the data of the test case are not gaussian we then introduce the transformed gaussian model proposed by De Oliveira *et al.* (1997). The Box and Cox family of power transformations is combined with the gaussian random field model. The gaussian model is defined for the transformed random variable. The transformation parameter is also treated as an unknown parameter. We will see that this model is quite difficult to use in practice as for each value of the transformation parameter the range of the transformed data is different. Moreover, Box and Cox (1964) and De Oliveira *et al.* (1997) suggest that a priori the parameters must be dependent. As we will see, such a prior is difficult to define.

The most delicate point of the Bayesian approach is the definition of the prior distributions for the unknown parameters as they are unobservable. Several prior distributions for the parameters are considered and the sensitivity of the posterior and predictive distributions is checked. This is important as it can happen that the model may be poorly identifiable from the available data and the choice of the prior may have a strong influence on the results. The Bayesian predictive distributions cannot always be computed analytically and they cannot be simulated directly. Markov Chain Monte Carlo (MCMC) algorithms must be used. In the past few years Bayesian analysis has seen a real growth of its applications, in particular in spatial prediction problems in epidemiology, air pollution, toxicology, image analysis. This is mostly due to the development of MCMC.

Finally, a more general approach to evaluate the reserves is defined. As it is usual practice in geostatistics the gaussian anamorphosis function is considered to transform the data in normally distributed values. Both a constant mean and a vertical drift are specified. As the number of observations is large, the type of covariance function is chosen while its parameters are let unknown. As for conditional simulations the gaussian hypothesis is needed for a transformation of Z. Classic geostatistical conditional simulations are carried out but now without a fixed covariance model. The two approaches differ in the treatment of unknown parameters.

The Bayesian and *plug-in* predictive distributions are compared. They differ mostly in the tails. This is important as we work with a cutoff.

Moreover, the Bayesian approach allows us to

- consider a random drift instead of an unknown drift;
- update the information on the unknown parameters by extracting information contained in the observations;
- incorporate external but related information in the estimation process.

1.1.3 Valuing flexibility: towards real options

The following example, which is standard in the real options literature (Trigeorgis, 1996), highlights the importance of taking account of management flexibility to react to potential losses when valuing an investment project. A company wants to invest in a mine that costs $ 100 million, of which $ 10 million is needed at the start of the project. The mine will produce a return of either $ 180 million or nothing with equal probability. The classic approach based on the net present value (NPV) would give an expected NPV of $ -10 million and the investment would be refused (Figure 1.3). However, the company can simply drop the project if an exploratory hole is barren (that is, there is no gold to be found). If exploration shows promising results, further investment may be less risky. The outcome may be between a loss of 10 and a profit of 80. The

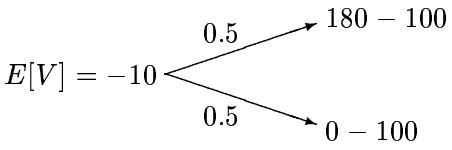


Figure 1.3: *Value of the project (in $ million) with classic NPV approach. The expected NPV is negative: the project is refused*

expected NPV is $ 35 million: this project is then worthwhile pursuing (Figure 1.4). Flexibility enhances the project value.

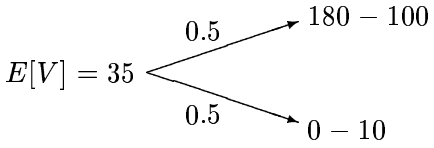


Figure 1.4: *Value of the project (in $ million) with NPV approach but taking into account flexibility. The expected NPV with flexibility is positive: the project is undertaken*

Analogously, we suppose that management reacts optimally to changing conditions of price and reserves when evaluating the case study. To value flexibility a decision theoretic approach is used.

A decision- making problem involves a temporal sequence of decisions, each alike in kind, but where the optimal action at each decision point may depend on time and the system state. The initial decisions we consider are

1. start immediately,
2. wait one year passively,
3. wait one year to acquire new information by extra drilling,
4. abandon the project.

Once the project is started the available option is

1. abandon the project.

Our goal is to develop a decision rule that guides management actions for each time and system state to be optimal for the objective function. The analysis of such decision problems requires specification of

1. predictive models of system dynamics formulated in terms of quantities relevant to management. for the reserves Conditional simulations or the Bayesian approach that was briefly introduced in the previous section are used for the reserves; a geometric Brownian motion is defined for the gold price;

2. an objective function for evaluating alternative management strategies. The expected NPV under risk-neutral probability is used to value the project. It accounts for management flexibility.

The risk- neutral price process is used and cash flows are discounted at the risk- free rate. Stochastic dynamic programming is used to carry out the computations because it allows us to value early exercise options. Although Monte Carlo simulations are carried out for the reserves, a discrete approximation of the geometric Brownian motion is considered for the gold price as it permits us to visualize the sequential problem without the burden of carrying out the simulations. Moreover, it helps highlighting possible management responses to changes in price and reserves. The Monte Carlo reserves allows us to define possible mine- types with an associated probability of occurrence (Cortazar *et al.*, 2001). Thus the project is evaluated for each simulated value of the reserves, that is the Monte Carlo value of the mine is obtained.

We are interested in the value of additional information. We assume that additional drilling only, if obtained, may modify the initial distribution of the reserves and that no information is provided by the production phase. Note that information on the reserves in a mine (or reservoir) develops over time. Hence the production management can learn more about the mine (or reservoir) behavior over time. Such learning structures should be incorporated in the model to provide solutions in the form of strategies. In this case technical risks evolve in time: the integrated approach by Smith and Nau (1995) or, if these risks seem important compared to market risks, the stochastic dynamic programming approach with an exogenous discount rate as in Lund (1999) could be used.

Other sources of uncertainty could also be included in the model. Costs are assumed unknown to value the lower part of large pit. However, having no information on the company exploiting the deposit we finally kept them constant to value the 2 projects.

We do not take into account market opportunities to hedge the project risks. Hedging eliminates the risk of fluctuating prices, but, if forward selling is considered, also means limiting the opportunity for future profits should prices move favorably.

Chapter 2

The case study: the Eldorado gold mine

The Eldorado mine is an open pit gold mine. About 90 inclined holes were drilled in the orebody and 3205 samples were analysed. The strike of the deposit is about $15°$ east of north, it dips at about $45°$ to the west and plunges northward at about $30°$. On the eastern side the deposit is cut by a fault so there is no mineralisation below this. The structure of the mineralisation is expected to be parallel to this plane.

In Section 1 the data are described. In particular, the hypothesis of stationarity is checked. In Section 2 the additional information needed to evaluate the third option is defined through conditional simulations. In Section 3 ordinary kriging is carried out to define the envelope containing the orebody. In Section 4 conditional simulations are carried out and in Section 5 conclusions are presented.

2.1 Stationarity

The orebody is limited to the right by a fault. The mineralisation lies along this fault. It does not continue up to the surface. There is waste above it as it can be seen from the basemaps presented in Figure 2.1 (with low grades shown in yellow and higher grades shown in darker colours). It has to be mined to access the ore. We will estimate this waste separately. Firstly we outlined the mineralised zone. The basemaps are presented in Figure 2.2. The data selected seem homogeneous. Unless explicitly stated in this work we will always refer to these data.

An extrapolation hypothesis is needed to decide whether the upper data inform us on the distribution of grade in the lower part or not. There exist several extrapolation hypothesis. At one extreme we could treat the data in the two parts of the deposit as if they were independent, that is the upper data do not inform us on the lower part, at the other we could suppose that the same random function is observed in both parts.

The first step was to determine whether the deposit can be considered as being stationary (second- order stationarity[1]). For this the mineralised zone was divided into a grid and the statistics were calculated for grid squares. As no significant trends were found, either perpendicular to the fault or parallel with it, this zone can be treated as being stationary. Figure 2.3(a) presents the average grade for each depth. Figure 2.3(b) presents the average grade for different levels of the coordinate in the eastern direction. Looking at the first graphic it can be noted that the average grade of the upper part is higher than the average grade of the lower part. A drift that takes this into account could be considered instead.

[1] A random function with a constant mean and a two-point covariance function that depends only on the distance between the two points is called second- order stationary

We assume that all the data are realisations of the same random function Z with a constant mean and a fixed covariance function[2]. It is a strong hypothesis. We note that to carry out kriging we just need to specify the first two moments of the random function.

2.1.1 The raw data

Table 2.1 gives the statistics of the 3205 samples together with the statistics of the 2390 samples inside the mineralised zone. The two histograms presented in Figure 2.4 differ only in the number of low values which depends on the selection considered. It can be noted that the mean grade in the upper part is higher than in the lower part. Looking at these values it is reasonable to think that the mineralisation extends in the lower part.

Variable	N	Minimum	Maximum	Mean	Standard deviation
Grades	3205	0.01	21.20	1.18	2.10
Grades	2390	0.01	21.20	1.51	2.32
Upper Part	2244	0.01	21.20	1.52	2.35
Lower Part	146	0.01	9.19	1.36	1.73

Table 2.1: *Statistics of all the grades and of the grades inside the mineralised zone*

Variogram model

Experimental variograms[3] were calculated in the three main directions of the orebody: the two directions parallel to the plane of mineralisation, d_1 and d_2, and the direction perpendicular to the plane of mineralisation, d_3. Looking at the experimental variograms in Figure 2.5 we can assume the same behaviour for the directions parallel to the plane of mineralisation but not for the perpendicular direction. We see that the range in d_3 is shorter. This can also be seen in the Basemaps (Figure 2.2). The experimental variograms tend to the same sill. This anisotropy can be modelled as a geometric anisotropy. Physically this corresponds to a rotation and stretching of the original spatial coordinates which then allow to revert to an isotropic model. The parameters corresponding to the anisotropy angle and to the anisotropy ratio must be added to the model: generally three rotation angles, needed to align the coordinate system with the principal directions, and two stretching parameters, needed to equalize the correlation length in all 3 directions. Let

[2] If the random function Z is stationary, its mean $m = E[Z(x)]$ is constant and its covariance function

$$K(h) = E[Z(x) - m][Z(x+h) - m]$$

depends only on the distance h. The lag h is a vector, that is the covariance depends on both its length and on its direction. The relation between covariance and variogram is

$$\gamma(h) = K(0) - K(h)$$

where $\gamma(h)$ is the semi-variogram. When the covariance depends only on the length of the distance it is said to be *isotropic*. That is there is no reason to distinguish one direction from another. The Euclidean distance will be considered

[3] The experimental variogram is given by

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i,j \in D(h)} (Z(x_i) - Z(x_j))^2$$

where $D(h) = \{(i,j) : |\mathbf{x}_i - \mathbf{x}_j| = h\}$ and $N(h)$ is the number of pairs in $D(h)$

$\mathbf{x} = (x_1, x_2, x_3)$ be the coordinates of the samples in the original directions, that is eastern, northern and depth and h the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. Let $\mathbf{x}' = (x'_1, x'_2, x'_3)$ be the coordinates in the three principal axes of the orebody and h' the Euclidean distance between $\mathbf{x}'_i$ and $\mathbf{x}'_j$. Thus first we need to transform the coordinates $\mathbf{x}$ into coordinates $\mathbf{x}'$, $\mathbf{x}' = T\mathbf{x}$ where T_{ij} is the cosine of the angle formed between the positive semiaxes x'_i and x_j. Then as in the plane parallel to the plane of mineralisation isotropy is assumed, we need to compute the ratio, k, between the major range of the ellipse that is for d_1 and d_2 and the range for the perpendicular direction d_3 ($k < 1$). The anisotropic model is given by

$$\gamma'(h) = \gamma\left(\sqrt{(h'_1)^2 + (h'_2)^2 + k^2(h'_3)^2}\right)$$

where γ is the isotropic model. A variogram model consisting of a nugget effect[4] plus an exponential was fitted. The isotropic exponential model has the general form

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ \alpha^{-1}\left(1 - \exp\left(-\frac{h}{\nu}\right)\right) & h \neq 0 \end{cases}$$

with sill $\alpha^{-1} \geq 0$ and scale parameter $\nu > 0$. The exponential model is continuous but not differentiable at the origin. The exponential variogram reaches its sill only asymptotically when $h \to \infty$. It reaches 95 % of the sill at about 3ν. This is called the practical range. When the variogram reaches a limiting value, the sill, this means that there is a distance beyond which $Z(x)$ and $Z(x+h)$ are uncorrelated. Table 2.2 shows the sills of each structure and their ranges in the three main directions.

	Sill	Range 1 (m)	Range 2 (m)	Range 3 (m)
Nugget effect	1			
Exponential 1	4.8	60	60	40

Table 2.2: *Variogram parameters of the fitted model (the main directions of the orebody are obtained through a rotation of (-165, 45, 150) that is indicated in the geological convention, relative to east)*

2.2 Generating grades on the fictive holes

The additional information is simulated.

When conditional simulations were first developed in the seventies, their main uses were mine planning and grade homogeneisation (Deraisme, 1977; Maréchal and Shrivastava, 1977). Simulations are still widely used for this (see for example Dowd, 1994; Sanguinetti *et al.*, 1997; Sahin and Fuseni, 1998). In a similar vein, simulations have also been used to compare different sampling patterns. For example, Kleingeld *et al.* (1997) used a Cox process to simulate the diamond distribution in marine placer deposits in order to test the efficiency of different sampling campaigns and to assign confidence intervals to block predictions.

By the early nineties, the emphasis had moved to evaluating mining projects. In his review paper, Ravenscroft (1994) distinguished between two types of uses of simulations

[4]The behavior of the variogram near the origin is linked to the continuity and to the spatial regularity of the regionalized variable. If the variogram is not continuous at the origin, we have a nugget effect that can be due to microvariability that is variability at a scale smaller than the sampling support or to measurement errors

- modelling local short-scale variability in a way not provided by any interpolation technique (applications listed in the preceding paragraph);
- generating alternative images of the deposit to allow for sensitivity and risk analysis.

Since then, risk analysis has been one of the leitmotivs of simulation papers. Analogously, we thought of simulating the additional information to generate alternative scenarios for this additional information. Several authors have focused on the interaction between geological uncertainty and openpit optimisation (Biver *et al.*, 1997; Thwaites, 1998; Coombes *et al.*, 2000).

Nowadays multiple conditional simulations are routinely used for evaluating projects in both mining and petroleum. When multiple simulations are used for volumetrics in the oil industry, it is standard practice to run several hundred reservoir simulations and then rank the simulated volumes in descending order and select three typical cases: a pessimistic case (10% probability of being below it), the most likely case (the 50% quantile) and an optimistic case (only 10% of exceeding this). These are denoted as P10, P50 and P90. One key difference between oil and mining is that ore reserves depend on two variables, grade and tonnage, rather than one.

In the 1980s the oil industry started to introduce probabilistic reserve definitions in addition to the usual *proved* and *probable* categories. Murtha (2001) explains what is meant by P10, P50 and P90: «P10 is the point that would split the area under the curve into 10% and 90% of the total area, meaning that there is only a 10% chance that the reserves are less than the P10 value». As the distribution of the reserves is obtained from a Monte Carlo simulation model, this means that 10% of the realisations in the model are less than the P10 value and 90% are greater. This requires being able to rank the results of the MC simulations in an unambiguous way.

The procedure used to define the new drill-holes can be divided into 4 steps

1. choose the number of fictive holes and their locations;
2. run 100 point simulations to assign a value to these drill-holes;
3. calculate the average grade of fictive plus real drill-holes for each simulation for the lower part of the ore-body;
4. find the three cases P10, P50 and P90.

Number and location of the fictive holes

Only six drill-holes intersect the lower part of the deposit. We decided to double the number of holes in that part of the deposit by adding six fictive holes located about half way between the real ones. Table 2.3 describes how we located these new drill-holes relative to the ones already existing: it was the quickest way to build them. Figure 2.6 highlights the location of these drill-holes.

2.2.1 Conditional simulations to inform the fictive holes

One hundred conditional simulations were run to inform the points on the six fictive drill-holes. Conditional simulations of a gaussian random function are carried out in the two following steps: first a non conditional

	Initial drill-hole	y_1	y_2	y_3
1	87	x_1	$x_2 - 65$	x_3
2	76	x_1	$x_2 - 60$	x_3
3	76	x_1	$x_2 + 30$	x_3
4	76	x_1	$x_2 + 70$	x_3
5	89	x_1	$x_2 - 50$	x_3
6	90	x_1	$x_2 - 50$	x_3

Table 2.3: *Location of the fictive drill-holes (where x_1, x_2, x_3 are the coordinates (in meters) of the initial drill-holes and y_1, y_2, y_3 are the coordinates (in meters) of the new ones)*

simulation is obtained, it is then conditioned using ordinary kriging. For the non conditional simulations turning bands are used. We do not describe here the turning bands algorithm (see Matheron, 1973; Chilès and Delfiner, 1999). This approach gives realisations of a gaussian random function. *Conditioning by kriging* is described in Appendix A.

The histogram in Figure 2.4 shows clearly that the data are not gaussian. We use an anamorphosis function to transform the data into normally distributed values, we then work with the gaussian variable, and finally back-transform the results to get them in the initial random variable space. We assume that grades after anamorphosis are gaussian.

The anamorphosis function is a one to one monotone mapping of the raw data to the standard gaussian values based on the cumulative distribution function. Let φ denote the anamorphosis function, $Z = \varphi(Y)$. This function associates each point z of the histogram of Z to a point y of the histogram of Y, such that $P[Z \leq z] = P[Y \leq y]$, that is it puts into relation the equivalent quantiles of the distributions F and G of the random variables Z and Y, respectively. It can be written $z = \varphi(y) = F^{-1} \circ G(y)$. The transformation function φ can be defined graphically from the empirical distribution.

The anamorphosed grades

To carry out the simulations raw grades had to be transformed to normality. The statistics of the gaussian variable are presented in Table 2.4. The transformation was obtained using a gaussian anamorphosis function.

Variable	N	Minimum	Maximum	Mean	Standard deviation
Gauss	2390	-2.05	3.56	0.00	0.99

Table 2.4: *Statistics of the gaussian values inside the mineralised zone*

We used the linear interpolator inversion letting the gaussian variable range from -10 to 10 and the gaussian inverse variable range from 0 to 30g/t (compared to a maximum sample of 21.20). Eighty Hermite polynomials were used to define the anamorphosis. Figure 2.7 shows the anamorphosis function with gaussian values along the X axis and actual values along the Y axis.

Variogram model

We calculated the experimental variograms of the anamorphosed grades in the three main directions of the orebody (with an angular tolerance of $35°$ and a lag-value of 25m for the directions parallel to the plane of mineralisation (d_1, d_2) and of $45°$ and a lag-value of 5m for the direction perpendicular to it (d_3)) and then fitted a model (Figure 2.9). This fits the experimental variograms quite well at least for distances up to 200m. The parameters of this model are given in Table 2.5. It is a nested variogram. This variogram model was used to generate 100 conditional simulations to inform the fictive holes.

	Sill	Range 1 (m)	Range 2 (m)	Range 3 (m)
Spherical	0.1	5	5	5
Exponential 1	0.7	100	100	60
Exponential 2	0.2	350	350	100

Table 2.5: *Variogram parameters of the fitted model of grades after anamorphosis (the main directions of the orebody are obtained through a rotation of (-165, 45, 150) that is indicated in the geological convention, relative to east)*

The points were simulated on a 5m $\times$ 5m $\times$ 5m grid. The parameters of this grid are in Table B.1. Given the size of the data set to allow the computations a neighbourhood has to be used that includes only a subset of the data for the prediction (that is for the kriging) of each grid node. The neighbourhood was set as an ellipsoid with ranges of 200m for d_1 and d_2 and of 20m for d_3 to take account of the anisotropy. The range of d_3 is much smaller than that for d_1 and d_2. This choice is given by the supposed structure of the orebody. Moreover, as it can be seen in Figure 2.9, at a distance of 20m for d_3 we already have the 70% of the total variation. And, as there are zones with few samples points, in particular for the lower part of the orebody, we had to define large ranges for d_1 and d_2. The simulated values were assigned to the samples locations of the fictive drill-holes that were closest to them. The mean grades of the real + fictive drill-holes for the lower part of the mineralised zone were calculated for each of the 100 simulations (Figure 2.10). Three typical cases were selected to represent the P10, P50 and P90 reserves. Table 2.6 gives the statistics for the real plus fictive drill-holes for the upper + lower part and for the lower part, for these 3 simulations together with the corresponding statistics before adding the fictive drill-holes. We now have 2642 data: 2390 real data and 252 fictive data.

Figure 2.11 presents the average grade for each depth for each new data set.

2.3 Recoverable reserves

For each pit we need predictions of the reserves. As we have seen in Chapter 1, for each block v we look for the recoverable reserves at cutoff z_c,

$$(Z(v) - z_c)\, 1_{Z(v) \geq z_c}.$$

This function is called the conventional income. It is of considerable economic importance. It allows us to decide whether block v is economically interesting. It is assumed that the selection of blocks above z_c is free and that z_c is chosen so that the quantity of metal recovered from a block of grade z_c pays for its marginal mining and processing costs. In particular, after having defined the mining blocks, we look for

Upper + lower part					
	N	Minimum	Maximum	Mean	Standard deviation
Initial 90 holes	2390	0.01	21.20	1.51	2.32
Rich (simu12)	2642	0.01	21.20	1.54	2.46
Average (simu46)	2642	0.01	21.20	1.50	2.33
Poor (simu66)	2642	0.01	21.20	1.47	2.27

Lower part					
	N	Minimum	Maximum	Mean	Standard deviation
Initial 90 holes	146	0.01	9.19	1.36	1.73
Rich (simu12)	323	0.01	20.16	1.87	3.20
Average (simu46)	323	0.01	16.72	1.54	2.29
Poor (simu66)	323	0.01	9.96	1.26	1.72

Table 2.6: *Statistics of the real + fictive data and of the real + fictive data in the lower part of the deposit (of the 252 new data 177 are in the lower part)*

- the ore tonnage (that is the number of blocks above cutoff)

$$\sum 1_{Z(v_i)\geq z_c};$$

- and the metal tonnage

$$\sum Z(v_i)\, 1_{Z(v_i)\geq z_c}.$$

The grades $Z(v)$ of the blocks are not known. The objective is to predict for all the blocks of the two pits

- the ore tonnage

$$T(z_c) = E[1_{Z(v)\geq z_c}] = P[Z(v) \geq z_c]$$

- and the average recovered grade

$$m(z_c) = E[Z(v)|Z(v) \geq z_c].$$

Note that the metal tonnage is given by $m(z_c)T(z_c)$ and the conventional income is given by $(m(z_c) - z_c)T(z_c)$. These functions are called *recovery functions* or *selectivity curves*. If the distribution of block grades was known we could predict all of these. Note that when the selection is made, $Z(v)$ is not known exactly. So the selection is made on predicted values, that is the recovered average grade is

$$E[Z(v)|Z^*(v) \geq z_c].$$

The predictions of recoverable reserves should take this information effect into account. As it was said in the previous chapter we suppose that the information effect can be neglected: the final predictions are obtained using blast hole information and can be assumed very close to the true grades.

Ordinary kriging is not considered. Kriging is a linear predictor. We need the distribution of blocks and not just their predicted values. One alternative is to consider multiple conditional simulations and compute for each simulation the quantities of interest. Another alternative is to predict directly for each block v its

distribution function, $P[Z(v) \leq z_c]$. This function can be predicted by disjunctive kriging. To approximate the tonnages and grades that might really be recovered we will consider conditional simulations. They give us a measure of the spread of values around predictions.

Kriging was used to define the envelope containing the orebody.

Pit contours

We have to set up the contours for the 2 pits. In industry, softwares such as «Whittle4D» are used to determine the outlines of the pits so as to maximise the profitability while respecting the slope stability constraints. Because we do not have access to this software we have chosen a simplified procedure for constructing the pit contours level by level. This was done through kriging conditional on the original data. Figure 2.12 presents the contours of the two pits for a given level of x_1.

2.4 Conditional simulations

Following Journel (1977) the objectives of kriging (predictions) and simulations differ, as

- a predictor $Z^*(x)$ gives a prediction that is the closest as possible to the true unknown grade $Z(x)$. The predictor is chosen so that it is unbiased and minimizes the mean squared error, $E[Z(x) - Z^*(x)]^2$. But these predictors have no reason to reproduce the spatial variability of the true grades. The minimisation produces a smoothing of the dispersions.

- On the contrary a conditional simulation reproduces the empirical distribution of the true grades, that is it identifies the principal caracteristics of dispersion of the true grades. But a simulated value of x is not the best predictor of $Z(x)$.

We want to simulate block grades given punctual data. The change of support must be considered. Predicting $1_{Z(x)\geq z_c}$ for a point support is not equivalent to predicting $1_{Z(v)\geq z_c}$ for a block. The change of support from point to block has a large impact on the quantity of ore above cutoff as it can be seen in Figure 2.8. The dispersion of grades of larger and larger blocks decreases. There are several models to take the change of support into account: the affine correction, the discrete gaussian model and the averaging of points in a block (see Matheron, 1978; Rivoirard, 1994).

We will consider the averaging of points in a block, which is the most straightforward and extensively applied. Simulations are carried out on a fine grid that is smaller than the block size. Then the simulated values that fall within a given block are averaged to obtain the block value. Each block v is discretised in n points x_i, $i = 1, \ldots, n$, and the true block grade, $Z(v) = \frac{1}{|v|}\int_v Z(y)dy$, is assimilated to the arithmetic mean of the n punctual grades, $Z^*(v) = \frac{1}{n}\sum_i Z(x_i)$ and it is this mean that we simulate. The discretisation of block v or the number n of points taken in v depends on the error $Z(v) - Z^*(v)$ we can accept. As Journel says, the variance of this error should not mask the structural variability of the real block grades. In practice it is sufficient to make sure that $E[Z(v) - Z^*(v)]^2$ is small compared to the nugget effect corresponding to v. If the block is large, the nugget effect is almost zero, it must be checked that $E[Z(v) - Z^*(v)]^2$ is much smaller compared to the prior variance of the dispersion of the grades $Z(v)$. An advantage of this approach is that no hypothesis is made on the law of the random function $Z(v)$. Simulations are based only on the available information. The disadvantage of this approach is that the number of points to estimate can become enormous.

A total of 4 sets of 50 simulations were run using, as for kriging, the following sets of conditioning data

- the original real data;
- the high grade valued data;
- the average grade valued data;
- the low grade valued data.

First information on the grades after anamorphosis (i.e. statistics and variogram model) is given. The results from conditional simulations are then presented. Finally comments on the simulations results are given.

2.4.1 Statistics of anamorphosed grades

To carry out the simulations the data have been transformed to normally distributed values beforehand. A different anamorphosis transformation was used for each set of data. Table 2.7 presents the statistics of these transformed values. As expected, they do not vary in the upper part but they do in the lower part.

Upper + lower part					
	N	Minimum	Maximum	Mean	Standard deviation
Rich (simu12)	2642	-2.13	3.62	0.00	1.00
Average (simu46)	2642	-2.14	3.58	0.00	1.00
Poor (simu66)	2642	-2.12	3.62	0.00	1.00

Lower part					
	N	Minimum	Maximum	Mean	Standard deviation
Rich (simu12)	323	-2.13	3.25	0.00	1.12
Average (simu46)	323	-2.14	2.93	0.00	1.02
Poor (simu66)	323	-2.12	2.24	-0.07	0.97

Table 2.7: *Statistics of the real + fictive gaussian values in the upper + lower part and in the lower part of the deposit*

Variogram models

The experimental and variogram models were then computed. Figure 2.13 presents the variograms of the anamorphosed grades for the two data sets rich and poor. The two variogram models have the same structures and the difference in the parameters is small. Adding new data changes the anamorphosis function and the variograms, but in this case we added only 252 data with respect to 2390 real data.

2.4.2 Simulations results

Blocks of size 10m $\times$ 25m $\times$ 10m were defined. The grid parameters are given in Table B.2. The neighbourhood is the same as the one used for simulations. The discretisation of the blocks is of $2 \times 5 \times 1$. Figure 2.14 presents simulated grades for two different depths. Table 2.8 presents the tonnage and Table 2.9 presents the

recovered average grade with a cutoff of 1 g/t. Adding the new data may increase the variance of the recovered grade as it can be seen for the rich and average additional information. On the contrary the variance of the tonnage decreases. This is particularly evident for the rich additional information. As expected, the average number of blocks above cutoff, in the lower part of the large pit, increases when adding new information.

Given the initial data, the large pit is expected to contain 326 tons of ore, while the small pit 238 tons of ore. Given the additional information, the large pit contains 351, 337 and 324 tons of ore while the small pit contains 246, 240 and 238 tons of ore for the rich, average and poor information, respectively.

Initial 90 drill-holes								
	N blocks	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	55259	17053	23780	**21028**	1269	19358	21181	22476
Lower part	11079	3229	6055	**4697**	599	3938	4776	5379
Small pit	41821	13470	18108	**15788**	958	14743	15776	16876

Initial 90 drill-holes + rich fictive data								
	N blocks	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	55605	18983	24258	**21298**	1012	20038	21389	22278
Lower part	11322	3785	5996	**4981**	486	4379	5024	5511
Small pit	41831	14105	18151	**15792**	802	14804	15740	16667

Initial 90 drill-holes + average fictive data								
	N blocks	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	55605	18527	25159	**21410**	1159	19864	21354	22472
Lower part	11322	3904	5897	**5006**	445	4360	5033	5519
Small pit	41831	13527	18625	**15868**	946	14726	15917	16832

Initial 90 drill-holes + poor fictive data								
	N blocks	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	55605	19066	24059	**21179**	1219	19815	20880	23009
Lower part	11322	3850	5958	**4831**	525	4227	4778	5543
Small pit	41831	14261	18129	**15805**	922	14675	15721	16955

Table 2.8: *Simulations results: tonnage (with a cutoff of 1g/t) for the whole large pit, lower part of large pit and small pit. The number of predicted blocks is 55259 for the large pit and is 41821 for the small pit. Given the additional information the number of predicted blocks is 55605 for the large pit and is 41834 for the small pit. The size of the blocks is 10m $\times$ 25m $\times$ 10m*

Initial 90 drill-holes							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.82	3.37	**3.10**	0.13	2.96	3.09	3.28
Lower part	2.90	4.09	**3.49**	0.32	3.06	3.48	3.96
Small pit	2.84	3.29	**3.01**	0.11	2.88	3.02	3.16

Initial 90 drill-holes + rich fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	3.02	3.65	**3.30**	0.14	3.12	3.29	3.50
Lower part	3.38	5.17	**4.01**	0.38	3.57	3.97	4.48
Small pit	2.94	3.37	**3.12**	0.11	2.98	3.11	3.25

Initial 90 drill-holes + average fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.92	3.44	**3.15**	0.13	3.00	3.17	3.36
Lower part	3.05	4.62	**3.60**	0.33	3.27	3.54	4.02
Small pit	2.82	3.31	**3.03**	0.11	2.87	3.03	3.16

Initial 90 drill-holes + poor fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.85	3.43	**3.06**	0.14	2.89	3.01	3.23
Lower part	2.82	3.94	**3.24**	0.23	2.94	3.22	3.54
Small pit	2.81	3.27	**3.01**	0.12	2.85	2.99	3.20

Table 2.9: *Simulations results: recovered average grade with a cutoff of 1g/t for the whole large pit, lower part of large pit and small pit. The number of predicted blocks is 55259 for the large pit and is 41821 for the small pit. Given the additional information the number of predicted blocks is 55605 for the large pit and is 41834 for the small pit. The size of the blocks is 10m $\times$ 25m $\times$ 10m*

2.5 Conclusions

The scope of this thesis is to value the additional information. We decided to simulate the additional data to work with different possible scenarios. Obviously, the computations become time consuming. It is perhaps more interesting when fewer data are available. Another approach could have been of assuming the percentage of reduction of the variance of the reserves due to extra information (Dias, 2002). But additional data do not always imply a reduction of variance.

We assumed that the data from both the upper and lower parts were realisations from the same random

function. It is a strong hypothesis. In addition a constant mean was considered. The parameters of the model are effectively defined by data in upper half. The data after anamorphosis are treated as being reduced gaussian, but the anamorphosis function is defined using the histogram of raw data thus of the upper data. A Bayesian approach that permits to introduce some uncertainty on the model parameters seems an interesting alternative. It allows to let some flexibility in the model. In particular, the uncertainty on the range of the covariance model allows to take into account different hypotheses concerning the correlation between the upper and lower part of the orebody.

After a description of the Bayesian approach to spatial prediction in Chapter 3, the application to the case study will be presented in Chapter 4. Compared to the Bayesian approach, conditional simulations will be called the *plug-in* approach. Finally, the project will be evaluated in Chapter 5.

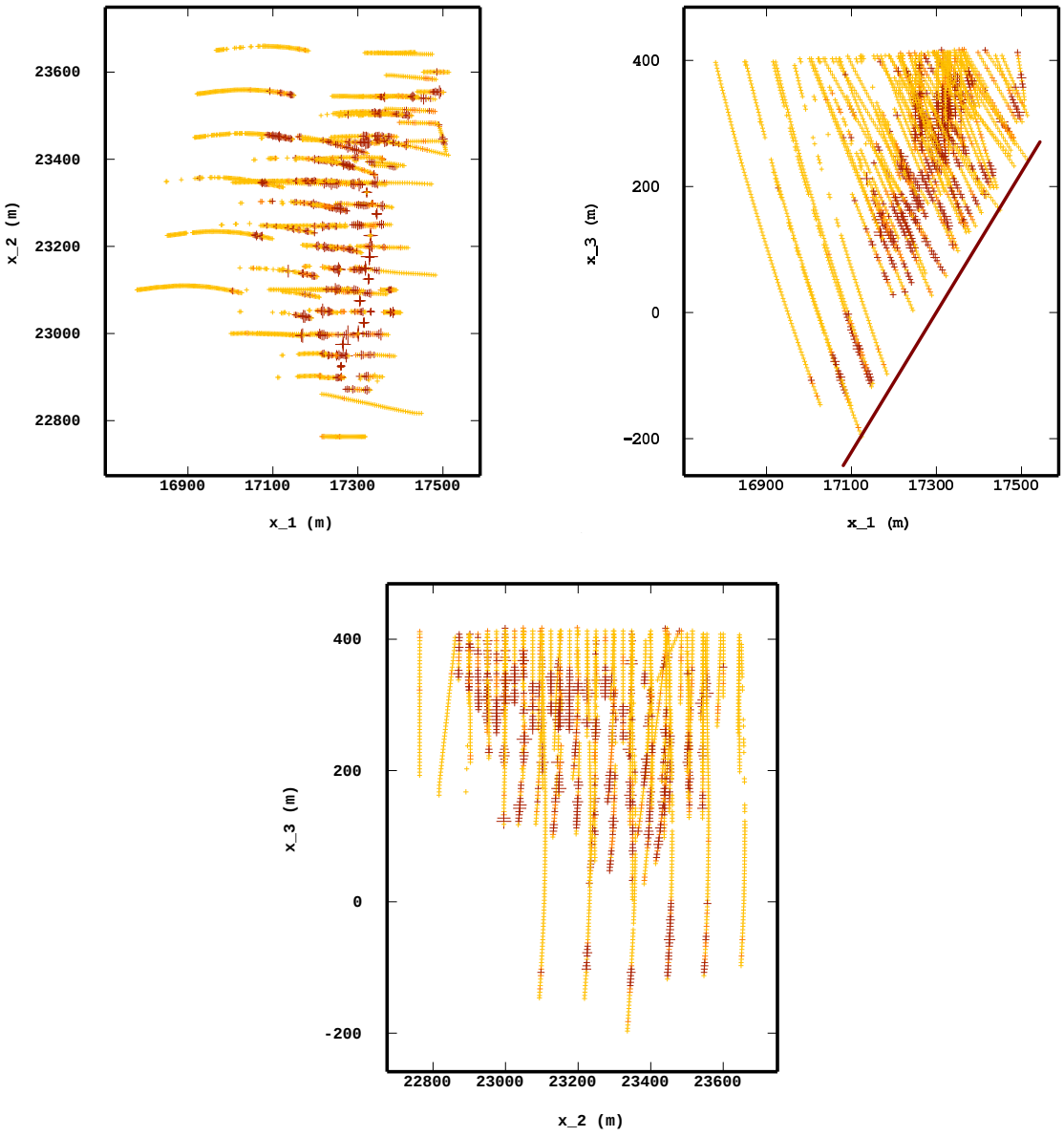


Figure 2.1: *Basemaps of the variable grade. The low values are indicated in light (yellow) colours, while the high values are highlighted in red/ brown.* x_1 *stands for the coordinate in the eastern direction,* x_2 *for the northern direction and* x_3 *for the depth. From the* $X_1 0 X_3$ *basemap we can see that the mineralisation lies along the fault (highlighted in the figure) and that there is waste above it. From the* $X_2 0 X_3$ *basemap we see that the 6 drill-holes that go into the lower part of the orebody may indicate the presence of a rich region*

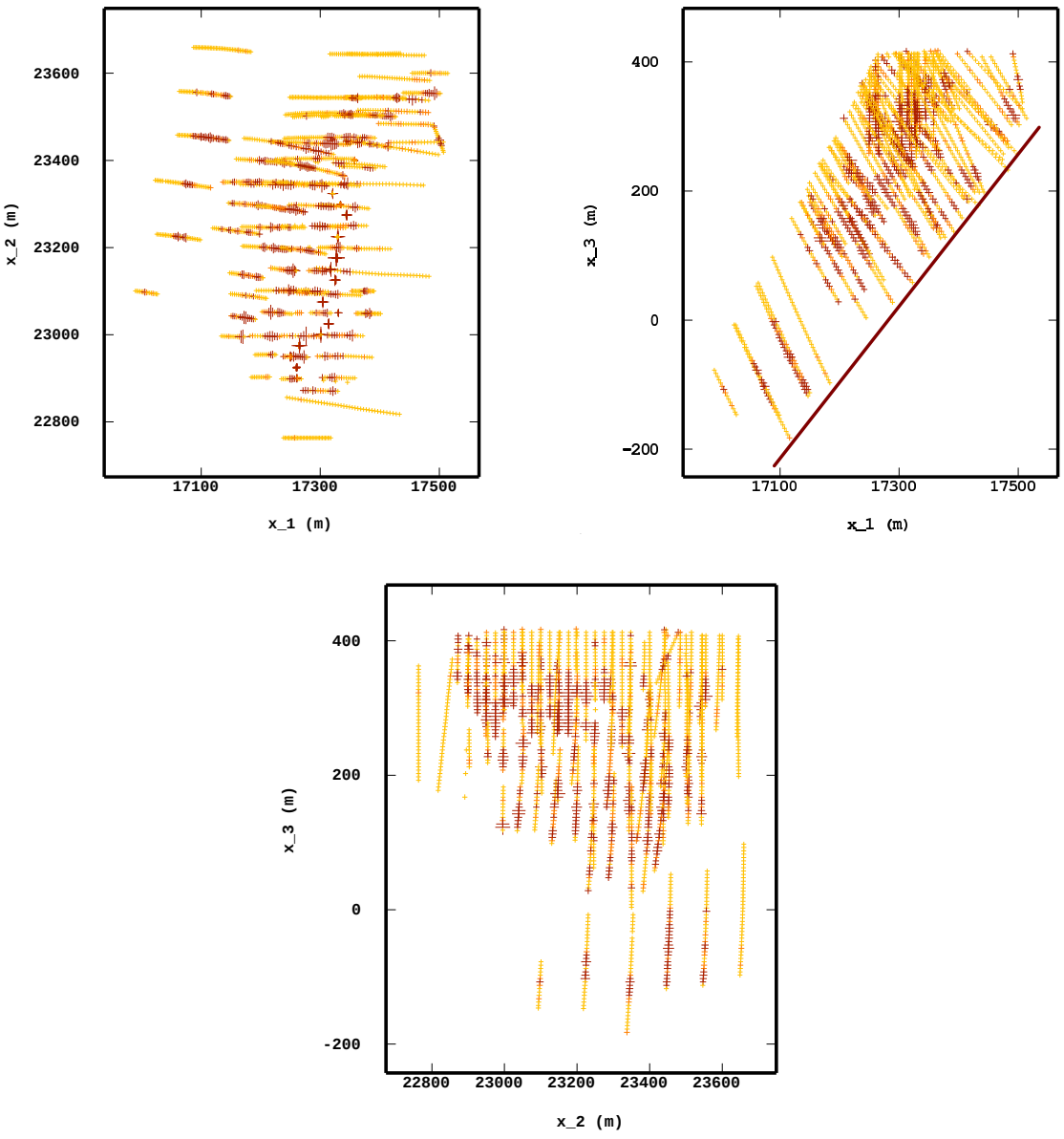


Figure 2.2: *Basemaps of the variable grade inside the mineralised zone. The size of the area under study is of 500m for x_1, 850m for x_2 and 600m for x_3*

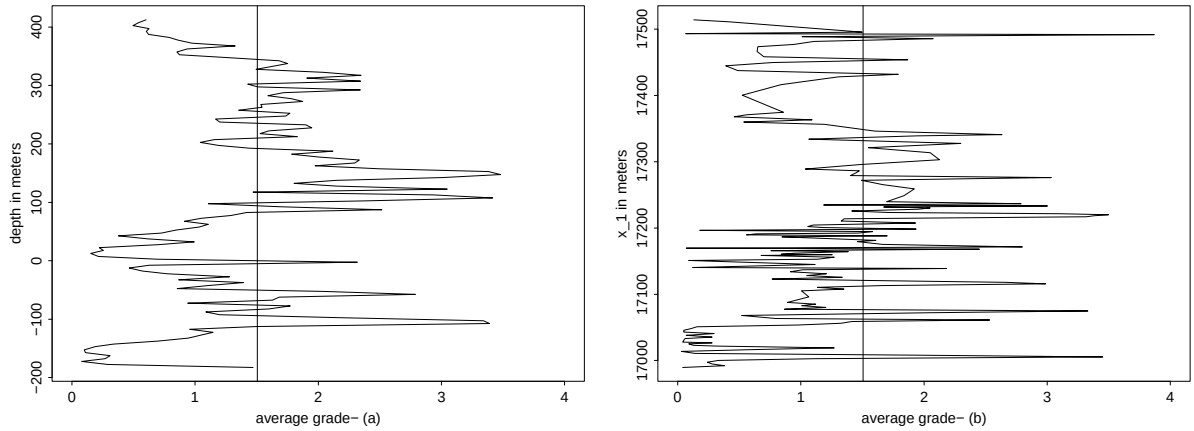


Figure 2.3: *Checking the hypothesis of stationarity. The average grade is computed for each vertical level in (a) and as a function of x_1 (i.e. the eastern direction) in (b). The vertical line in both graphics is the mean of the data. From these graphs it is plausible to treat the zone as stationary. From (a) we can note that the average grade in the vertical zone between 100m and 0m drops. As it can be seen in the previuous X_10X_3 basemap for each vertical level in this zone few observations (just 123) are available. These data come in part from the drillholes that test the lower part of the deposit. It can be noted that the average grade of the upper part is higher than the average grade of the lower part. A drift that takes this into account could be considered instead*

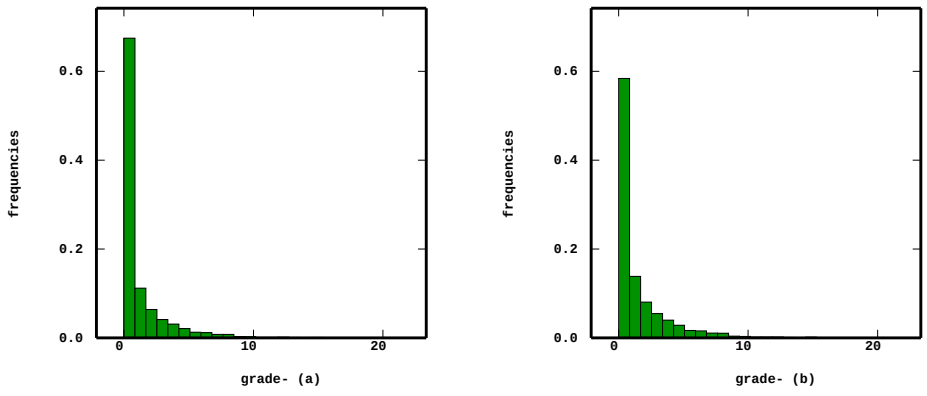


Figure 2.4: *Histograms of the variable grade (a) and of the variable grade inside the mineralised zone (b). Note that more than 50 % of the grades are smaller than 1g/t*

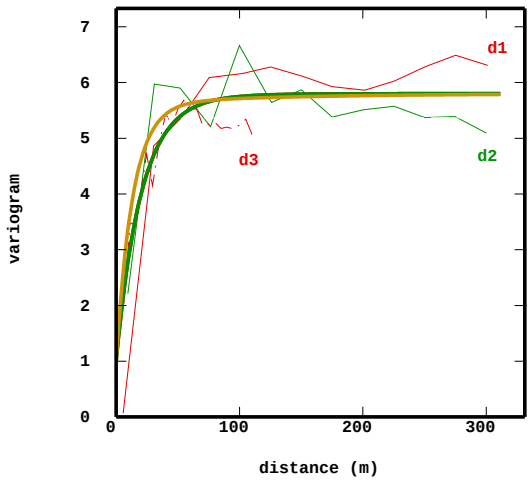


Figure 2.5: *Experimental and variogram model for the grade variable. The directions d_1 and d_2 lie in the plane parallel to the plane of mineralisation, d_3 is the direction perpendicular to this plane. A smaller range is defined for this vertical direction*

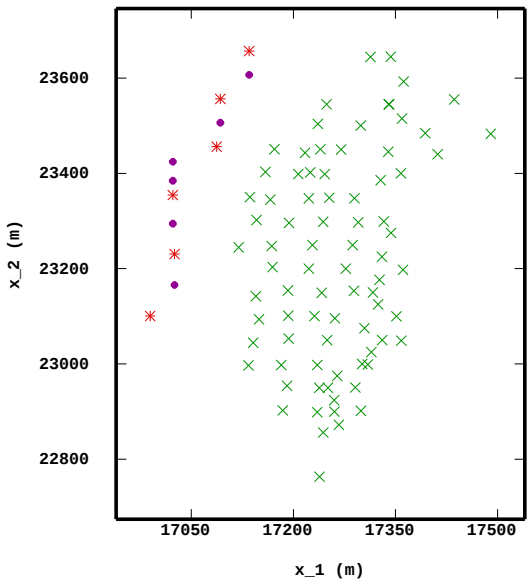


Figure 2.6: *The $X_1 0 X_2$ basemap of the variable grade. The stars represent the collars of the initial drillholes that test the lower zone, while the circles represent the collars of the new drillholes*

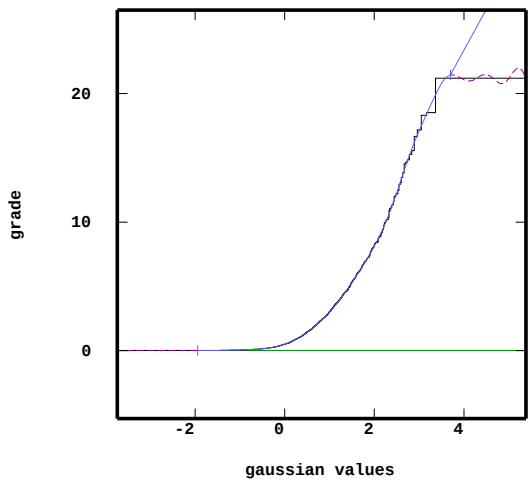


Figure 2.7: *Anamorphosis function with gaussian values along the X axis and actual values along the Y axis. Note the linear increase from the largest experimental value, 21.2, corresponding to a gaussian value of 3.56, to 30 at a maximum of 10*

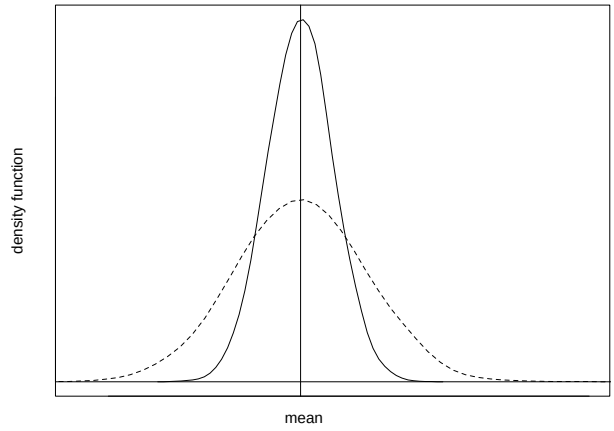


Figure 2.8: *The distribution of block grades (continuous line) is narrower than the distribution of point grades (dashed line). Both distributions have the same mean (vertical line)*

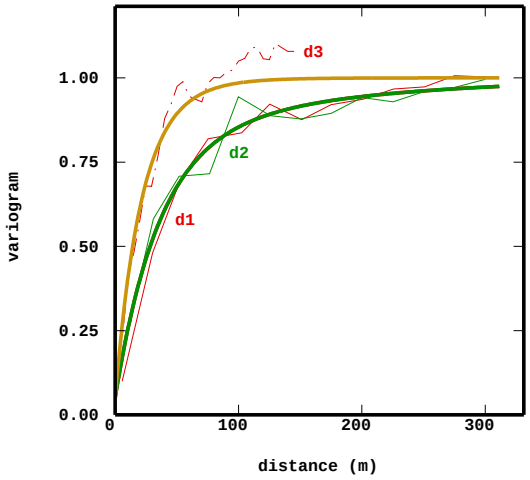


Figure 2.9: *Experimental and variogram model for the gaussian variable*

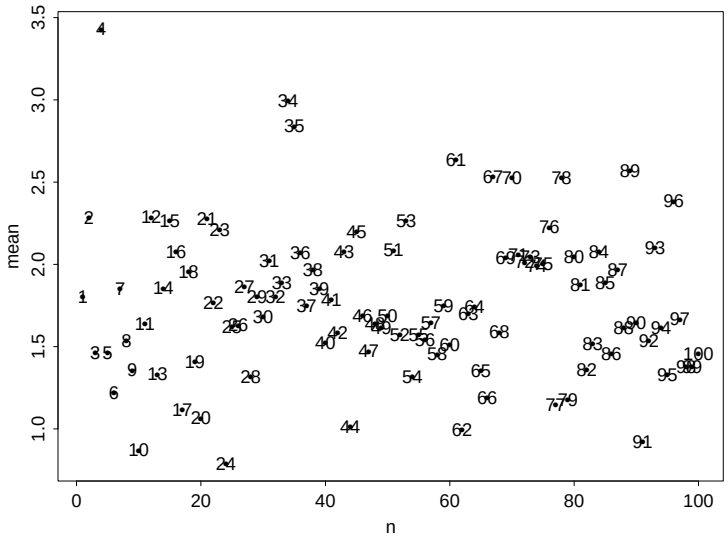


Figure 2.10: *Scatter diagram of the average grades of real + fictive drill-holes for the lower part of the orebody. Simulations 12, 46 and 66 were chosen to represent P90, P50 and P10, respectively*

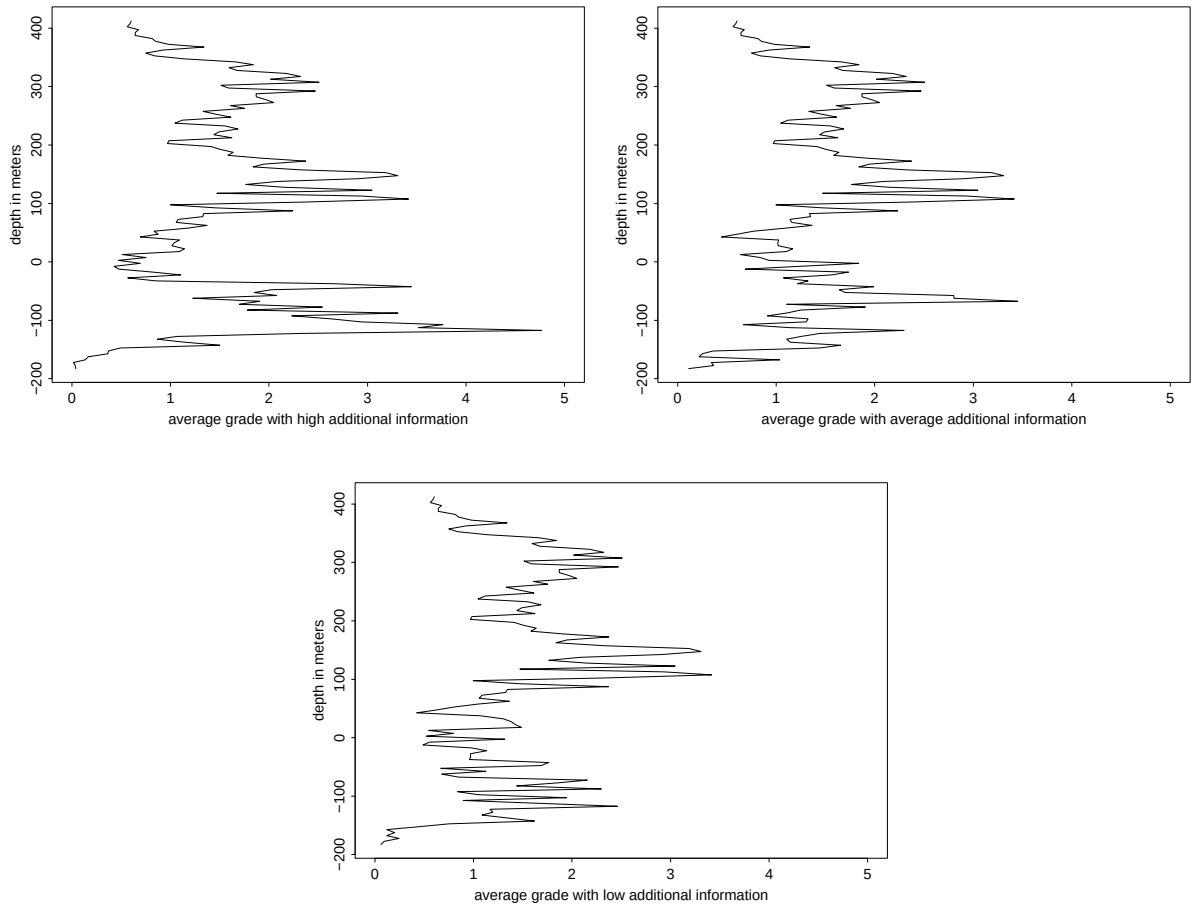


Figure 2.11: *The average grade is computed at each vertical level for the three new data sets. It can be noted that the three scenarios are quite different, as it was expected, for depths* < 0*m. Although a constant mean was assumed, with this additional data it is true that the hypothesis of a drift may be interesting (note that the data were generated with a constant mean!)*

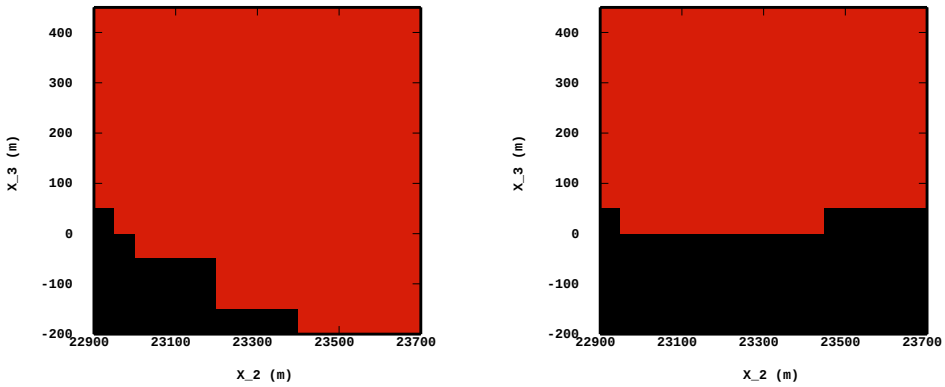


Figure 2.12: *The contours of the two pits, large and small respectively, for a given level of* x_1

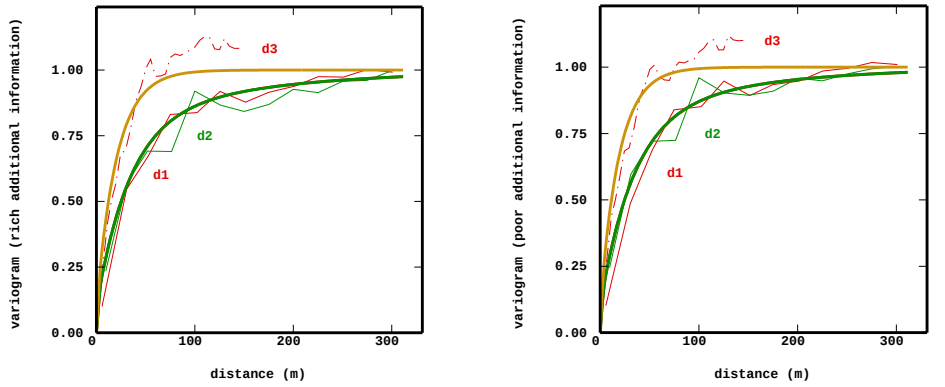


Figure 2.13: *Experimental and variogram model for the gaussian variable for the rich and the poor data sets*

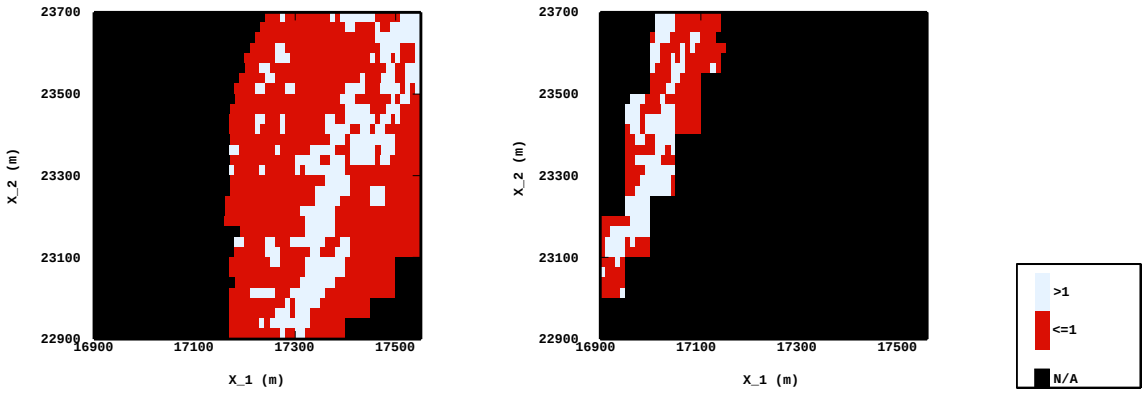


Figure 2.14: *Blocks with simulated grade above cutoff for two different depths*

Chapter 3

Bayesian approach to estimation and prediction of random fields

3.1 Introduction

Geostatistical predictions require the covariance structure as input parameters. As these are inferred from data they are subject to error. Particularly if not enough data are available this error should be taken in consideration. Bayesian geostatistical analysis (Kitanidis, 1986; Omre and Halvorsen, 1989; Handcock and Stein, 1993; Diggle *et al.*, 1998; Papritz and Moyeed, 2001) specifically recognises uncertainty on model parameters. Its effect on estimates and decision process can be evaluated. In addition it provides a statistical method for updating estimates as new information becomes available. Another advantage is that geological information or previous knowledge, that is the user normally has expectations about the general behavior of the phenomenon studied, can be included in the estimation process.

Kitanidis (1986) recognized the importance of accounting for covariance uncertainty parameters in predictive inference for gaussian random fields. He proposed a full Bayesian analysis, but did not implemented the approach, expressing the predictor and prediction variance in terms of integrals which cannot be computed analytically. Omre (1987) and Omre and Halvorsen (1989) proposed a mixed approach that they called Bayesian kriging: the mean function is defined as a random variable, while the covariance parameters are obtained using an estimator for the covariance function conditional on the mean parameters. The authors define only the first two moments of the mean parameters, while their entire distribution is needed to calculate the prediction density. They show that the choice of prior distribution for the mean leads to a continuum of methods between simple and universal kriging. Handcock and Stein (1993) proposed a full Bayesian approach for prediction in gaussian random fields that explicitly accounts for uncertainty of the covariance parameters. Their analysis was restricted to linear kriging; Diggle *et al.* (1998) extended the same approach to non-linear spatial prediction by embedding the linear kriging methodology within a more general distributional framework that is similar to the structure of a generalized linear model. De Oliveira *et al.* (1997) observe that these models seem to be well suited for image restoration inverse problems. In order to consider non gaussian random fields, they introduce a Bayesian Transform Gaussian (BTG) model which considers a parametric family of monotonic transformations and accounts for parameter uncertainty in the predictions.

Let $\{Z(x), x \in \mathcal{D}\}$, $\mathcal{D} \subseteq \mathbb{R}^3$, be the random field of interest and $\mathbf{Z} = (Z(x_1), \ldots, Z(x_n))^T$ a set of n observations from a single realisation of this random field, where $x_1, \ldots, x_n$ are known distinct locations in $\mathcal{D}$. Based on $\mathbf{Z}$ we want to predict the unobserved random vector $\mathbf{Z}(\mathbf{x_0}) = \mathbf{Z_0} = (Z(x_{01}), \ldots, Z(x_{0k}))^T$, where $\mathbf{Z_0}$ comes from the same realization as the data vector $\mathbf{Z}$ and $x_{01}, \ldots, x_{0k}$ are known distinct locations in $\mathcal{D}$. Now, let $\boldsymbol{\theta}$ be the vector of model parameters, and $f(\mathbf{Z_0}|\mathbf{Z}, \boldsymbol{\theta})$ the predictive density of $\mathbf{Z_0}$ given $\mathbf{Z}$ and $\boldsymbol{\theta}$. The parameters are viewed as random variables with prior distribution $\pi(\boldsymbol{\theta})$, and Bayes' Theorem is used to revise their probabilities. Their posterior distribution is thus

$$\pi(\boldsymbol{\theta}|\mathbf{Z}) \propto \pi(\boldsymbol{\theta})\, f(\mathbf{Z}|\boldsymbol{\theta}). \tag{3.1}$$

The prior distribution reflects beliefs about $\boldsymbol{\theta}$ prior to experimentation, the posterior distribution reflects the updated beliefs about $\boldsymbol{\theta}$ after observing the sample $\mathbf{Z}$. The posterior distribution combines the prior beliefs about $\boldsymbol{\theta}$ with the information contained in the sample $\mathbf{Z}$ to give the final beliefs about $\boldsymbol{\theta}$. But in practice how these priors and their parameters should be chosen? Or how a qualified guess can be made? It can be based for example, from a different but related data set. In petroleum reservoir description the general knowledge of the geologist is of great importance and seismic information may be available too. For other applications, such as meteorology and air pollution analysis, underlying drifts may be determined by rough deterministic models and used as qualified guesses for expected surfaces (Omre, 1987).

The Bayesian prediction is based on

$$f(\mathbf{Z_0}|\mathbf{Z}) = \int f(\mathbf{Z_0}|\mathbf{Z}, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta}|\mathbf{Z})\, d\boldsymbol{\theta} \tag{3.2}$$

which takes into account the parameter uncertainty by averaging over the parameter space the conditional distribution $f(\mathbf{Z_0}|\mathbf{Z}, \boldsymbol{\theta})$, with weights given by the posterior distribution for the model parameters, $\pi(\boldsymbol{\theta}|\mathbf{Z})$. It is obtained from both the prior and data information.

The *plug-in* prediction is based on

$$f(\mathbf{Z_0}|\mathbf{Z}, \hat{\boldsymbol{\theta}}) \tag{3.3}$$

where $\hat{\boldsymbol{\theta}}$ is an estimate of the unknown parameters $\boldsymbol{\theta}$. Thus, the plug-in approach consists of two steps: first the data are used to estimate the unknown parameters $\boldsymbol{\theta}$, and second the data and the estimates $\hat{\boldsymbol{\theta}}$ are both used to produce a predictor for $\mathbf{Z_0}$ by plugging $\hat{\boldsymbol{\theta}}$ in the place of $\boldsymbol{\theta}$. As Handcock and Wallis (1994) state, the Bayesian prediction can be interpreted as a weighted average of plug-in predictions. In comparison with likelihood based methods, the Bayesian predictive distribution takes into account the complete likelihood surface rather than focusing on the maximum likelihood estimates of the covariance parameters (Handcock and Wallis, 1994).

We let $\pi(\cdot)$ denote any parameter density function. Its arguments identify the random variables in question.

The problem of spatial prediction in random fields is now formulated as a decision problem as in Cressie (1993). The reference for decision theory and Bayesian methods is Berger (1985). We must recall, as it was said in Chapter 1, that we do not only need the prediction of $\mathbf{Z_0}$ but the entire predictive density function.

Prediction as a decision problem

Let $\mathcal{P}$ be the class of predictors, $\tilde{\mathbf{Z}}_\mathbf{0}$, for $\mathbf{Z_0}$, where $\tilde{\mathbf{Z}}_\mathbf{0}$ is a function of the data $\mathbf{Z}$. Let $L(\mathbf{Z_0}, \tilde{\mathbf{Z}}_\mathbf{0})$ be the loss function which measures the loss incurred when predicting $\mathbf{Z_0}$ with $\tilde{\mathbf{Z}}_\mathbf{0}$. An optimal predictor $\tilde{\mathbf{Z}}_\mathbf{0} \in \mathcal{P}$

is the one that minimizes the Bayes risk

$$r(\tilde{\mathbf{Z}}_\mathbf{0}) = E[L(\mathbf{Z_0}, \tilde{\mathbf{Z}}_\mathbf{0})],$$

where the expectation is taken with respect to the joint distribution of $\mathbf{Z_0}$ and $\tilde{\mathbf{Z}}_\mathbf{0}$. It is a well known result in Bayesian decision theory that regardless of the loss function used and if $\mathcal{P}$ is the class of all functions of the data, then minimizing the Bayes risk is equivalent to minimizing the Bayesian expected loss (Berger, 1985; p 159)

$$\begin{aligned} \rho(\tilde{\mathbf{Z}}_\mathbf{0}) &= E[L(\mathbf{Z_0}, \tilde{\mathbf{Z}}_\mathbf{0})|\mathbf{Z}] \\ &= \int L(\mathbf{Z_0}, \tilde{\mathbf{Z}}_\mathbf{0})\, f(\mathbf{Z_0}|\mathbf{Z})\, d\mathbf{Z_0}. \end{aligned}$$

If the squared error loss function is used, $L(\mathbf{Z_0}, \tilde{\mathbf{Z}}_\mathbf{0}) = (\mathbf{Z_0} - \tilde{\mathbf{Z}}_\mathbf{0})^2$, the optimal point predictor is the conditional expectation

$$\hat{\mathbf{Z}}_\mathbf{0} = E[\mathbf{Z_0}|\mathbf{Z}]. \tag{3.4}$$

This predictor is called the *least squares predictor*. The Bayes risk and Bayesian expected loss become respectively the unconditional and conditional mean squared prediction errors and are given by

$$r(\hat{\mathbf{Z}}_\mathbf{0}) = E[(\mathbf{Z_0} - \hat{\mathbf{Z}}_\mathbf{0})^2] = \text{Var}[\mathbf{Z_0}] - \text{Var}[\hat{\mathbf{Z}}_\mathbf{0}]$$

$$\rho(\hat{\mathbf{Z}}_\mathbf{0}) = E[(\mathbf{Z_0} - \hat{\mathbf{Z}}_\mathbf{0})^2|\mathbf{Z}] = \text{Var}[\mathbf{Z_0}|\mathbf{Z}].$$

The Bayesian expected loss is going to be used as the measure of prediction uncertainty and $f(\mathbf{Z_0}|\mathbf{Z})$ as the predictive density function. This density function is the central quantity for predictive inference. The result in (3.4) is valid not only for $\mathbf{Z_0}$ but also for other quantities of interest, such as, for example, the overall mean or the mean within any subarea.

There are two possible approaches to prediction: the Bayesian approach and the *plug-in* approach. In the following sections a detailed description of the Bayesian approach is presented, and for comparison purposes the *plug-in* approach is briefly described. We will concentrate on the results for unknown mean parameters because of their relation with conventional geostatistical methods, where the mean is filtered and the covariance parameters are estimated by some method and plugged-in for predictions. The theoretical kriging predictor which filters the mean and assumes the covariance function is known, is optimal among all unbiased linear predictors and if $Z(\cdot)$ is gaussian among all unbiased predictors. But the kriging predictor that is computed in practice is based on an estimate of the covariance function. The effect of this estimation is still not well understood. Putter and Young (2001) show that the effect of estimation is negligible asymptotically if the joint gaussian distributions of $Z(\cdot)$ at $x_1, \ldots, x_n$ under the true and estimated covariance are contiguous almost surely[1]. It is not easy to verify the conditions for contiguity. The importance of taking into consideration the uncertainty on the covariance parameters will be highlighted.

In Section 2 the gaussian random model is presented. We follow Handcock and Stein's work. Informative prior distributions for the parameters of the mean and covariance functions are considered. In Section 3 the transformed gaussian random field is treated. These approaches are applied to the test case. For this

[1]For every n, let $(\mathcal{U}_n, \mathcal{A}_n)$ be a measurable space and let $\{Q_n\}$ and $\{Q_n^{'}\}$ be two sequences of probability measures on $(\mathcal{U}_n, \mathcal{A}_n)$. The sequence $\{Q_n^{'}\}$ is contiguous with respect to $\{Q_n\}$ if, for every $A_n \in \mathcal{A}_n$, $Q_n(A_n) \to 0$ implies $Q_n^{'}(A_n) \to 0$ (Putter and Young, 2001)

only the lower part of the deposit and its 146 observations are used, while the upper data are treated as prior information.

In the next chapter the Bayesian approach will be used to obtain the recoverable reserves.

3.2 Gaussian random fields

3.2.1 The model

We assume that $Z(\cdot)$ is a gaussian random field with mean $\boldsymbol{\beta}^T f(x)$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, $\boldsymbol{\beta} \subseteq \mathbb{R}^p$, and $f(x)$ is a $p \times 1$ vector of known location-dependent covariates. Its covariance is $\alpha^{-1} K_\nu(x, x')$, $x, x' \in \mathcal{D}$, where α is the precision and $\boldsymbol{\nu}$ is a vector of structural parameters controlling the range and the smoothness of the random field, $\boldsymbol{\nu} \subseteq \mathbb{R}^q$. $K_\nu(\cdot)$ is the correlation function assumed to be positive definite $\forall \boldsymbol{\nu}$. The likelihood of the model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha, \boldsymbol{\nu})^T$ based on the observed data $\mathbf{z} = (z_1, \ldots, z_n)^T$, $z_i = Z(x_i)$, is given by

$$L(\boldsymbol{\theta}|\mathbf{z}) \propto \alpha^{\frac{n}{2}} |K_\nu|^{-\frac{1}{2}} \exp\left(-\frac{\alpha}{2}(\mathbf{z} - F\boldsymbol{\beta})^T K_\nu^{-1}(\mathbf{z} - F\boldsymbol{\beta})\right) \tag{3.5}$$

where $|K_\nu|$ is the determinant of K_ν and $F = \{f_j(x_i)\}_{n \times p}$ is assumed to have full rank.

The Bayesian predictive distribution is given in (3.2). By the stated assumptions we have that

$$(\mathbf{z}_0, \mathbf{z}|\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_{n+k}\left(\begin{pmatrix} F_0\boldsymbol{\beta} \\ F\boldsymbol{\beta} \end{pmatrix}, \alpha^{-1}\begin{pmatrix} N_\nu & L_\nu \\ L_\nu^T & K_\nu \end{pmatrix}\right)$$

where N_ν and L_ν are $k \times k$ and $k \times n$ correlation matrices. It follows that

$$(\mathbf{z}_0|\mathbf{z}, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_k(M, \alpha^{-1}\,\Sigma) \tag{3.6}$$

where

$$M = L_\nu K_\nu^{-1}\mathbf{z} + (F_0 - L_\nu K_\nu^{-1} F)\boldsymbol{\beta}$$

and

$$\Sigma = N_\nu - L_\nu K_\nu^{-1} L_\nu^T.$$

To perform the integration in (3.2) note that

$$\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}|\mathbf{z}) = \pi(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu})\, \pi(\alpha|\mathbf{z}, \boldsymbol{\nu})\, \pi(\boldsymbol{\nu}|\mathbf{z}).$$

In general, it is not possible to write down meaningful closed form expressions for the parameters posterior distributions and for the predictive distribution, and MCMC methods are required to do the computations. In particular Gibbs sampling is a natural tool to sample from a posterior of this form.

We start by defining the prior distribution for the parameters and compute their posterior distribution. In particular, as we use the predictive distribution, $f(\mathbf{z}_0|\mathbf{z}, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$, given in (3.6), we look for the following conditional distributions

- $\pi(\boldsymbol{\nu}|\mathbf{z})$;
- $\pi(\alpha|\mathbf{z}, \boldsymbol{\nu})$;

- $\pi(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu})$.

Although we use (3.6), we present the predictive distribution $f(\mathbf{z}_0|\mathbf{z}, \boldsymbol{\nu})$ obtained integrating with respect to $\boldsymbol{\beta}$ and α. Depending on the choice of the prior we have analytical forms for $\pi(\boldsymbol{\nu}|\mathbf{z})$ and $\pi(\alpha|\mathbf{z}, \boldsymbol{\nu})$. If not available the algorithm for simulations is Gibbs sampling. In this case we need the following conditional distributions

- $\pi(\boldsymbol{\nu}|\mathbf{z}, \boldsymbol{\beta}, \alpha)$;

- $\pi(\alpha|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\nu})$;

- $\pi(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu})$.

The kriging and the more general *plug- in* predictor are then described and compared to the Bayesian predictor. Two examples are given to highlight the importance of accounting for parameters uncertainty. Finally the application to the test case is presented.

The choice of the parameters prior distribution

The choice of priors is a delicate issue in Bayesian inference. If the investigator has real prior knowledge about the parameters then the issue is the expression of that knowledge in terms of distributions. The prior distribution for the parameters is unobservable. It is easier to express prior knowledge directly in terms of potentially observable Z_0 than the parameters themselves. An approach to the selection of prior distributions is the so called device of *imaginary training samples*. The distribution of Z_0 is

$$m(Z_0) = \int f(Z_0|\boldsymbol{\beta}, \alpha, \boldsymbol{\nu})\, \pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu})\, d\boldsymbol{\beta}\, d\alpha\, d\boldsymbol{\nu}.$$

Hence given that our prior knowledge about Z_0 can be expressed as $m(Z_0)$ we can then indirectly evaluate a prior for $(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$, $\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$, by solving this integral equation. This will be difficult to achieve in practice, but analysis might suggest families of priors to be explored by other methods.

The incorporation of geological knowledge into kriging studies is very situation dependent.

Conjugate priors can be computationally convenient although this alone should not justify their choice. Two extreme cases for prior choice are

- when parameters are perfectly known, the priors can be regarded as degenerate distributions on the parameters values;

- when the prior knowledge about the parameters is vague, noninformative or flat priors can be adopted.

For spatial random fields it is usual to regard the mean parameter $\boldsymbol{\beta}$ independently of the covariance parameters (Handcock, 1989). We will suppose that the **parameters** are **a priori independent** and that information is available to define the prior distribution. We start by defining the priors for $\boldsymbol{\beta}$ and α. The prior distribution for $\boldsymbol{\nu}$ is not treated here but in the applications.

For comparison purposes the following prior distributions will be considered

1. A conjugate prior. The conjugate prior family for $\boldsymbol{\beta}$ and α is the Normal- Gamma which assumes dependence between $\boldsymbol{\beta}$ and α,

$$(\boldsymbol{\beta}, \alpha|\boldsymbol{\nu}) \sim \mathcal{N}(\mathbf{b}, \alpha^{-1}S)\,\Gamma(a_1, 1/a_2).$$

   This is equivalent to the product of the distributions

$$(\boldsymbol{\beta}|\alpha, \boldsymbol{\nu}) \sim \mathcal{N}(\mathbf{b}, \alpha^{-1}S) \quad \text{and} \quad (\alpha|\boldsymbol{\nu}) \sim \Gamma(a_1, 1/a_2).$$

   That is, the prior mean and variance for $(\boldsymbol{\beta}|\alpha, \boldsymbol{\nu})$ are respectively $\mathbf{b}$ and $\alpha^{-1}S$. The prior index and scale parameter for $(\alpha|\boldsymbol{\nu})$ are respectively a_1 and $1/a_2$.

2. A noninformative prior. We chose as a noninformative prior

$$\pi(\boldsymbol{\beta}, \alpha|\boldsymbol{\nu}) \propto \frac{1}{\alpha}.$$

   As we can see in Appendix C it is Jeffreys prior for $(\boldsymbol{\beta}, \alpha)$. It is used to represent ignorance about the value of the parameters $\boldsymbol{\beta}$ and α. The prior

$$\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \propto \frac{\pi(\boldsymbol{\nu})}{\alpha}$$

   was proposed by Kitanidis (1986) and Handcock and Stein (1993), but without specific guidance as to the choice of $\pi(\boldsymbol{\nu})$. In practice, it is common to use $\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \propto 1/\alpha$, but Berger *et al.* (2000) show that this choice results in an improper posterior distribution for $(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$. This holds for other common choices of noninformative priors, such as $\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \propto 1/\alpha\,\boldsymbol{\nu}$ and $\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \propto 1$. Berger *et al.* (2000) observe that commonly used improper priors for $\boldsymbol{\nu}$ should not be chosen as they lead to improper posterior distributions. They consider objective Bayesian analysis of spatial data that utilizes noninformative or conventional prior distributions, such as Jeffreys, for unknown parameters of Gaussian random fields. In spatial models, noninformative priors are often used because of the difficulty of interpreting and hence eliciting a subjective prior for the correlation parameters. The authors provide noninformative prior distributions that result in proper posterior distributions. They define a suitable noninformative prior, the reference prior, that can be used in default Bayesian analyses. They take $\pi^R(\boldsymbol{\beta}|\alpha, \boldsymbol{\nu}) = 1$ and compute $\pi^R(\alpha, \boldsymbol{\nu})$ using Jeffreys rule for the marginal model defined via the integrated likelihood[2].

3. An informative prior which assumes independence of $\boldsymbol{\beta}$ and α. To simplify the calculations, we define a gaussian prior for $\boldsymbol{\beta}$ and a gamma prior for α. We will refer to this prior as the independent prior in contrast to the conjugate prior.

We will limit our attention to isotropic covariance functions. We will begin with only $\boldsymbol{\beta}$ unknown as it permits to see the relationship between the Bayesian approach and geostatistical methods. As Omre and Halvorsen (1989) showed if the prior for $\boldsymbol{\beta}$ is taken as noninformative, the Bayesian predictor coincides with the universal kriging (or ordinary kriging) predictor. If on the contrary a degenerate distribution is taken for $\boldsymbol{\beta}$ too, the Bayesian predictor coincides with the simple kriging predictor.

[2]The integrated likelihood is given by $L^1(\alpha, \boldsymbol{\nu}|\mathbf{z}) = \int L(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}|\mathbf{z})\,\pi^R(\boldsymbol{\beta}|\alpha, \boldsymbol{\nu})\,d\boldsymbol{\beta}$, that is the product of the likelihood and the prior $(\boldsymbol{\beta}|\alpha, \boldsymbol{\nu})$ over $\boldsymbol{\beta}$

3.2.2 Parameters posterior distribution

For the three priors we defined in the previous section we now compute the conditional distributions for the model parameters, that is $\pi(\boldsymbol{\beta}|\mathbf{z},\alpha,\boldsymbol{\nu})$, $\pi(\alpha|\mathbf{z},\boldsymbol{\nu})$ and $\pi(\boldsymbol{\nu}|\mathbf{z})$. These distributions can be computed analytically for the conjugate and noninformative priors, while for the independent prior we can only compute the full conditional distributions, that is $\pi(\boldsymbol{\beta}|\mathbf{z},\alpha,\boldsymbol{\nu})$, $\pi(\alpha|\mathbf{z},\boldsymbol{\beta},\boldsymbol{\nu})$ and $\pi(\boldsymbol{\nu}|\mathbf{z},\boldsymbol{\beta},\alpha)$.

We recall that

$$\begin{aligned}\pi(\boldsymbol{\beta},\alpha,\boldsymbol{\nu}|\mathbf{z}) &\propto \pi(\boldsymbol{\beta},\alpha,\boldsymbol{\nu})\, f(\mathbf{z}|\boldsymbol{\beta},\alpha,\boldsymbol{\nu}) \\ &= \pi(\boldsymbol{\beta},\alpha,\boldsymbol{\nu})\, \alpha^{\frac{n}{2}}\, |K_\nu|^{-\frac{1}{2}}\, \exp\left(-\frac{\alpha}{2}(\mathbf{z}-F\boldsymbol{\beta})^T K_\nu^{-1}(\mathbf{z}-F\boldsymbol{\beta})\right).\end{aligned}$$

Let

$$R(\boldsymbol{\beta},\boldsymbol{\nu}) = (\mathbf{z}-F\boldsymbol{\beta})^T K_\nu^{-1}(\mathbf{z}-F\boldsymbol{\beta}) = ||\mathbf{z}-F\boldsymbol{\beta}||^2_{K_\nu^{-1}}$$

and

$$\hat{\boldsymbol{\beta}} = \min_\beta R(\boldsymbol{\beta},\boldsymbol{\nu}),$$

that is

$$\hat{\boldsymbol{\beta}} = (F^T K_\nu^{-1} F)^{-1} F^T K_\nu^{-1}\mathbf{z}.$$

It is then possible to write

$$R(\boldsymbol{\beta},\boldsymbol{\nu}) = R(\hat{\boldsymbol{\beta}},\boldsymbol{\nu}) + ||\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}||^2_{F^T K_\nu^{-1} F}. \tag{3.7}$$

1. Conjugate prior. From Bayes' theorem we have

$$\begin{aligned}\pi(\boldsymbol{\beta}|\mathbf{z},\alpha,\boldsymbol{\nu}) &\propto \pi(\boldsymbol{\beta}|\alpha,\boldsymbol{\nu})\, f(\mathbf{z}|\boldsymbol{\beta},\alpha,\boldsymbol{\nu}) \\ &\propto \text{const}\, \exp\left(-\frac{\alpha}{2}\left(R(\boldsymbol{\beta},\boldsymbol{\nu}) + ||\boldsymbol{\beta}-\mathbf{b}||^2_{S^{-1}}\right)\right).\end{aligned}$$

It is useful to observe that

$$R(\boldsymbol{\beta},\boldsymbol{\nu}) + ||\boldsymbol{\beta}-\mathbf{b}||^2_{S^{-1}} = S(\boldsymbol{\nu}) + ||\boldsymbol{\beta}-\mathbf{b}_c||^2_{S_c^{-1}} \tag{3.8}$$

where

$$\mathbf{b}_c = S_c\,(F^T K_\nu^{-1}\mathbf{z} + S^{-1}\mathbf{b}), \quad S_c = (S^{-1} + F^T K_\nu^{-1} F)^{-1} \tag{3.9}$$

and

$$S(\boldsymbol{\nu}) = \mathbf{b}^T S^{-1}\mathbf{b} + \mathbf{z}^T K_\nu^{-1}\mathbf{z} - \mathbf{b}_c^T\, S_c^{-1}\, \mathbf{b}_c.$$

Recognising the factors involving $\boldsymbol{\beta}$ in $\pi(\boldsymbol{\beta}|\mathbf{z},\alpha,\boldsymbol{\nu})$ it is evident that

$$(\boldsymbol{\beta}|\mathbf{z},\alpha,\boldsymbol{\nu}) \sim \mathcal{N}_p(\mathbf{b}_c, \alpha^{-1} S_c).$$

Now to obtain the conditional distribution of the precision α recall that

$$\begin{aligned}\pi(\boldsymbol{\beta},\alpha|\mathbf{z},\boldsymbol{\nu}) &\propto \pi(\boldsymbol{\beta},\alpha|\boldsymbol{\nu})\, f(\mathbf{z}|\boldsymbol{\beta},\alpha,\boldsymbol{\nu}) \\ &\propto \alpha^{\frac{p}{2}}\, \alpha^{a_1+\frac{n}{2}}\, \exp\left(-\frac{\alpha}{2}(R(\boldsymbol{\beta},\boldsymbol{\nu}) + ||\boldsymbol{\beta}-\mathbf{b}||^2_{S^{-1}})\right)\, \exp(-a_2\,\alpha).\end{aligned}$$

Using (3.8) and integrating over $\boldsymbol{\beta}$ we have $(\alpha|\mathbf{z},\boldsymbol{\nu}) \sim \Gamma(a_{11}, 1/a_{22})$. The parameters of this conditional gamma distribution are

$$a_{11} = a_1 + \frac{n}{2} \quad \text{and} \quad a_{22} = a_2 + \frac{S(\boldsymbol{\nu})}{2}. \tag{3.10}$$

The choice of a_1 is of little importance when the sample size n is large. The posterior distribution for $\boldsymbol{\nu}$, $\pi(\boldsymbol{\nu}|\mathbf{z})$, can be obtained using the relation

$$\pi(\boldsymbol{\nu}|\mathbf{z}) = \frac{\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}|\mathbf{z})}{\pi(\boldsymbol{\beta}, \alpha|\mathbf{z}, \boldsymbol{\nu})}$$

which follows directly from Bayes' theorem. The posterior distribution for $\boldsymbol{\nu}$ is of the form

$$\pi(\boldsymbol{\nu}|\mathbf{z}) \propto \pi(\boldsymbol{\nu})\, |K_\nu|^{-\frac{1}{2}}\, |S_c|^{\frac{1}{2}}\, \left(a_2 + \frac{S(\boldsymbol{\nu})}{2}\right)^{-(a_1+\frac{n}{2})}. \tag{3.11}$$

2. Noninformative prior. Here we have

$$\pi(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu}) \quad \propto \quad \text{const} \exp\left(-\frac{\alpha}{2} R(\boldsymbol{\beta}, \boldsymbol{\nu})\right)$$

and using (3.7) it is evident that the conditional distribution of $\boldsymbol{\beta}$ is gaussian. Recognising the factors involving $\boldsymbol{\beta}$ we obtain the conditional mean and covariance of the mean parameter $\boldsymbol{\beta}$, that is

$$\mathbf{b}_c = \hat{\boldsymbol{\beta}} = S_c\, (F^T K_\nu^{-1} \mathbf{z}) \quad \text{and} \quad S_c = (F^T K_\nu^{-1} F)^{-1} \tag{3.12}$$

which coincides with the maximum likelihood estimator. The same result can be obtained with the conjugate prior and $S^{-1} = 0$: note that here $S(\boldsymbol{\nu}) = ||\mathbf{z} - F\mathbf{b_c}||^2_{K_\nu^{-1}} = ||\mathbf{z} - F\hat{\boldsymbol{\beta}}||^2_{K_\nu^{-1}} = R(\hat{\boldsymbol{\beta}}, \boldsymbol{\nu})$.

In this case

$$\pi(\boldsymbol{\beta}, \alpha|\mathbf{z}, \boldsymbol{\nu}) \quad \propto \quad \alpha^{-1}\, \alpha^{\frac{n-p}{2}}\, \alpha^{\frac{p}{2}}\, \exp\left(-\frac{\alpha}{2} R(\boldsymbol{\beta}, \boldsymbol{\nu})\right).$$

Thus integrating over $\boldsymbol{\beta}$ we have the conditional distribution of the precision α, which is gamma with parameters

$$a_{11} = \frac{n-p}{2} \quad \text{and} \quad a_{22} = \frac{R(\hat{\boldsymbol{\beta}}, \boldsymbol{\nu})}{2}. \tag{3.13}$$

The posterior distribution for $\boldsymbol{\nu}$ is

$$\pi(\boldsymbol{\nu}|\mathbf{z}) \propto \pi(\boldsymbol{\nu})\, |K_\nu|^{-\frac{1}{2}}\, |S_c|^{\frac{1}{2}}\, R(\hat{\boldsymbol{\beta}}, \boldsymbol{\nu})^{-\frac{n-p}{2}}. \tag{3.14}$$

It is interesting to consider the influence on the posterior distribution of $\boldsymbol{\nu}$ of the prior distributions for $\boldsymbol{\beta}$ and α. The terms in (3.11) that involve the prior parameters for $\boldsymbol{\beta}$ and α are

$$|S^{-1} + F^T K_\nu^{-1} F|^{-\frac{1}{2}}\, \left(a_2 + \frac{S(\boldsymbol{\nu})}{2}\right)^{-(a_1+\frac{n}{2})}.$$

As the prior variance for α decreases, that is either a_1 gets smaller or a_2 gets larger, this quantity tends to zero. Thus going to a noninformative prior for the precision, the distribution of ν becomes more dispersed. On the contrary, the influence of the prior for the mean coefficients $\boldsymbol{\beta}$ is little: going to a noninformative prior for $\boldsymbol{\beta}$, that is S^{-1} tends to zero, the distribution of ν becomes slightly more dispersed.

Remark. We have seen that for the conjugate and noninformative priors we have that

$$(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_p(\mathbf{b}_c, \alpha^{-1} S_c) \quad \text{and} \quad (\alpha|\mathbf{z}, \boldsymbol{\nu}) \sim \Gamma(a_{11}, 1/a_{22}),$$

then

$$(\boldsymbol{\beta}, \alpha|\mathbf{z}, \boldsymbol{\nu}) \sim \mathcal{N}_p(\mathbf{b}_c, \alpha^{-1} S_c)\, \Gamma(a_{11}, 1/a_{22}).$$

Note that

$$\pi(\boldsymbol{\beta}|\mathbf{z}, \boldsymbol{\nu}) \propto |S_c|^{-\frac{1}{2}} \int \alpha^{\frac{p}{2}} \exp\left(-\frac{\alpha}{2} ||\mathbf{b}_c - \mathbf{b}||^2_{S_c^{-1}}\right) \exp(-\alpha\, a_{22})\, \alpha^{a_{11}-1}\, a_{22}^{a_{11}}\, d\alpha$$

and recognising the terms involving α we have that

$$\pi(\boldsymbol{\beta}|\mathbf{z}, \boldsymbol{\nu}) \propto |S_c|^{-\frac{1}{2}} \left[2a_{22} + ||\mathbf{b}_c - \mathbf{b}||^2_{S_c^{-1}}\right]^{-(a_{11}+\frac{p}{2})}.$$

Thus

$$(\boldsymbol{\beta}|\mathbf{z}, \boldsymbol{\nu}) \sim t_\eta(\mathbf{b}_c, s_1^2 S_c). \tag{3.15}$$

This is a p- variate t density with $\eta = 2a_{11}$ degrees of freedom, location parameter $\mathbf{b}_c$ and scale parameter $s_1^2 S_c$, where $s_1^2 = 2a_{22}/\eta$. The parameters $(\mathbf{b}_c, S_c)$ and (a_{11}, a_{22}) depend on the prior distribution.

3. Independent prior. Now if $\boldsymbol{\beta}$ and α are *a priori* independent we cannot compute analytically the previous conditional distributions. To sample from the parameters posterior distribution Gibbs sampling can be used. Its implementation requires sampling from the full conditional ditributions $(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu})$, $(\alpha|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\nu})$ and $(\boldsymbol{\nu}|\mathbf{z}, \boldsymbol{\beta}, \alpha)$.

From Bayes' Theorem we have

$$\begin{aligned} \pi(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu}) &\propto \pi(\boldsymbol{\beta}|\alpha, \boldsymbol{\nu})\, f(\mathbf{z}|\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \\ &\propto \text{const} \exp\left(-\frac{\alpha}{2}\left(R(\boldsymbol{\beta}, \boldsymbol{\nu}) + ||\boldsymbol{\beta} - \mathbf{b}||^2_{(\alpha S)^{-1}}\right)\right). \end{aligned}$$

It is now evident that

$$(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_p(\mathbf{b}_c, \alpha^{-1} S_c)$$

where

$$\mathbf{b}_c = S_c\,(F^T K_\nu^{-1} \mathbf{z} + \alpha^{-1} S^{-1} \mathbf{b}) \quad \text{and} \quad S_c = (\alpha^{-1} S^{-1} + F^T K_\nu^{-1} F)^{-1}. \tag{3.16}$$

For the conditional distribution of α, we observe, from Bayes' theorem,

$$\begin{aligned} \pi(\alpha|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\nu}) &\propto \pi(\alpha|\boldsymbol{\beta}, \boldsymbol{\nu})\, f(\mathbf{z}|\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \\ &\propto \text{const}\, \alpha^{a_1 - 1} \exp(-\alpha a_2)\, \alpha^{\frac{n}{2}} \exp\left(-\frac{\alpha}{2} R(\boldsymbol{\beta}, \boldsymbol{\nu})\right). \end{aligned}$$

Thus

$$(\alpha|\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\nu}) \sim \Gamma(a_{11}, 1/a_{22})$$

where

$$a_{11} = a_1 + \frac{n}{2}, \quad a_{22} = a_2 + \frac{R(\boldsymbol{\beta}, \boldsymbol{\nu})}{2}.$$

Analogously we have that

$$\pi(\boldsymbol{\nu}|\mathbf{z}, \boldsymbol{\beta}, \alpha) \propto \pi(\boldsymbol{\nu})\, |K_\nu|^{-\frac{1}{2}} \exp\left(-\frac{\alpha}{2} R(\boldsymbol{\beta}, \boldsymbol{\nu})\right).$$

The choice of the correlation function

Handcock and Stein (1993) and Handcock and Wallis (1994) considered the **Matérn class** of correlation functions. It is a general class of isotropic correlation functions. It is used for the wide range of behaviors covered at the origin and the interpretability of the parameters (for more details see Stein (1999)). These correlation functions have the general form

$$K_\nu(h) = \frac{1}{2^{\nu_2-1}\Gamma(\nu_2)} \left(\frac{h}{\nu_1}\right)^{\nu_2} \mathcal{K}_{\nu_2}\left(\frac{h}{\nu_1}\right)$$

where h is the Euclidean distance, Γ is the gamma function[3], $\nu_1 > 0$ is a scale parameter controlling the range of correlation that is how fast the correlation decays with distance, $\nu_2 > 0$ is the parameter controlling the smoothness or roughness of the field that is the geometrical properties of the random field (such as continuity and differentiability), and $\mathcal{K}_{\nu_2}$ is the modified Bessel function of the second kind of order ν_2 (Abramowitz and Stegun, 1965). The case $\nu_2 = 1/2$ corresponds to the exponential model. Integer values of ν_2 indicate the number of times the process is differentiable.

De Oliveira *et al.* (1997) used the **general exponential** correlation function given by

$$K_\nu(h) = \exp\left(-\nu h^{\nu_2}\right)$$

which contains the exponential and the squared exponential correlation functions as two of its members. Let $\nu_1 = \exp(-\nu) \in (0, 1)$. This parameter controls the range of correlation. For any fixed ν_2, the correlation between observations decays with distance faster for small values of ν_1 compared to large values. The parameter $\nu_2 \in (0, 2]$ controls the roughness of the random field. This family is easy to compute and is parametrized by physically interpretable quantities. However, as Stein (in Diggle *et al.*, 1998) notes, the processes are not mean square differenciable unless $\nu_2 = 2$.

The Matérn class is thus the best choice.

In the applications we will choose the covariance function by graphically fitting the experimental variogram. We will not investigate how the choice of the covariance influences the predicting performance of the model.

Example 1

Simulations are carried out to obtain a 1 dimensional data set, **z**, of 30 observations from a standard gaussian random function Z with an exponential covariance with scale parameter 30m. The data are simulated on a grid of 290m with a mesh of 10m. The mean of the data is 0.19 and the variance is of 0.81. The experimental variogram is presented in Figure 3.1. This first example is to show the dependence of the posterior distribution of the model parameters on their prior distribution. In particular the difference in the results obtained with a noninformative prior and an informative prior is considered. For this we defined a small data set. Given the 30 observations we can assume that Z has a constant mean and an exponential covariance. The model parameters are $\boldsymbol{\theta} = (\beta, \alpha, \nu)$, where ν is the scale parameter[4]. We look for the posterior distribution of $\boldsymbol{\theta}$. The non informative prior distribution is taken for β and α. For ν a discrete uniform prior is defined between 10 and 150. The resulting posterior distributions are presented in Figure 3.2 (continuous

[3]The gamma function is defined as: $\Gamma(\nu_2) = \int_0^\infty x^{\nu_2-1}e^{-x}dx$

[4]The practical range (when the correlation is only of 5%) of the exponential model is about three times the scale parameter

curve). The mean values of the posterior distributions are, respectively, 0.23 for β, 0.97 for α and 46.01 for ν. Now suppose we have some prior information on β and α and that this translates into taking a conjugate prior for these parameters. Thus a gamma distribution with mean 1 and variance 1 is taken for α and a gaussian distribution with mean 0 and variance $1/\alpha$ is taken for $(\beta|\alpha)$. The posterior distributions obtained with this prior are shown in Figure 3.2 (dashed curves). The mean values of the posterior distributions are, respectively, 0.18 for β, 1.05 for α and 38.92 for ν. We note that the posterior mean of β gives a weight to the prior information, in this case to the value 0 for β. In addition using the prior information the estimate for the precision α increases. As expected the three posterior variances are lower when prior information is used than in the noninformative case. The posterior distributions with an informative gaussian prior for ν, of mean 30 and variance 10, are also presented in the figure (dotted curves). Thus when few data are available the influence on the results of prior information is important.

Figure 3.3 presents the posterior distribution of β with fixed covariance parameters, with ν fixed but α unknown and with both covariance parameters unknown. In (a) the noninformative prior was used, while in (b) the conjugate prior for β and α and the gaussian prior for ν were used. From the figure it can be noted that it is the scale parameter ν that mostly influences the posterior distribution of β. As expected, for the conjugate prior the differences berween the posteriors are smaller.

Figure 3.4 shows the posterior distribution for ν obtained with the two different priors. In particular in (a) different prior means and variances (conjugate prior) for β are considered. It can be seen that they do not influence the posterior distribution of ν. In (b) different prior means and variances for α are considered. It is evident that it is α and its prior parameters that mostly influence the posterior distribution of ν.

Example 2

Simulations are carried out to obtain a 1 dimensional data set, **z**, of 300 observations from a standard gaussian random function Z with a nugget effect of 0.4 and exponential covariance with scale parameter 30m. The model used to simulate the data is presented in Figure 3.5(a). The data are simulated on a grid of 299m with a mesh of 1m. The mean of the data is 0.15 and the variance is of 1.07. The experimental variogram is also presented in Figure 3.5(a). This example is to show whether it is useful or not to introduce a Bayesian approach when relatively many data are available and no prior information is given. In particular, the effect of accounting for the uncertainty on the nugget effect is investigated.

The likelihood function of $\boldsymbol{\theta} = (\beta, \alpha, \tau, \nu)$ is now given by

$$L(\boldsymbol{\theta}|\mathbf{z}) \propto |\tau^2 I + \alpha^{-1} K_\nu|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{z} - F\beta)^T (\tau^2 I + \alpha^{-1} K_\nu)^{-1} (\mathbf{z} - F\beta)\right)$$

where τ^2 is the nugget effect and I the identity matrix. For the mean parameter β the non informative prior is considered. Instead of letting the covariance parameters vary continuously we chose 15 possible sets of covariance parameters, which are supposed *a priori* to be equally likely. Two of these are shown in Figure 3.5(b). Gibbs sampling is used and simulations from the conditional distributions are obtained, but how this will be seen in more detail later in this chapter.

Figure 3.6 presents the posterior distribution of β. The dotted curve is obtained accounting for the uncertainty on the covariance parameters. The continuous and large dashed curves are obtained for two different sets of covariance parameters: true model with a nugget effect of 0.4 and a model with a nugget effect of

0.1, respectively. The Bayesian curve is close to that obtained with the true model. The posterior distributions of β obtained plugging- in the covariance parameters tend to overestimate the uncertainty. Usually in geostatistics one covariance model is chosen to fit the experimental variogram even if more than one model seems plausible. If the nugget effect is not defined while it should be included in the model the posterior distribution of β greatly overestimates the uncertainty. The Bayesian approach could thus be useful to model the uncertainty on the short scale behaviour. This should be checked in more detail for predictions.

This simplified approach still permits to take into account the uncertainty on the parameters.

3.2.3 Prediction of Z_0

For simplicity we consider the prediction of the single random variable $Z_0 = Z(x_0)$. If the parameters $\boldsymbol{\theta}$ were known, the optimal predictor for $Z(x_0)$ would be given by

$$\hat{Z}_\theta(x_0) = E_\theta[Z(x_0)|\mathbf{z}] = \mathbf{k}_\nu^T K_\nu^{-1}\mathbf{z} + (f(x_0)^T - \mathbf{k}_\nu^T K_\nu^{-1} F)\boldsymbol{\beta}. \tag{3.17}$$

which is conditionally and unconditionally unbiased[5], where $\mathbf{k}_\nu$ is a $n \times 1$ vector, $k_{\nu,i} = K_\nu(x_0, x_i)$. Under the gaussian model **the simple kriging** predictor coincides with $\hat{Z}_\theta(x_0)$. Thus under the gaussian model the simple kriging predictor is optimal. The prediction variance (and kriging variance) is

$$\text{Var}_\theta[Z(x_0)|\mathbf{z}] = E_\theta[(Z(x_0) - \hat{Z}_\theta(x_0))^2|\mathbf{z}] = \alpha^{-1}(1 - \mathbf{k}_\nu^T K_\nu^{-1}\mathbf{k}_\nu) \tag{3.18}$$

which does not depend on the data but just on their location. This is true only for the gaussian model. Note that the conditioning on the data reduces the prediction variance. Thus gaussianity is a sufficient condition for the linear SK to be optimal, if all the parameters are considered known. If the parameters are unknown the optimal predictor is in general non-linear . In general, the parameters $\boldsymbol{\theta}$ are not known. When using simple kriging, the mean is known and the covariance parameters are estimated by some method and plugged-in for predictions, as if they were the truth parameters. It must be noted that if we misspecify α we will misspecify the prediction variance proportionately. If we misspecify $\boldsymbol{\nu}$ both the predictor and the prediction variance will be affected.

As it was said earlier, if both the mean and covariance parameters are unknown, it is not possible to write down closed form expressions for the predictive distribution and so MCMC methods must be used. For example, Gibbs sampling can be used. If just the mean parameter is unknown and as $(\boldsymbol{\beta}|\mathbf{z}, \alpha, \boldsymbol{\nu})$ follows a normal density, the predictive distribution, $f(Z(x_0)|\mathbf{z}, \alpha, \boldsymbol{\nu})$, is a normal distribution, $\mathcal{N}(M_\nu, \alpha^{-1}\Sigma_\nu)$. The parameters depend on the prior. The mean and variance of the predictive distribution are

$$M_\nu = E[Z(x_0)|\mathbf{z}, \alpha, \boldsymbol{\nu}] = \mathbf{k}_\nu^T K_\nu^{-1}\mathbf{z} + (f(x_0)^T - \mathbf{k}_\nu^T K_\nu^{-1} F)\mathbf{b}_c \tag{3.19}$$

and

$$\alpha^{-1}\Sigma_\nu = \text{Var}[Z(x_0)|\mathbf{z}, \alpha, \boldsymbol{\nu}] = \alpha^{-1}[(1 - \mathbf{k}_\nu^T K_\nu^{-1}\mathbf{k}_\nu) + (f(x_0)^T - \mathbf{k}_\nu^T K_\nu^{-1} F)S_c(f(x_0)^T - \mathbf{k}_\nu^T K_\nu^{-1} F)^T] \tag{3.20}$$

where $\mathbf{b}_c$ and S_c are given in

[5] It is conditionally and unconditionally unbiased, that is

$$E_\theta[\hat{Z}_\theta(x_0)] = E_\theta[Z(x_0)] \quad \text{and} \quad E_\theta[\hat{Z}_\theta(x_0)|\mathbf{z}] = E_\theta[Z(x_0)|\mathbf{z}]$$

1. Equation 3.9 for the conjugate prior;

2. Equation 3.12 for the non informative prior and in

3. Equation 3.16 for the independent prior.

The increase in variance with respect to (3.18) is due to the fact of having taken into account the uncertainty on β.

Once a sample of the covariance parameters is available, we sample from $f(Z(x_0)|\mathbf{z}, \alpha, \boldsymbol{\nu})$. However if just the mean or the mean and the precision parameters are unknown, depending on the prior specified, the predictive distribution can be calculated analytically. This is the case for the conjugate and for the noninformative prior. The predictive distribution is $f(Z(x_0)|\mathbf{z}, \boldsymbol{\nu})$. It is a t distribution with η degrees of freedom, with location parameter M_ν and scale parameter $s_1^2\Sigma_\nu$, that is $t_\eta(M_\nu, s_1^2\Sigma_\nu)$. Note that now the predictive variance depends on the data and not only on their location. The parameters depend on the prior as we have seen previously. The ignorance about the precision parameter α is expressed by the difference between the variance of the t distribution, $\eta s_1^2\Sigma_\nu/(\eta-2)$, and the variance of the gaussian distribution, $\alpha^{-1}\Sigma_\nu$. As s_1^2 is the natural estimate of α^{-1} the ratio of the variances is approximately given by $\eta/(\eta-2)$. Thus if η is large the uncertainty on α will not change the predictive distribution very much. Thus the uncertainty in α manifests itself in the conversion from a gaussian to a t distribution. The uncertainty on $\boldsymbol{\nu}$ manifests itself through the weighting of each of these t distributions by the posterior for $\boldsymbol{\nu}$. Depending on the influence of the correlation parameters on the spread and location of the t distribution, the Bayesian posterior might be wider or narrower than the predictive plug-in distribution.

Example 1

Reconsider Example 1 introduced previously. Given the 30 observations we want to predict the random variable z_0 at an unobserved location x_0. We first compare the Bayesian predictive distributions obtained with the noninformative prior for β and α and the gaussian prior for ν and with the conjugate prior for β and α and the gaussian prior for ν. Figure 3.7 shows clearly that when the uncertainty about z_0 is great, prior information plays an important role, and on the contrary when the uncertainty about z_0 is small prior information has a smaller weight compared to data. The noninformative prior is then used to highlight the importance of taking into account the uncertainty on the covariance parameters. The *plug- in*, that is the results we would obtain if simple kriging or ordinary kriging were carried out, and the Bayesian predictive distributions are compared in Figure 3.8. The figure presents these different predictive distributions at two unknown locations. In (a) x_0 is 55m that is close to the 30 observations and in (b) x_0 is 400m that is far from the observed data. The influence of taking into account the uncertainty on the mean and covariance parameters is particularly important in (b). On the contrary it can be noted that taking into account the uncertainty on the mean is not relevant in (a).

Figure 3.9 presents the predictive distribution at $x_0 = 450$m. In (a) it can be noted that the Bayesian predictive distribution tends to overestimate the uncertainty with respect to the predictive distributions with the covariance parameters known. The difference is larger when the uncertainty on z_0 is larger. In (b) predictive distributions obtained with β unknown and with different values of the covariance parameters are compared. The Bayesian predictive distribution is a weighted mean of these distributions.

Example 2

This example is considered here to see what is the effect on the predictions of misspecifying the nugget effect. Looking at the experimental variogram (Figure 3.5) we can assume that there is a nugget effect. Figure 3.10 presents the predictive distributions at $x_0 = 55.5$m and $x_0 = 400$m. The Bayesian predictive distribution is compared to the predictive distribution with β unknown and fixed covariance parameters. In (a) and (c) the predictive distribution is defined using the true model. Looking at these graphs it can be said that when the uncertainty on z_0 is small as in (a) the Bayesian predictive distribution tends to overestimate the uncertainty. On the contrary when the uncertainty on z_0 is great as in (c) the Bayesian predictive distribution tends to underestimate the uncertainty. In (b) and (d) the predictive distribution is defined using a model that seems to fit well the experimental variogram and has a smaller nugget effect but still a total variance of 1. This model is one of the fifteen we defined for the covariance function. Looking at (b) it can be seen that a smaller nugget effect greatly reduces the predictive variance. In this case we would underestimate the uncertainty on z_0. This does not happen in (d) as the range of correlation is small compared to the distance between z_0 and the observed data and thus just the total variance is relevant. The Bayesian predictive distribution prevents us of these kind of errors. The Bayesian predictive distribution is a weighted mean of the *plug- in* predictive distributions.

3.2.4 The *plug-in* approach

We consider the maximum likelihood (ML) approach for estimating $\boldsymbol{\theta}$, that is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} L(\boldsymbol{\theta}|\mathbf{z}).$$

Once the model parameters are estimated, they are plugged-in in the predictive distribution to obtain the predictor and prediction variance of Z_0 given respectively in (3.19) and (3.20).

Only the mean parameter is unknown. As we can see in Equations 3.19 and 3.20 we do not need to estimate the mean parameters $\boldsymbol{\beta}$. However, for any fixed value of α and $\boldsymbol{\nu}$, the ML estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, is given in (3.12) which is the *generalized least squares* estimate of $\boldsymbol{\beta}$, even when the random field is not gaussian. In this case $\hat{Z}(x_0) = E[Z(x_0)|\mathbf{z}]$ is (3.19) which is also the *best linear unbiased* predictor of $Z(x_0)$ and is known as the **universal kriging** predictor if F and $f(x_0)$ are trend matrices with rows given by data coordinates or a function of them (**ordinary kriging** if $\boldsymbol{\beta} = \beta_0$), or as the **kriging with external trend** predictor if F and $f(x_0)$ are trend matrices with covariates measured at data and prediction locations, respectively. Its prediction variance (or kriging variance) is given in (3.20) which depends on both α and $\boldsymbol{\nu}$.

In this case the Bayesian predictor coincides with the kriging predictor.

The mean parameter and the precision parameter are unknown. For any fixed value of $\boldsymbol{\nu}$, and substituing $\hat{\boldsymbol{\beta}}$ in the place of $\boldsymbol{\beta}$ in the likelihood function (3.5) we obtain the following profile likelihood

$$L_p(\alpha|\mathbf{z}, \boldsymbol{\nu}) \propto \alpha^{\frac{n}{2}} |K_\nu|^{-\frac{1}{2}} \exp\left(-\frac{\alpha}{2} R(\hat{\boldsymbol{\beta}}, \boldsymbol{\nu})\right).$$

Thus $\hat{\alpha}$ which is given by

$$\arg\max_{\alpha} L_p(\alpha|\mathbf{z}, \boldsymbol{\nu})$$

is

$$\hat{\alpha} = \left(\frac{R(\hat{\boldsymbol{\beta}}, \nu)}{n}\right)^{-1}.$$

This could be compared to the mean of the conditional gamma obtained with a noninformative prior and with parameters given in (3.13). As $E[Z(x_0)|\mathbf{z}]$ does not depend on α it is unchanged, but the prediction variance is now given by $\hat{\alpha}^{-1}\,\Sigma$.

All the parameters are unknown. Replacing $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$ in the likelihood function gives the following profile likelihood

$$L_p(\boldsymbol{\nu}|\mathbf{z}) \propto |K_\nu|^{-\frac{1}{2}}\,\hat{\alpha}^{\frac{n}{2}}$$

and the ML estimate of $\boldsymbol{\nu}$, $\hat{\boldsymbol{\nu}}$, if it exists, is obtained by iterative numerical methods.
Thus, the predictor $E[Z(x_0)|\mathbf{z}]$ is

$$\hat{Z}(x_0) = \mathbf{k}_{\hat{\nu}}^T K_{\hat{\nu}}^{-1}\,\mathbf{z} + (f(x_0)^T - \mathbf{k}_{\hat{\nu}}^T K_{\hat{\nu}}^{-1} F)\hat{\boldsymbol{\beta}}_{\hat{\nu}} \tag{3.21}$$

which has been named the empirical or estimated best linear unbiased predictor of $Z(x_0)$. In general, it is a non linear predictor since $\hat{\boldsymbol{\nu}}$ is a non linear estimator. The variance of $\hat{Z}(x_0)$ has no closed form, and the usual estimate of it is obtained by plugging $\hat{\alpha}$ and $\hat{\boldsymbol{\nu}}$ in (3.20). This would in general result into an overestimation of the predictor's precision and in consequence the prediction intervals for $Z(x_0)$ tend to be too optimistic. The estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$ are then computed with $\hat{\boldsymbol{\nu}}$. The estimation of $\boldsymbol{\nu}$ by ML has been criticized because of the occurrence of multimodal likelihoods. As De Oliveira *et al.* (1997) state for the Bayesian approach, in principle, a multimodal likelihood is not an issue for the prediction of $Z(x_0)$.

The most common alternative for the estimation of covariance parameters is the weighted least squares method for curve fitting as proposed in Cressie (1993, p.94). The resulting fits are often visually convincing, but this raises the question of whether matching theoretical and experimental variogram is optimal in any sense. The ML methods have the advantage of being optimal under stated assumptions although they are computationally expensive for large data sets. As we have seen they also play a central role in Bayesian inference.

The plug-in posterior can be wider or narrower than the Bayesian posterior depending on the estimates used. In general, it tends to underestimate the uncertainty by a small amount. The weak point of the *plug-in* approach is that it does not account for the uncertainty about the covariance parameters on the estimation of the prediction uncertainty. This has been one of factors that motivated the development of the Bayesian approach to prediction in random fields. The Bayesian approach fully accounts for model uncertainty in the computation of predictors and their associated uncertainty measures. However, it is more computationally demanding. Chilès and Delfiner (1999, pp 175–176) observe that with the Bayesian approach two different sources of uncertainty, the spatial variability and the uncertainty on the magnitude of this variability, are mixed. Moreover, to obtain the UK (or OK) predictor no distributional assumptions are needed, while for the Bayesian approach the conditional distribution of $Z(x_0)$ given $\mathbf{z}$ and the parameters $\boldsymbol{\theta}$ is required. On the other side, linear kriging predictors are compelling only under the gaussian assumption. Papritz and Moyeed (2001) compare Bayesian and model-based kriging with *plug-in* kriging methods for two different data sets. Given their results they doubt that there is a gain by using Bayesian approaches to kriging. They found that it did not matter whether parameter uncertainty was taken into account or not. The authors suggest that this depends on the validity of the modelling assumptions, that are quite strong for the Bayesian approach.

3.2.5 Application to the test case

As we have seen the data are clearly not gaussian. Thus the gaussian model presented in this chapter is not realistic. Although the mean $F\boldsymbol{\beta}$ is positive and Z centered around this mean does not take negative values,

once the new information is available the change in the mean could be important and negative values of Z be obtained.

To work with this model we need to transform the data into gaussian values. The anamorphosed grades, introduced in Chapter 2 and presented in Table 3.1, will be used to describe the gaussian model. We use the observations of the lower part and use the upper data as prior information.

	N	Minimum	Maximum	Mean	Standard deviation
	2390	-2.05	3.56	0.00	0.99
Upper part	2244	-2.05	3.56	0.00	0.99
Lower part	146	-2.05	2.13	0.00	0.92

Table 3.1: *Statistics of the gaussian values*

The model

We chose the following drift function

$$E[Z(x)|\boldsymbol{\beta}] = \beta_0 + f_1(x)\beta_1 + f_2(x)\beta_2$$

where $f_1(x)$ is the vertical coordinate (or depth) of $Z(x)$ and $f_2(x)$ is the vertical coordinate squared. We did not consider the coordinates in the horizontal plane. It must be noted that the data are in this plane in a rather small area and the coordinates are very close. The covariance model is the same exponential as in Table 2.5. This model was chosen after fitting the experimental variograms of the upper data (Figure 3.11). Note that the 146 observations of the lower part of the orebody do not influence the experimental variograms. For simplicity, the omnidirectional variogram is considered instead (Figure 3.12). The model parameters are thus $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha, \nu)$.

As it was said earlier, the parameters are *a priori* independent. The following prior distibutions for the model parameters were defined

- a gaussian distribution for $\boldsymbol{\beta}$ with mean $\mathbf{b} = \mathbf{0}$ and covariance S equal to the identity matrix multiplied by 10;
- a gamma distribution for α of index $a_1 = 1$ and scale parameter $a_2 = 1$, that is of mean 1 and variance 1, and
- a discrete prior for ν on the interval $(10, 100)$. More weight is given to values in the interval $(20, 40)$, so that the mean is 30.

The mean parameters of the priors for α and ν were computed by eye fitting the experimental variogram (Figure 3.12). To highlight our ignorance on $\boldsymbol{\beta}$, its variance is set equal to 10.

Parameters posterior distribution. The algorithm used to sample from the parameters posterior distribution consists of the following steps

1. choose the initial values for α and ν. These starting values are chosen in the range specified by the prior.

2. Update all the components of the parameter vector $\boldsymbol{\beta}$

   - choose a new value, $\boldsymbol{\beta}'$, from the conditional probability function $\pi(\boldsymbol{\beta}|\mathbf{z}, \alpha, \nu)$.

3. Update α

   - choose a new value, α', from the conditional probability function $\pi(\alpha|\mathbf{z}, \boldsymbol{\beta}, \nu)$.

4. Update ν

   - compute the approximate discrete posterior distribution in the support chosen for ν, that is $\tilde{\pi}(\nu|\mathbf{z}, \boldsymbol{\beta}, \alpha)$;
   - choose a new value, ν', from $\tilde{\pi}(\nu|\mathbf{z}, \boldsymbol{\beta}, \alpha)$.

We observe that at each iteration K_ν^{-1} must be computed and as we have 146 observations it is quite time consuming. The advantage of taking a discrete prior for ν is thus clear[6].

Steps 2-4 were iterated until the chain was judged to have reached its equilibrium distribution. Then we sampled every kth ($= 50$) realisation of the chain. Increasing the value of k reduces the serial correlation in the resulting output sample (Diggle *et al.*, 1998). Figure 3.13 presents the samples of $(\boldsymbol{\beta}, \alpha, \nu)$ given $\mathbf{z}$ and the evolution of the Monte Carlo estimates of these parameters as a function of the number of iterations. The chains seem stationary. Figure 3.14 shows the posterior distribution for the model parameters. The posterior mean for $\boldsymbol{\beta}$ is $(-0.558, -0.011, -6.203 \times 10^{-5})^T$. The posterior mean of the precision α is 0.87 and of the range ν is 33.44. The correlation between the parameters is shown in Figure 3.15. As expected, there is a strong dependence between the two parameters of the covariance, α and ν (the range increases as the variance increases).

Figure 3.16 presents the estimated drift. The solid line represents the average grade measured at each depth from -200m to 0m. The figure contains 3 dotted lines which correspond to different estimated drifts, that is with a different number of drift parameters.

Sensitivity to the prior. We compare the results for the different priors. The prior distribution for ν is always the same gaussian defined previously. We restrict the analysis to a constant drift. Figure 3.17 shows the posterior distribution for the model parameters. The mean values of these posterior distributions are almost identical

- with the independent prior we have -0.20, 0.91 and 30.98 for β, α and ν, respectively;
- with the conjugate prior we have -0.20, 0.90 and 32.10 for β, α and ν, respectively;
- with the noninformative prior we have -0.22, 0.87 and 34.04 for β, α and ν, respectively.

However, the variance and in particular the behavior in the tails differ. As it is expected, the posterior variance is larger and the tails are heavier with a noninformative prior. Note that the precision is slightly smaller in the noninformative case. As the choice of prior distributions is extremely difficult, it is important to check the inference for the sensitivity to the priors chosen. In our case we must admit that the posterior distributions are very close and the choice of the prior has little influence on the results.

[6] If we do not discretise the posterior distribution of the correlation parameter ν, then to sample from $\pi(\nu|\mathbf{z}, \boldsymbol{\beta}, \alpha)$ a Metropolis-Hastings algorithm can be used

The importance of taking into account the uncertainty on the covariance parameters is highlighted in Figure 3.17(d). The posterior distributions for the mean for three cases are compared: with the 2 covariance parameters unknown, with α unknown and ν fixed and finally with both α and ν fixed. The third and second curves are very close, while the first one has a higher variance and is skew. The importance of taking into account the uncertainty on α seems thus moderate. Integrating out α, in the conjugate case, the distribution of β goes from a gaussian to a t distribution with the same mean but a different variance (see Equation 3.15). The ratio of the variances of $(\beta|\mathbf{z},\nu)$ and $(\beta|\mathbf{z},\alpha,\nu)$ is 1.01.

The fact of accounting for the uncertainty on the covariance parameters would be more evident if fewer data were available. For this we kept just 20 observations and computed the posterior distribution for β (Figure 3.18). The posterior distribution of $(\beta|\mathbf{z},\nu)$ is thus a t distribution with $\nu = 21$ degrees of freedom. The ratio of the variances of $(\beta|\mathbf{z},\nu)$ and $(\beta|\mathbf{z},\alpha,\nu)$ is 1.10. In addition note that in this case the prior distribution has a larger impact on the posterior distribution (Figure 3.18(a)).

Prediction of Z_0. For each $(\boldsymbol{\beta},\alpha,\nu)$ obtained previously, we sample from $f(Z(x_0)|\mathbf{z},\boldsymbol{\beta},\alpha,\nu)$. Figure 3.20 shows the predictive distributions at the 4 selected locations presented in Figure 3.19 for different levels of uncertainty. The dotted curve is obtained with the model parameters fixed at their posterior mean. The predictor corresponds to the SK predictor. The dashed curve is obtained taking into account the uncertainty on $\boldsymbol{\beta}$. This predictor corresponds to the UK predictor. SK and UK produce very similar results. Accounting for the uncertainty on $\boldsymbol{\beta}$ increases the prediction variance. The large dashed curve is obtained taking into account the uncertainty on the precision α. The prediction variance increases. Finally, the continuous curve represents the Bayesian predictive distribution. In general a higher variance is obtained when taking into account the uncertainty on the correlation parameter ν, but it is not always the case as it can be seen in Figure 3.20(c). This can be due to the fact that the Bayesian predictive distribution is a weighted mean of the predictive distributions with fixed model parameters.

Comparison with the plug-in **approach.** Figure 3.21 compares the Bayesian and the *plug-in* predictive distributions obtained fixing *by eye* the covariance parameters. The Bayesian predictive distribution has heavier tails. Thus the *plug-in* approach tends to underestimate the prediction variance.

The model parameters can be estimated by ML. The profile likelihood for ν is shown in Figure 3.22. The ML estimate of ν is 21. The corresponding ML estimates of α and $\boldsymbol{\beta}$ are 1.17 and $(-0.547, -0.013, -7.238 \times 10^{-5})^T$. Note that the estimate of the precision α is much larger than the value obtained with the Bayesian approach (0.87). This higher precision is due to the fact that the uncertainty on the parameters is not taken into account and the prediction variance, for the 4 selected points presented in Figure 3.19, is much smaller. This can be seen in Figure 3.23.

Cross validation is often used to compare different prediction methods (Papritz and Moyeed, 2001). We did not use any measure to rank the *plug-in* and Bayesian predictions as we are interested in the entire predictive distribution.

3.2.6 Conclusions

Kriging when the mean is of a known regression form can be given a Bayesian interpretation. Hence except for the differences in interpretation we end up with the same analysis as the traditional or *plug-in* approach.

The exponential model was chosen instead of using a more general family of covariance functions such as the Matérn. However, we wanted to highlight the importance of accounting for the uncertainty on the covariance parameters. Moreover, we decided that enough data were available to choose the covariance function. The calculations are time consuming and thus we preferred to allow for the uncertainty on a minimum number of parameters.

The choice of the prior for the model parameters is a crucial point in Bayesian analysis. We have seen that for the test case this choice does not have a large impact on the results.

The sampling algorithm requires at each iteration the computation of the inverse of the covariance matrix. This is time consuming and if the dimension of the data set is large (more than 100) this becomes unfeasible. One possibility is to use a reduced covariance matrix. As we will see in the next chapter, we discretised the range of the correlation parameter ν and computed the covariance matrix and its inverse once for all.

The data we are working on are clearly not gaussian. Thus the gaussian model presented in this chapter is not realistic. Although the mean $F\beta$ is supposed positive and the gaussian distribution of Z centered around this mean does not take negative values, once the new information is available the change in the mean could be important and negative values of Z be obtained. For this reason we consider in the next section the transformed gaussian model proposed by De Oliveira *et al.* (1997).

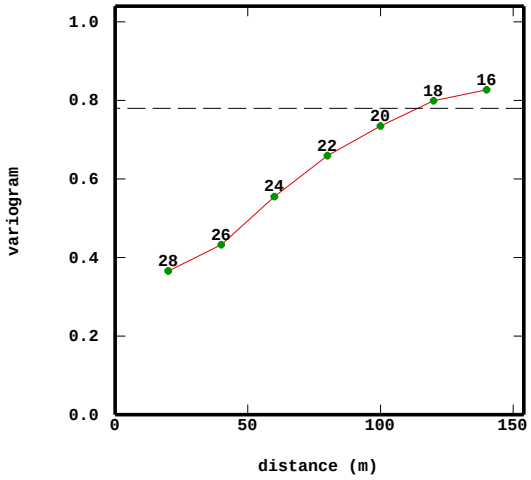


Figure 3.1: *Experimental variogram. The horizontal dashed line represents the variance of the data. The number of pairs for each distance is shown*

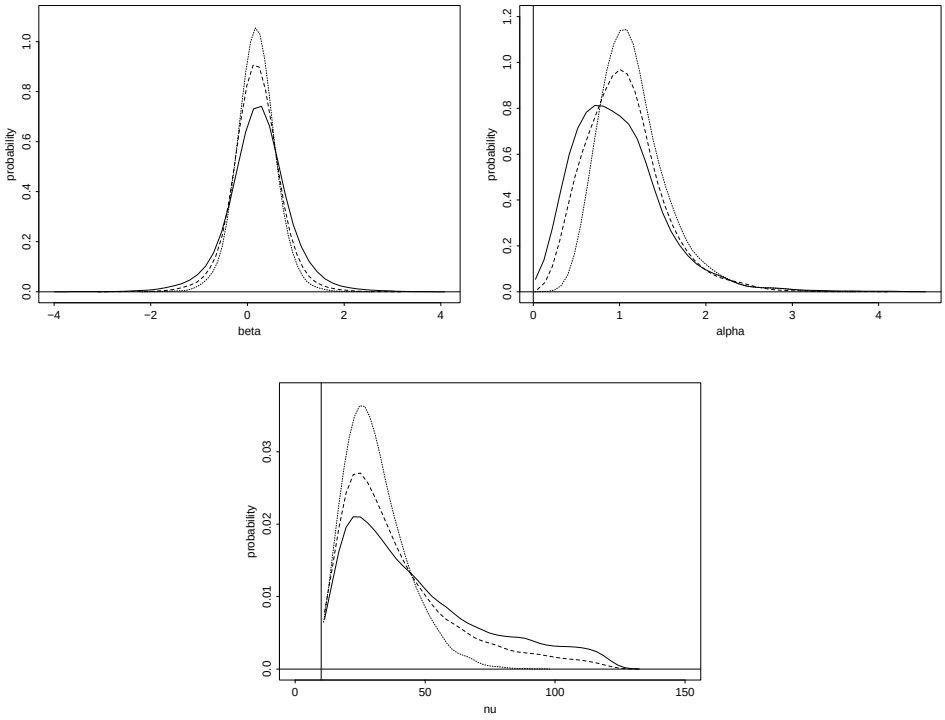


Figure 3.2: *Posterior distribution of the model parameters. The continuous curve is obtained with the noninformative prior for β and α and a uniform prior for ν. The dashed curve is obtained with the conjugate prior for β and α and a uniform prior for ν. The dotted line is obtained with the conjugate prior for β and α and with a gaussian prior for ν. From these graphs it is possible to see how the prior information changes the posteriors. In particular it is interesting to note that the change in the prior for ν influences greatly the posterior of α, while the effect on β seems more moderate (not for its variance)*

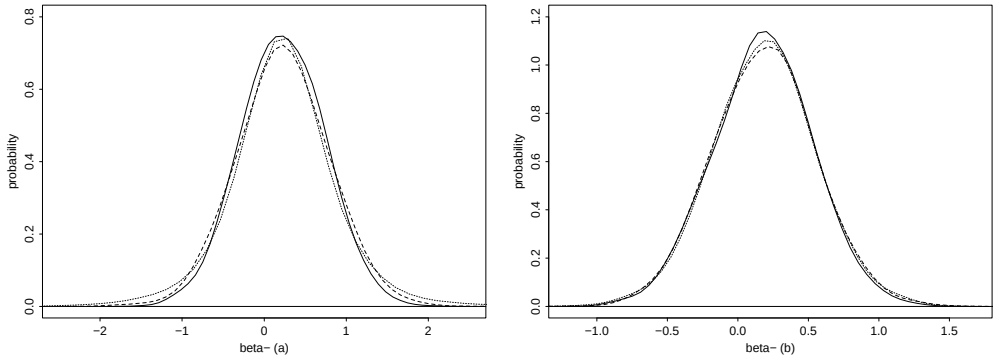


Figure 3.3: *Posterior distribution of the mean β. The continuous curve represents the posterior obtained with the covariance parameters α and ν fixed. This posterior is a gaussian distribution. The dashed curve is obtained letting α unknown. This posterior is a t distribution that has the same mean as the previous gaussian but a different and slightly higher variance. The ratio of the variance of the t distribution and the gaussian is approximately $\eta/\eta - 2$, that is 1.07. The dotted curve represents the posterior with both the covariance parameters unknown. In (a) the noninformative prior is used and to obtain the continuous curve α is set to 0.97 and ν to 46.01m. In (b) the conjugate prior is used and to obtain the continuous curve α is set to 1.19 and ν to 28.92m (these values are the posterior means of α and ν). The scales in the two figures are not the same. This was done to highlight the differences between the curves in each figure*

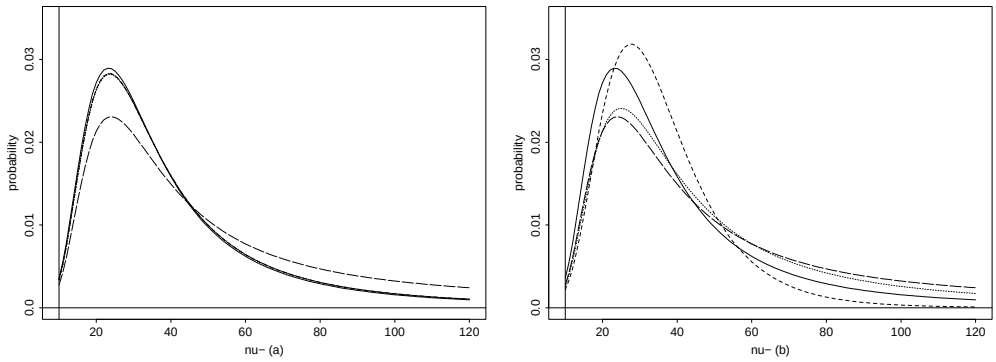


Figure 3.4: *Posterior distribution of the scale parameter ν obtained with both the noninformative (large dashed line) and conjugate prior (continuous curve) for β and α. A larger posterior variance is obtained with the noninformative prior. In (a) we let the prior mean and variance of the conjugate prior for β vary. The posterior distribution of ν is insensible to these changes: the curves obtained almost coincide. It is clear that the posterior distribution of ν depends mostly on the prior for α. This is evident in (b). The dashed curve is obtained with a prior for α that has the same mean (=1) but a smaller variance than the prior used to obtain the continuous curve. The posterior mean and variance of ν change for changes in the prior variance of α. The dotted curve is introduced to show that a change in the prior mean of α also changes the posterior of ν*

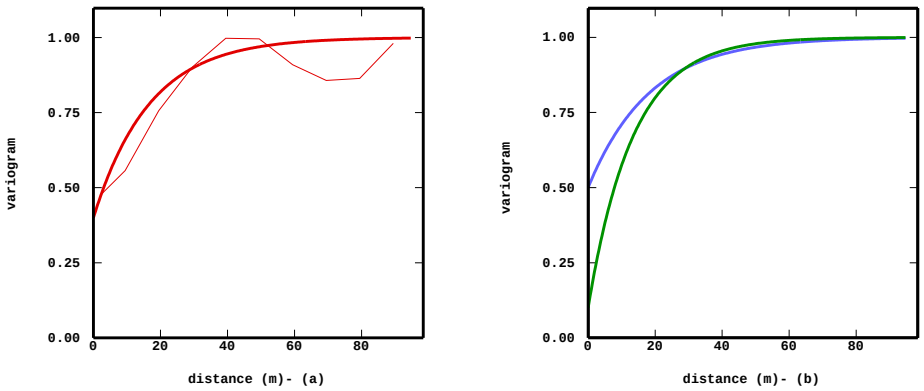


Figure 3.5: *Variogram model used for the simulations: a nugget effect of 0.4 plus an exponential model with a sill of 0.6 and practical range of 90m (a). The experimental variogram of the data is also presented. Given the data different variogram models may be chosen. Two of these are presented in (b). For example, if the first point is neglected in the experimental variogram as the number of pairs is low compared to the second and third points, the nugget effect would be set at around 0.1, while if the first point is kept it would be set at around 0.4*

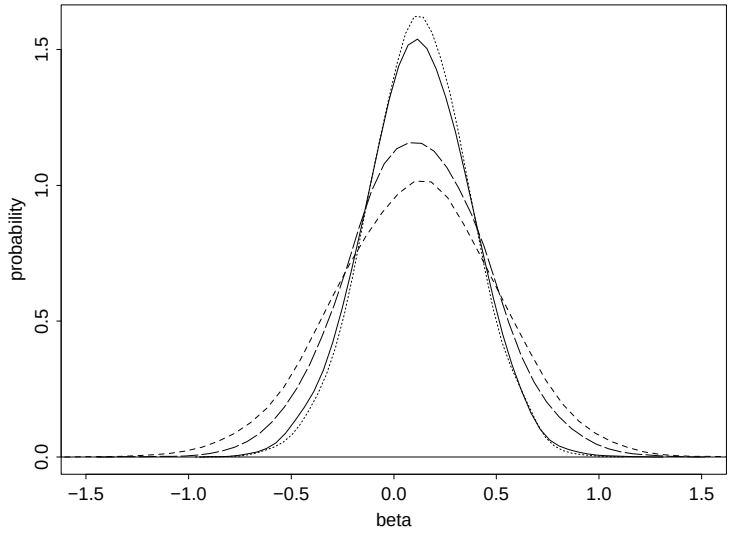


Figure 3.6: *Posterior distribution of the mean parameter β. The continuous curve is obtained using the true model. The dotted curve is obtained taking into account the uncertainty on the covariance parameters. The two curves are almost identical. The large dashed curve is obtained using a different model that graphically seemed to fit well the experimental variogram but has a smaller nugget effect ($= 0.1$). And finally the dashed curve is obtained defining a covariance model with no nugget effect. The fact of not considering a nugget effect when on the contrary it should be taken into account gives a posterior distribution of β that largely overestimates the uncertainty*

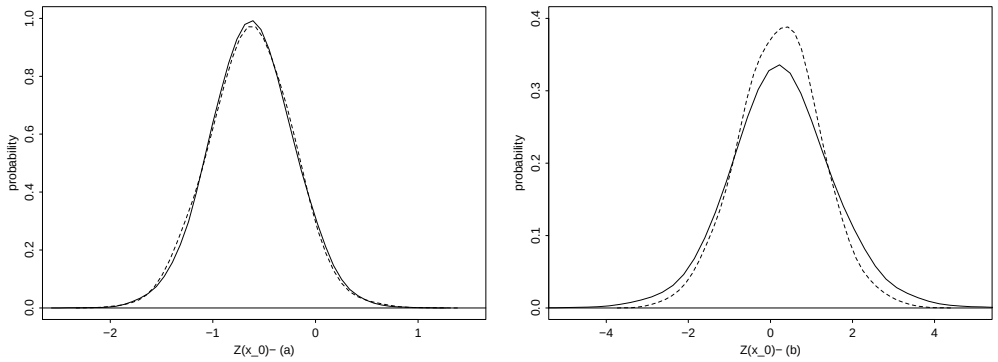


Figure 3.7: *Predictive distribution. The continuous curve is obtained with the noninformative prior for β and α and the gaussian prior for ν. The dashed curve is obtained with the conjugate prior for β and α and the gaussian prior for ν. In (a) the predictive distribution is computed for $x_0 = 55m$. In (b) the predictive distribution is computed for $x_0 = 400m$. It is evident that when the uncertainty about z_0 is great prior information has an important role*

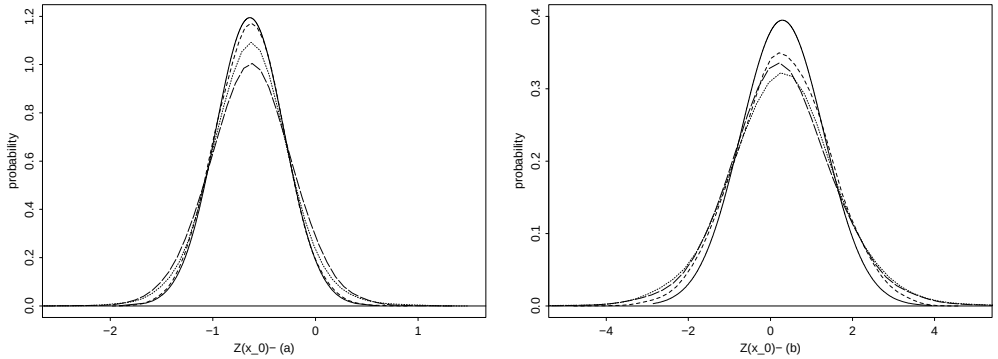


Figure 3.8: *Predictive distributions obtained with the noninformative prior for β and α and the gaussian prior for ν at (a) $x_0 = 55m$ and at (b) $x_0 = 400m$. The continuous curve is computed with the 3 parameters fixed and thus supposed known. The parameters are set equal to the mean of their posterior distributions, that is for the mean β 0.23, for the precision α 0.97 and for the scale parameter ν 46.01m (values obtained in the previous section). The predictor corresponds to the simple kriging predictor. The dashed curve is computed letting β unknown. The predictor corresponds to the ordinary kriging predictor. The dotted curve is obtained letting both β and α unknown. And finally the large dashed curve is obtained letting the 3 parameters unknown. The predictor corresponds to the Bayesian predictor. The Bayesian predictive distribution has a larger variance and heavier tails*

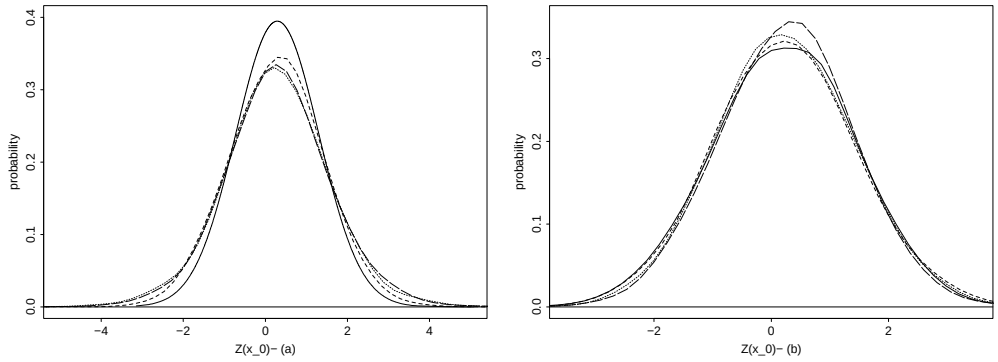


Figure 3.9: *Predictive distribution at $x_0 = 450m$ obtained with the noninformative prior for β and α and the gaussian prior for ν. In (a), as in the previous graph, the continuous curve is obtained with the 3 parameters fixed at their posterior mean. The dashed curve is computed letting β unknown. The dotted curve is obtained letting both β and α unknown. And finally the large dashed curve is obtained letting the 3 parameters unknown. This last distribution, the Bayesian predictive distribution, has a larger variance and heavier tails. In (b) predictive distributions obtained with β unknown and with different values for α and ν are compared. The corresponding predictors are the ordinary kriging predictors*

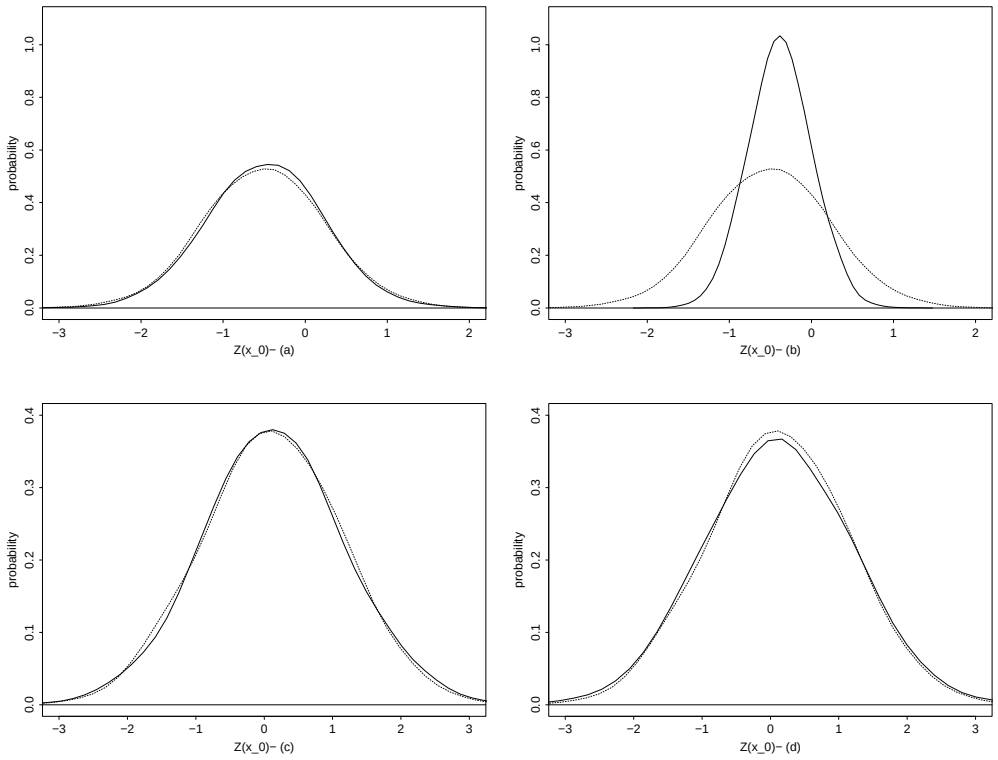


Figure 3.10: *Predictive distributions with the noninformative prior prior for β and α and the gaussian prior for ν. In (a) and (b) $x_0 = 55.5m$; in (c) and (d) $x_0 = 400m$. The continuous curve represents the predictive distribution with β unknown and fixed covariance parameters. The dotted curve represents the Bayesian predictive distribution. In (a) and (c) the Bayesian predictive distribution is compared to the predictive distribution based on the true model. In (b) and (d) the Bayesian predictive distribution is compared to the* plug-in *distribution based on a model with a smaller nugget effect*

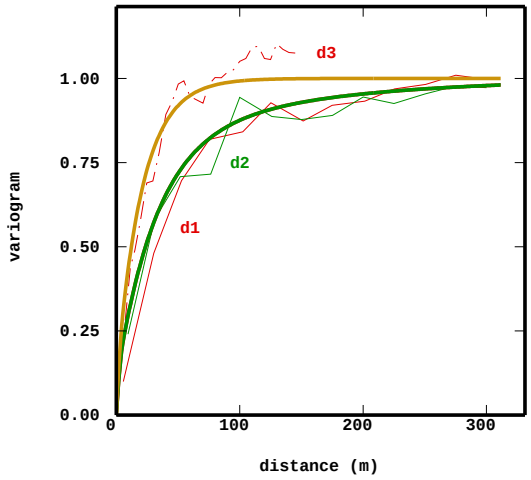


Figure 3.11: *Experimental and variogram model for the gaussian variable in the upper zone*

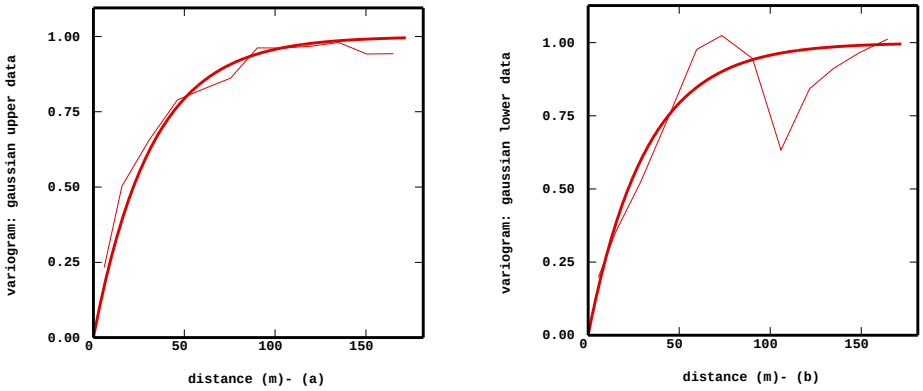


Figure 3.12: *Experimental and variogram model for the gaussian variable in (a) the upper zone and in (b) the lower zone. An exponential model with a sill of 1 and a practical range of 90 fits both variograms*

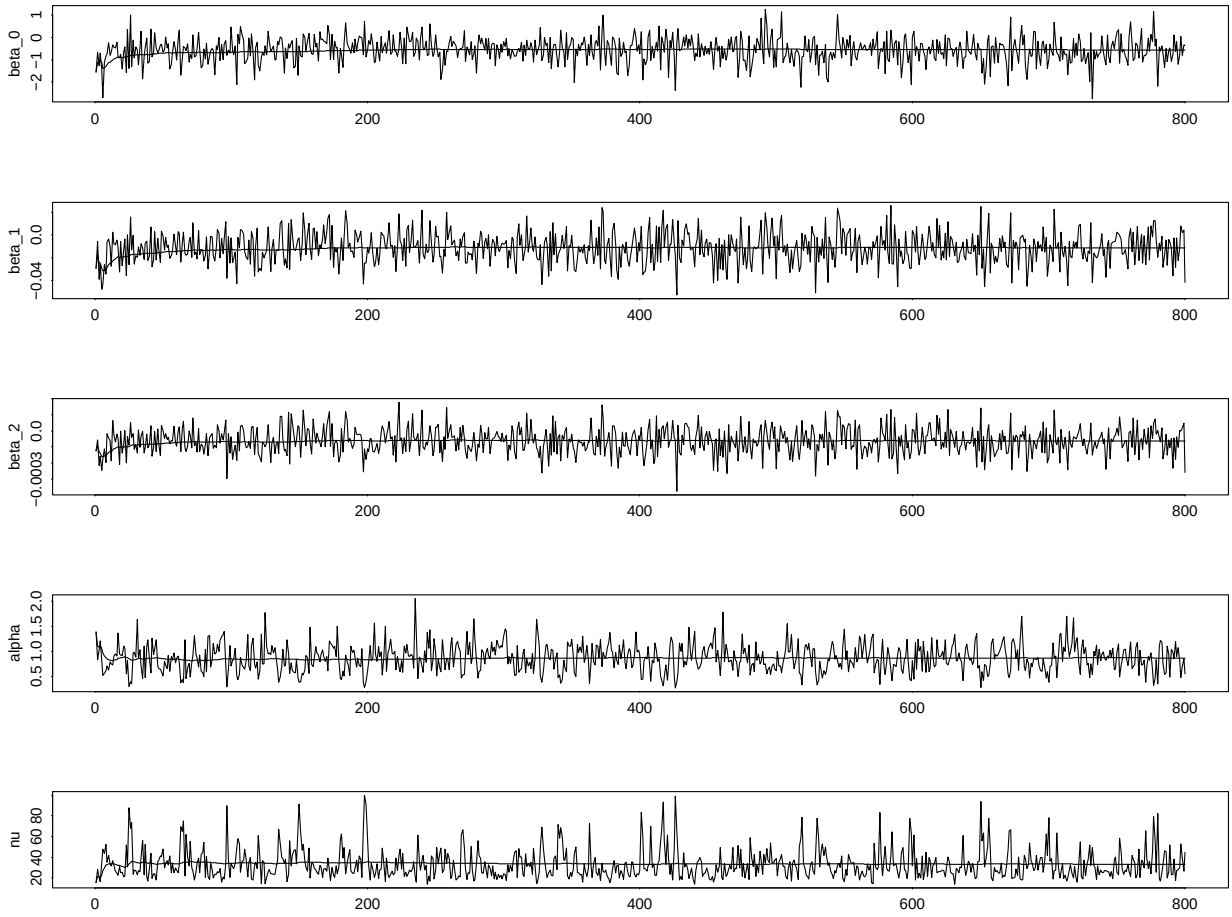


Figure 3.13: *Time series plot of the MCMC output every 50th iteration of the model parameters given* $\mathbf{z}$ *and the evolution of their estimates as a function of the number of iterations. The chains seem stationary. The Monte Carlo estimate of the posterior mean of* α *is 0.87 and of* ν *is 33.44*

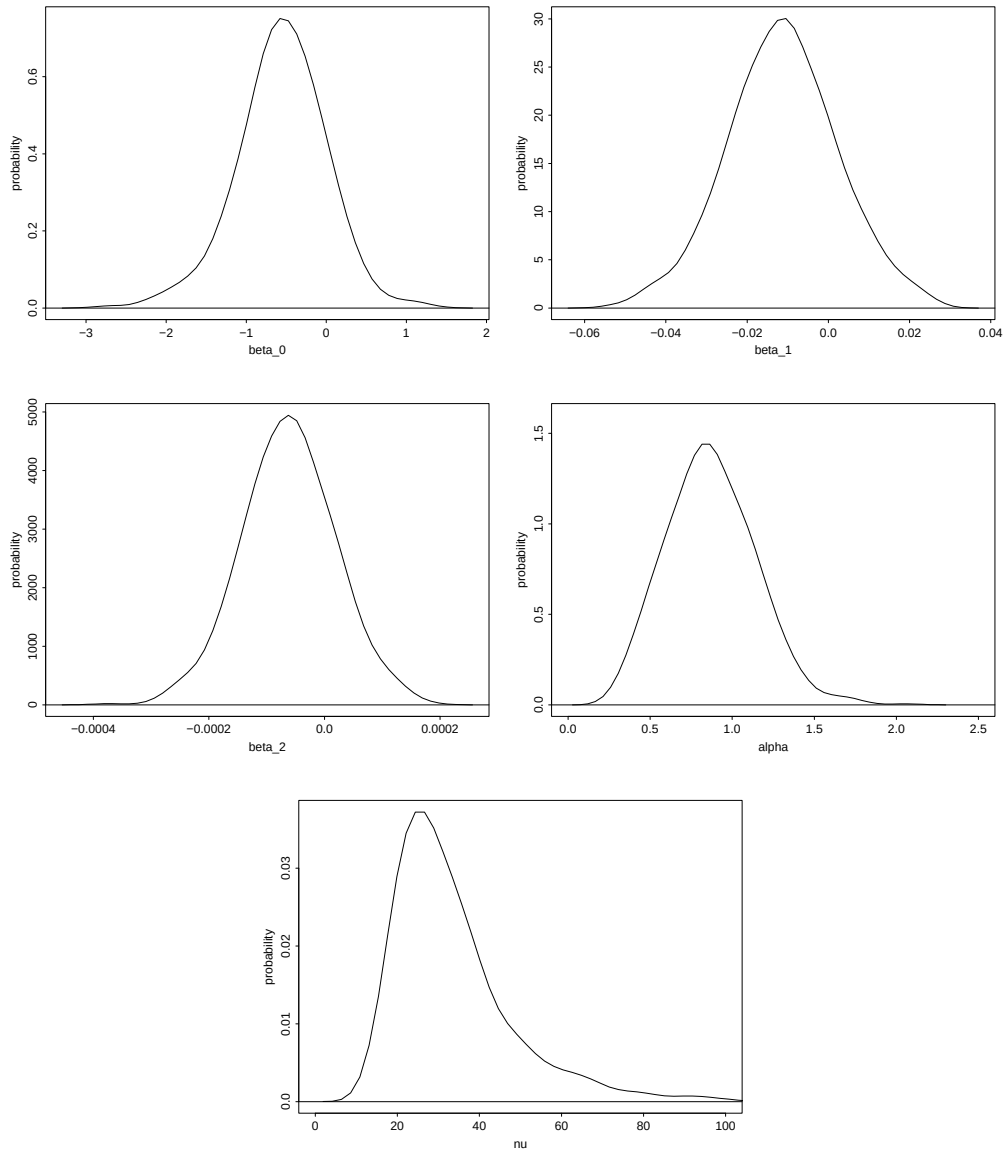


Figure 3.14: *Posterior distribution for the model parameters*

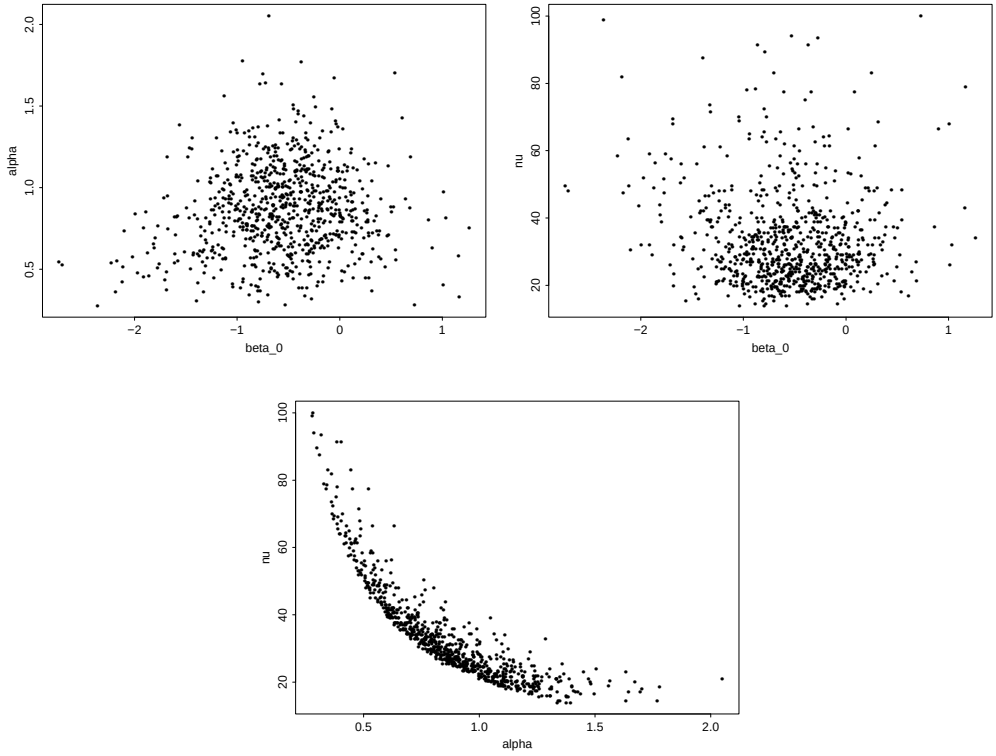


Figure 3.15: *Correlation between the parameters using samples from the posterior distribution. The samples of the mean parameter β_0 seem independent of the covariance parameters. Note the strong dependence between α and ν*

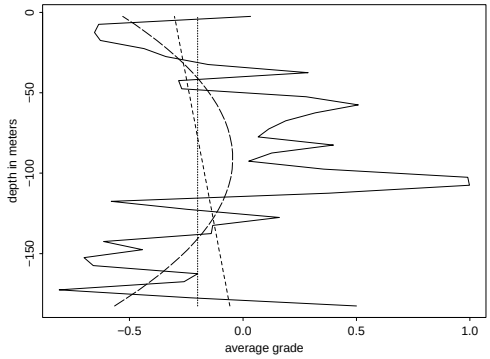


Figure 3.16: *The continuous line represents the average grade for each depth x_3. The figure shows the following estimated drifts: $\beta_0 + \beta_1 x_3 + \beta_2 x_3^2$ (large dashes), $\beta_0 + \beta_1 x_3$ (dashed line), β_0 that is a constant mean (dotted line)*

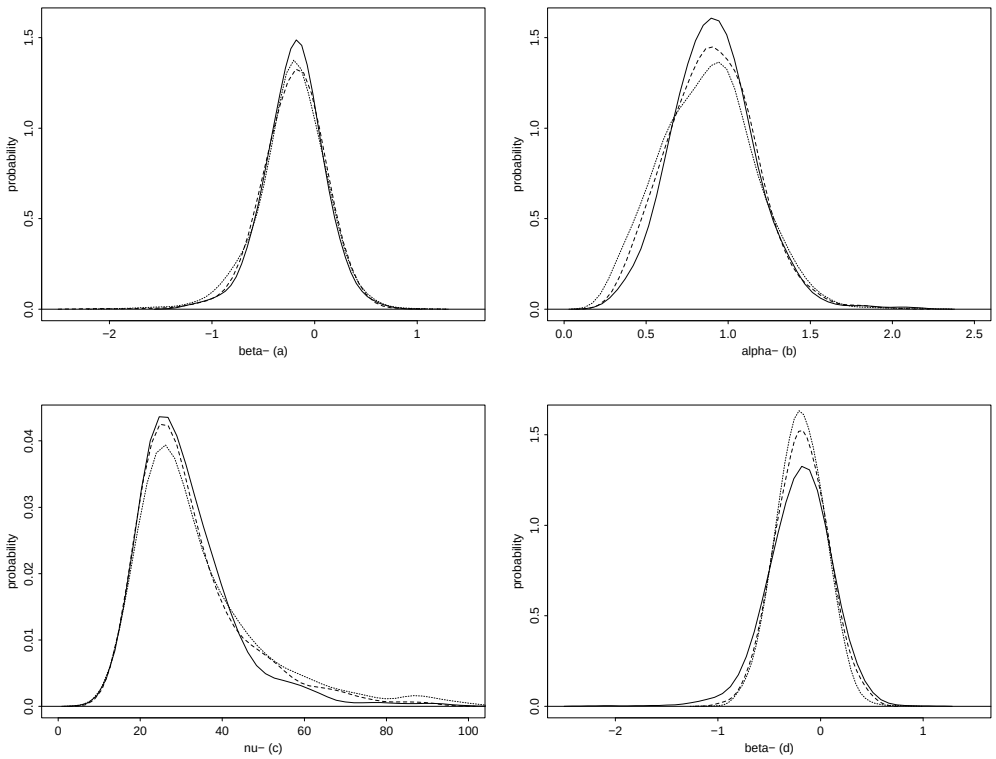


Figure 3.17: *Posterior distribution for the model parameters. In (a), (b) and (c) the posterior distributions for the mean parameter β, the precision parameter α and the correlation parameter ν are presented, respectively for the independent case (continuous line), conjugate case (dashed line) and the noninformative case (dotted line). In (d) the influence of accounting for the uncertainty on the covariance parameters is highlighted (conjugate case): the distributions of β with the 2 covariance parameters unknown (continuous line), with α unknown and ν fixed at 30 (dashed line) which is given in (3.15) and with α and ν fixed at 1 and 30 respectively (dotted line) which is a gaussian distribution, are compared. The last two curves have both a mean of -0.20 and variance of 0.07*

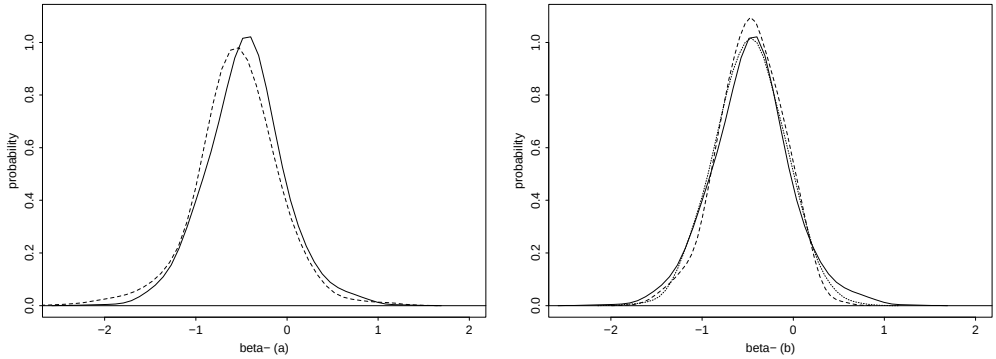


Figure 3.18: *Posterior distribution for the mean parameter β given only 20 observations. In (a) the distributions for the conjugate case (continuous line) and for the noninformative case (dashed line) are compared. In (b) the importance of taking into account the uncertainty on the covariance parameters is highlighted (conjugate case): the distributions of β with the 2 covariance parameters unknown (continuous line), with α unknown and ν fixed (dashed line) and with both α and ν fixed (dotted line) are compared*

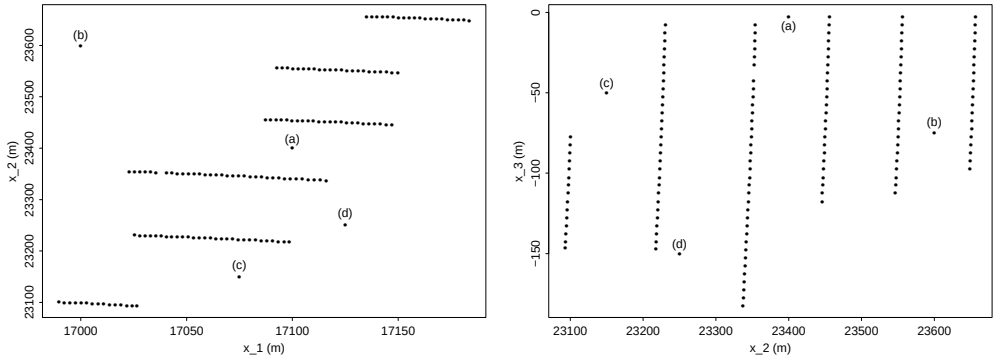


Figure 3.19: *Location of the lower data and of the four points where prediction is carried on*

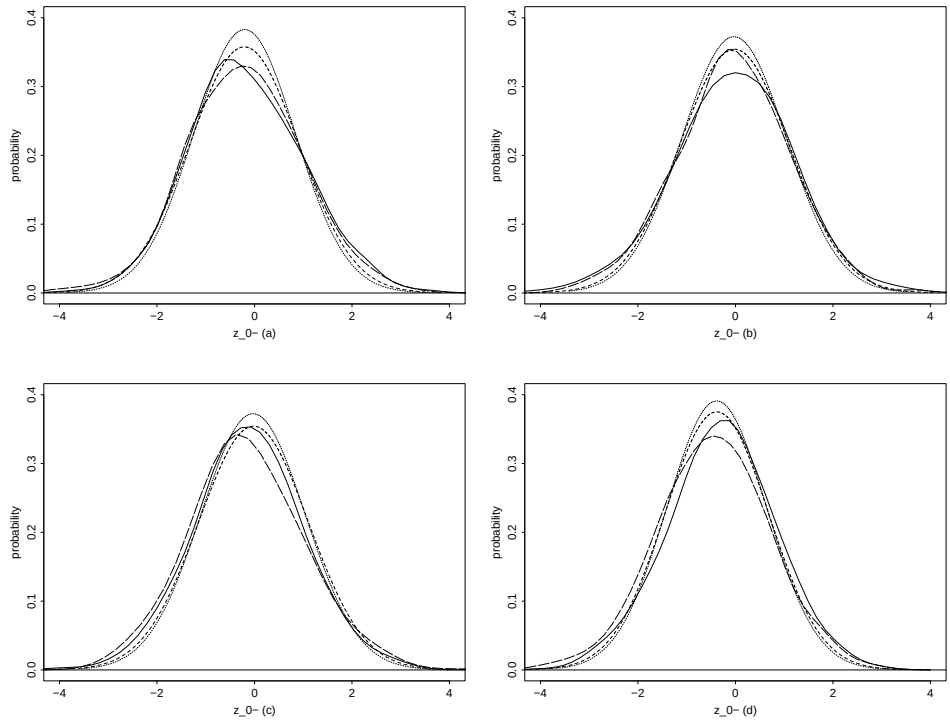


Figure 3.20: *Predictive distributions at (a)* $x_0 = (17100, 23400, -2.5)$*, (b)* $x_0 = (17000, 23600, -75)$*, (c)* $x_0 = (17025, 23150, -50)$ *and (d)* $x_0 = (17125, 23250, -150)$*. The continuous curve is obtained with uncertainty on all parameters taken into account. The large dashed curve is obtained with uncertainty on* β *and* α *taken into account while* ν *is fixed at its posterior mean. The dashed curve is obtained with uncertainty on* β *taken into account while* α *and* ν *are fixed at their posterior mean: as a noninformative prior is defined for* β *the predictor corresponds to the UK predictor and the prediction variance to the kriging variance. Finally the dotted curve is obtained with* β*,* α *and* ν *fixed at their posterior mean. The predictor corresponds to the SK predictor and the prediction variance to the kriging variance. Note that in (b) the Bayesian predictive distribution has heavier tails than the* plug-in *distributions. In (a), (c) and (d) they are very close*

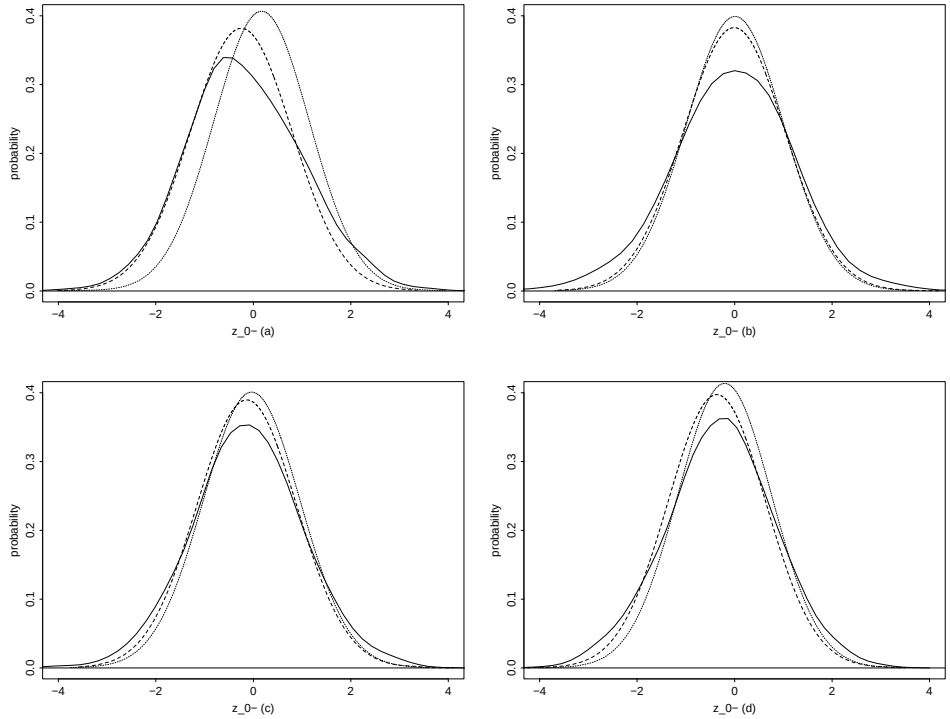


Figure 3.21: *Predictive distributions at (a)* $x_0 = (17100, 23400, -2.5)$*, (b)* $x_0 = (17000, 23600, -75)$*, (c)* $x_0 = (17025, 23150, -50)$ *and (d)* $x_0 = (17125, 23250, -150)$*. The continuous curve is obtained with uncertainty on all parameters taken into account. The dashed curve is obtained with uncertainty on* $\boldsymbol{\beta}$ *taken into account while* $\alpha = 1$ *and* $\nu = 30$*: as a noninformative prior is defined for* $\boldsymbol{\beta}$ *the posterior mean corresponds to the UK predictor and the variance to the kriging variance. Finally the dotted curve is obtained with* $\boldsymbol{\beta} = (0, 0, 0)^T$*,* $\alpha = 1$ *and* $\nu = 30$*: the posterior mean corresponds to the SK predictor and the posterior variance to the kriging variance*

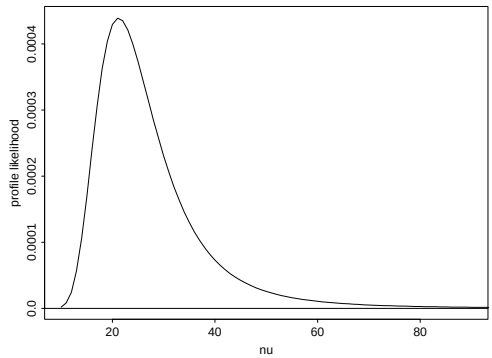


Figure 3.22: *Profile likelihood function for* ν*. The ML estimate is 21*

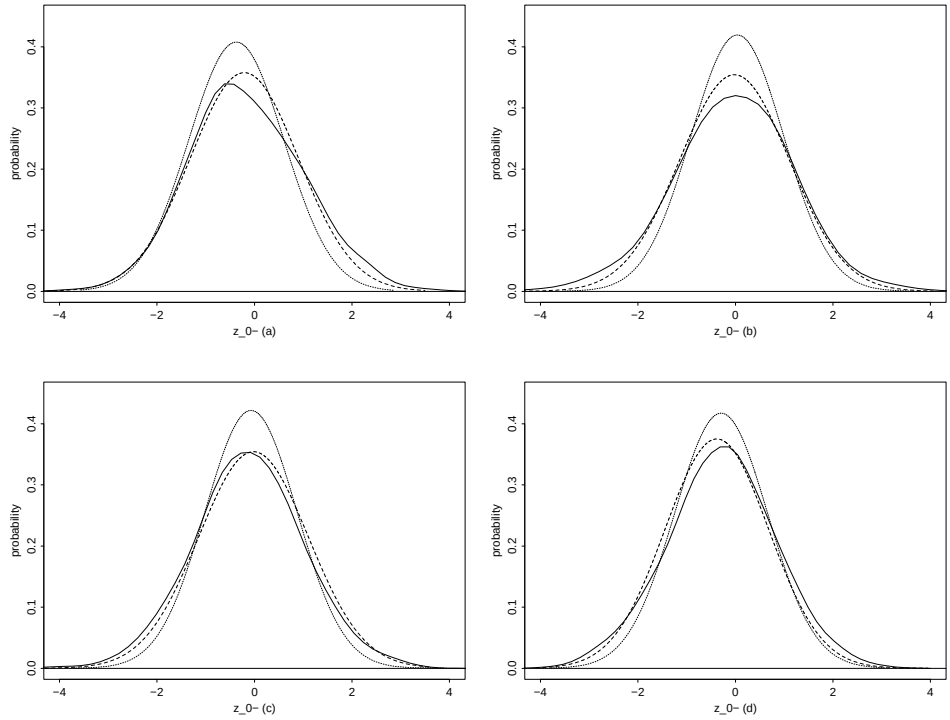


Figure 3.23: *Predictive distributions at (a)* $x_0 = (17100, 23400, -2.5)$*, (b)* $x_0 = (17000, 23600, -75)$*, (c)* $x_0 = (17025, 23150, -50)$ *and (d)* $x_0 = (17125, 23250, -150)$*. The continuous curve represents the Bayesian predictive distribution. The dashed curve represents the plug-in distribution with* α *and* ν *fixed at their posterior mean. Finally the dotted curve represents the plug-in distribution with* α *and* ν *fixed at their ML estimate*

3.3 Transformed gaussian random fields

The Gaussian assumption is often inappropriate for analysing geostatistical data. Transformations can then be used in an attempt to get nearly gaussian behaviour. Diggle *et al.* (1998) proposed a generalised linear mixed model which, like the original generalised linear model, accommodates non gaussian data. However, in the geostatistical setting there are technical difficulties in fitting the model (Christensen *et al.*, 2000). The simplicity of the transformed gaussian model, proposed by De Oliveira *et al.* (1997), makes it an attractive alternative. The Bayesian transformed gaussian model combines the Box- Cox family of power transformations with the Gaussian random field model in a Bayesian framework. However, the range of the transformed data depends on the transformation parameter. Thus letting this parameter unknown, the range of the transformed data varies at each step of MCMC.

The Box- Cox family of power transformations (Box and Cox, 1964) is frequently used to normalize positive data

$$g_\lambda(z) = \begin{cases} \frac{z^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log z & \text{if } \lambda = 0 \end{cases}$$

It must be observed that the model is compatible with normality only for $\lambda = 0$. For any $\lambda \neq 0$, the range of $g_\lambda(\cdot)$ is not $(-\infty, \infty)$, but a proper subinterval of the real line: if $\lambda > 0$, the range of $g_\lambda(\cdot)$ is $(-\frac{1}{\lambda}, \infty)$, while if $\lambda < 0$ the range of this transformation is $(-\infty, \frac{1}{\lambda})$. Strictly speaking, a truncated gaussian model in which one replaces $G(dy)$ by $\frac{1}{G(\mathcal{D})}\, I_{y\in\mathcal{D}}\, G(dy)$ where $\mathcal{D}$ is the image $g_\lambda(\mathbb{R}^+)$ should be used. Nevertheless, this model is a sensible approximation provided we restrict consideration to appropriate values of λ. We need that $(\lambda g_\lambda(z) + 1) > 0$, thus for $\lambda > 0$, we need that $P[g_\lambda(z) < -\frac{1}{\lambda}] \approx 0$, while if $\lambda < 0$ we need that $P[g_\lambda(z) > \frac{1}{\lambda}] \approx 0$.

The resulting parametric family of probability distributions is fairly rich. It includes sampling distributions that are gaussian- like (for λ close to 1), but also sampling distributions that are skewed with tails heavier than the gaussian (for λ close to 0). Figure 3.24 presents the histogram of the upper data (2244 observations) and of these transformed data for 3 different values of λ. It can be noted that for λ close to 0.2 we obtain a nearly gaussian behavior, while for λ close to 0 we get negative skewed values and for λ closer to 1 positive skewed values. Note that this approach permits to test the lognormal model. Another approach could be that of limiting λ to a few values, for example $(-1, 0, 0.5, 1)$, so that the transformed predictors can be interpreted as the reciprocal, the logarithm, the square root, and the untransformed response (Hoeting and Ibrahim, 1998).

3.3.1 The model

We assume that for some unknown transformation parameter λ, $\{Y(x) = g_\lambda(Z(x)), x \in \mathcal{D}\}$ is close to a gaussian random field with mean $\boldsymbol{\beta}^T f(x)$ and covariance $\alpha^{-1}\, K_\nu(x, x')$. Let $\mathbf{Y} = (g_\lambda(z_1), \ldots, g_\lambda(z_n))^T$, we then have that

$$(\mathbf{Y}|\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_n(F\boldsymbol{\beta}, \alpha^{-1} K_\nu).$$

The likelihood function of the model parameters $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})^T$ based on the original untransformed observations $\mathbf{z}$ is given by

$$L(\boldsymbol{\theta}|\mathbf{z}) \propto \alpha^{\frac{n}{2}}\, |K_\nu|^{-\frac{1}{2}}\, \exp\left(-\frac{\alpha}{2}||g_\lambda(\mathbf{z}) - F\boldsymbol{\beta}||^2_{K_\nu^{-1}}\right)\, J(\lambda, \mathbf{z})$$

where $J(\lambda, \mathbf{z})$ is the Jacobian of the transformation,

$$J(\lambda, \mathbf{z}) = \prod_{i=1}^{n} \left| \frac{dg_\lambda(z_i)}{dz_i} \right| = \prod_{i=1}^{n} z_i^{\lambda-1}.$$

The Bayesian predictive density function is given in (3.2). It can be observed that in this approach the uncertainty about λ, that is the uncertainty about the sampling distribution, is transferred to the final prediction. The transformation parameter λ is another of the uncertain parameters. As it was said earlier, one limitation of this approach is that for each λ the range of $g_\lambda(\cdot)$ changes.

By the stated assumptions we have that

$$(g_\lambda(\mathbf{z}_0), g_\lambda(\mathbf{z})|\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_{k+n}\left(\begin{pmatrix} F_0\boldsymbol{\beta} \\ F\boldsymbol{\beta} \end{pmatrix}, \alpha^{-1} \begin{pmatrix} N_\nu & L_\nu \\ L_\nu^T & K_\nu \end{pmatrix} \right)$$

where N_ν and L_ν are $k \times k$ and $k \times n$ correlation matrices.

It follows that $(g_\lambda(\mathbf{z}_0)|\mathbf{z}, \lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_k(M_{\lambda,\beta,\nu}, \alpha^{-1}\,\Sigma_{\lambda,\beta,\nu})$ where

$$M_{\lambda,\beta,\nu} = L_\nu K_\nu^{-1} g_\lambda(\mathbf{z}) + (F_0 - L_\nu K_\nu^{-1} F)\boldsymbol{\beta}$$

and

$$\Sigma_{\lambda,\beta,\nu} = N_\nu - L_\nu K_\nu^{-1} L_\nu^T.$$

We then have

$$f(\mathbf{z}_0|\mathbf{z}, \lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \propto \alpha^{\frac{k}{2}}\, |\Sigma_{\lambda,\beta,\nu}|^{-\frac{1}{2}}\, \exp\left(-\frac{\alpha}{2} ||g_\lambda(\mathbf{z}_0) - M_{\lambda,\beta,\nu}||^2_{\Sigma^{-1}_{\lambda,\beta,\nu}} \right) J(\lambda, \mathbf{z}_0). \tag{3.22}$$

We can compute $f(\mathbf{z}_0|\mathbf{z}, \lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$ only for n and k small. If n is large a neighborhood must be defined. Observe that it is a simplification of geostatistical conditional simulations, that were used in Chapter 2, which allow to sample $Z_0 \sim \mathcal{L}(Z_0|\mathbf{z}, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$ for any law $\mathcal{L}$ through an anamorphosis transformation.

First we define the parameters prior distribution. The parameters posterior distribution is then computed. We use $f(\mathbf{z}_0|\mathbf{z}, \lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$ for predictions. Thus we look for the following conditional distributions

- $\pi(\lambda, \boldsymbol{\nu}|\mathbf{z})$;
- $\pi(\alpha|\mathbf{z}, \lambda, \boldsymbol{\nu})$;
- $\pi(\boldsymbol{\beta}|\mathbf{z}, \lambda, \alpha, \boldsymbol{\nu})$.

Gibbs sampling is used to sample from $\pi(\lambda, \boldsymbol{\nu}|\mathbf{z})$. Finally an example and the application to the test case are presented.

The choice of the parameters prior distribution

The choice of the prior distributions for the parameters needs some care as a change in λ magnifies or shrinks all the data and changes the interpretation of $\boldsymbol{\beta}$, α and $\boldsymbol{\nu}$. Box and Cox suggest to take conditional priors of $\boldsymbol{\beta}$, α and $\boldsymbol{\nu}$ given λ, that is

$$\pi(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) = \pi(\lambda)\, \pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}|\lambda),$$

where $\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}|\lambda)$ depends on λ as the general size and range of the transformed observations $g_\lambda(\mathbf{z})$ may depend strongly on λ. If the conditional prior was assumed independent of λ, Box and Cox state that "nonsensical" results would be obtained. However, note that this is the prior distribution. The relationship between the parameters is also defined by the data.

Two different prior distributions will be considered: the noninformative prior proposed by Box and Cox and an informative prior. In addition, using the device of *imaginary training samples* we will define an alternative noninformative prior to that of Box and Cox which is, as we will see, arbitrarily defined.

1. Box and Cox prior. Box and Cox suppose that for a given λ the parameters $(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$ are *a priori* independent. They take

$$\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}|\lambda) \propto g(\lambda)\, \frac{\pi(\boldsymbol{\nu})}{\alpha}, \tag{3.23}$$

where $\pi(\boldsymbol{\nu})$ is the marginal prior of $\boldsymbol{\nu}$ and $g(\cdot)$ is a function to be determined. That is $\boldsymbol{\beta}$ and $\log(\alpha)$ are uniform over the range for which the likelihood is appreciable. The dependence on λ is introduced through a multiplicative factor. The authors determine $g(\lambda)$ in the following way. Let λ_1 be any reference value of λ for which the likelihood is appreciable. For all λ in some neighborhood of λ_1, Box and Cox assume that $g_\lambda(Z)$ is approximately linearly related to $g_{\lambda_1}(Z)$, that is

$$g_\lambda(Z) = \text{const} + l_\lambda g_{\lambda_1}(Z)$$

where l_λ has to be specified. They assume that the mean parameters $\boldsymbol{\beta}$ and $\log(\alpha)$ are uniform over the range for which the likelihood is appreciable. Now the function $g(\lambda)$ is chosen so that when the previous equation holds, the conditional prior (3.23) involving λ and λ_1 is consistent. For α we have that

$$\log(\alpha_\lambda) = \text{const} + \log(\alpha_{\lambda_1})$$

and, hence, to this order, the prior density of α_λ is independent of λ. Analogously for $\boldsymbol{\nu}$. However, the coefficients $\boldsymbol{\beta}$ are linear combinations of the expected values of $g_\lambda(Z)$ so that

$$\frac{d\boldsymbol{\beta}_\lambda}{d\boldsymbol{\beta}_{\lambda_1}} = l_\lambda.$$

Since there are p independent components of $\boldsymbol{\beta}$, it follows that $g(\lambda) = 1/l_\lambda^p$. They argue that $g(\lambda) = J(\lambda,\ \mathbf{z})^{-p/n}$ is a pragmatic choice: in passing from λ_1 to λ, a small element of volume of the n dimensional sample space is multiplied by $J(\lambda,\ \mathbf{z})/J(\lambda_1,\ \mathbf{z})$. An average scale change for a single z component is the nth root of this and since λ_1 is only a standard reference value we have $l_\lambda = J(\lambda,\ \mathbf{z})^{1/n}$.

The prior is finally given by

$$\pi(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) \propto \pi(\lambda)\, J(\lambda, \mathbf{z})^{-\frac{p}{n}}\, \frac{\pi(\boldsymbol{\nu})}{\alpha}.$$

This is an improper distribution. Note that it is outcome- dependent[7]. The choice for l_λ is qualified by Box and Cox as "somewhat arbitrary". This prior distribution is also used by De Oliveira *et al.* (1997). It must be observed that $g(\lambda)$ is a multiplicative factor in the parameters prior distribution. Thus it will influence directly only the posterior distribution of λ that will then influence the other parameters posterior distribution. For this and as we dispose of prior information, we decided to use the upper data to define the prior distribution for the parameters.

2. Informative prior. The prior distribution for the parameters is defined using the upper data $\mathbf{z}(x^*) = \mathbf{z}^*$ (of sample size m). It is given by

$$\pi(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) = \pi(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}|\mathbf{z}^*) = \frac{f(\mathbf{z}^*|\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})\, \pi^N(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})}{\int f(\mathbf{z}^*|\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})\, \pi^N(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})\, d\lambda\, d\boldsymbol{\beta}\, d\alpha\, d\boldsymbol{\nu}}$$

where $\pi^N(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$ is the standard noninformative prior, that is

$$\pi^N(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}) = \frac{1}{\alpha\, \boldsymbol{\nu}}.$$

The prior distributions are then

- $(\boldsymbol{\beta}|\mathbf{z}^*, \lambda, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_p(\mathbf{b}, \alpha^{-1} S)$ where

  $$\mathbf{b} = S\,(F_0^T K_{0,\nu}^{-1} g_\lambda(\mathbf{z}^*)), \quad S = (F_0^T K_{0,\nu}^{-1} F_0)^{-1},$$

  $F_0 = \{f_j(x_i^*)\}_{m\times p}$ and $K_{0,\nu}$ is the correlation matrix for $\mathbf{z}^*$;

- $(\alpha|\mathbf{z}^*, \lambda, \boldsymbol{\nu}) \sim \Gamma(a_1, 1/a_2)$ where

  $$a_1 = \frac{m-p}{2}, \quad a_2 = \frac{R(\lambda, \boldsymbol{\nu})}{2}$$

  and $R(\lambda, \boldsymbol{\nu}) = ||g_\lambda(\mathbf{z}^*) - F_0\mathbf{b}||^2_{K_{0,\nu}^{-1}}$;

- $\pi(\lambda, \nu|\mathbf{z}^*)$ is given by

  $$\pi(\lambda, \boldsymbol{\nu}|\mathbf{z}^*) = \frac{\pi(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu}|\mathbf{z}^*)}{\pi(\boldsymbol{\beta}, \alpha|\lambda, \boldsymbol{\nu}, \mathbf{z}^*)}$$

  and thus

  $$\pi(\lambda, \boldsymbol{\nu}|\mathbf{z}^*) \propto \frac{1}{\boldsymbol{\nu}}\, J(\lambda, \mathbf{z}^*)\, |K_{0,\nu}|^{-\frac{1}{2}}\, |S|^{\frac{1}{2}}\, R(\lambda, \boldsymbol{\nu})^{-\frac{m-p}{2}}. \tag{3.24}$$

Now *a priori* the parameters are not independent. Given λ and $\boldsymbol{\nu}$ we have a conjugate prior for $\boldsymbol{\beta}$ and α. As in the previous section a quadratic drift and an exponential covariance function are defined. Figure 3.25 presents these priors for β_0 and α for different values of λ. It can be noted that both β_0 and α depend strongly on λ.

[7] Pericchi (1981) and Sweeting (1984) suggest different prior distributions which do not depend on the data. Pericchi's prior has some major drawbacks discussed by Sweeting. In particular, for a given λ, the improper prior for the parameters arises as a limiting case of the family of conjugate prior distributions. This is inappropriate because the posterior distribution cannot be thought of as an approximation to a posterior distribution based on a vague but proper conjugate prior. As Sweeting states, this arises because the passage to the limit is taken before application of Bayes's theorem, and the resulting posterior distribution is not the probability limit of any sequence of posterior distributions based on proper conjugate priors. The family of conditional prior distributions derived by Sweeting leads to a class of scale- invariant prior distributions as limiting cases. In particular, there is no need to choose arbitrarily a function $g(\lambda)$. It is an interesting alternative to the prior proposed by Box and Cox

3.3.2 Parameters posterior distribution

The parameters posterior distributions for the two priors previously defined are now presented.

1. Box and Cox prior. The conditional distributions $(\boldsymbol{\beta}|\mathbf{z},\lambda,\alpha,\boldsymbol{\nu})$ and $(\alpha|\mathbf{z},\lambda,\boldsymbol{\nu})$ were given in (3.12) and (3.13), respectively. Note that $\mathbf{z}$ must be replaced by $g_\lambda(\mathbf{z})$. In particular, the joint posterior distribution of the transformation and correlation parameters, λ and $\boldsymbol{\nu}$, is

$$\pi(\lambda,\boldsymbol{\nu}|\mathbf{z}) \propto \pi(\lambda)\,\pi(\boldsymbol{\nu})\,J(\lambda,\mathbf{z})^{1-\frac{p}{n}}\,|K_\nu|^{-\frac{1}{2}}\,|S_c|^{\frac{1}{2}}\,R(\lambda,\boldsymbol{\nu})^{-\frac{n-p}{2}}$$

where $R(\lambda,\boldsymbol{\nu}) = ||g_\lambda(\mathbf{z}) - F\mathbf{b}_c||^2_{K^{-1}}$. As it was said earlier $g(\lambda)$ only influences the posterior distribution of λ. To sample from this posterior distribution Gibbs sampling is used.

2. Informative prior. As we have seen this prior corresponds to a conjugate prior for $\boldsymbol{\beta}$ and α. We suppose that Z^* and Z are two independent random variables although identically distributed. It is an approximation as we have supposed (see Chapter 2) that the data are spatially correlated and come from the same random variable. That is, we compute

$$\pi(\lambda,\boldsymbol{\beta},\alpha,\boldsymbol{\nu}|\mathbf{z},\mathbf{z}^*) \propto \pi(\lambda,\boldsymbol{\beta},\alpha,\boldsymbol{\nu}|\mathbf{z}^*)f(\mathbf{z}|\lambda,\boldsymbol{\beta},\alpha,\boldsymbol{\nu})$$

instead of

$$\pi(\lambda,\boldsymbol{\beta},\alpha,\boldsymbol{\nu}|\mathbf{z},\mathbf{z}^*) \propto \pi(\lambda,\boldsymbol{\beta},\alpha,\boldsymbol{\nu}|\mathbf{z}^*)f(\mathbf{z}|\mathbf{z}^*,\lambda,\boldsymbol{\beta},\alpha,\boldsymbol{\nu}).$$

The conditional distributions $(\boldsymbol{\beta}|\mathbf{z},\mathbf{z}^*,\lambda,\alpha,\boldsymbol{\nu})$ and $(\alpha|\mathbf{z},\mathbf{z}^*,\lambda,\boldsymbol{\nu})$ were given in (3.9) and (3.10), respectively. Note that the prior parameters $\mathbf{b}, S$ and a_1, a_2 now depend on $\mathbf{z}^*$, λ and $\boldsymbol{\nu}$. The joint posterior distribution for λ and $\boldsymbol{\nu}$ is given by

$$\pi(\lambda,\boldsymbol{\nu}|\mathbf{z},\mathbf{z}^*) \propto \pi(\lambda,\boldsymbol{\nu}|\mathbf{z}^*)\,J(\lambda,\mathbf{z})\,|K_\nu|^{-\frac{1}{2}}\,|S_c|^{\frac{1}{2}}\,\left(a_2 + \frac{S(\lambda,\boldsymbol{\nu})}{2}\right)^{-(a_1+\frac{n}{2})}$$

where $S(\lambda,\boldsymbol{\nu}) = \mathbf{b}^T S^{-1}\mathbf{b} + g_\lambda(\mathbf{z})^T K_\nu^{-1} g_\lambda(\mathbf{z}) - \mathbf{b}_c^T\,S_c^{-1}\,\mathbf{b}_c$.

Now if the data $\mathbf{z}^*$ were not available, the device of *imaginary training samples* and the expected- posterior prior could be used to convert improper noninformative priors into proper distributions. In this case the conditional prior of (3.23) could be obtained without having to use the prior suggested by Box and Cox. The expected- posterior prior is given by

$$\pi^*(\boldsymbol{\theta}) = \int \pi^N(\boldsymbol{\theta}|\mathbf{z}^*)\,m^*(\mathbf{z}^*)\,d\mathbf{z}^*$$

where $m^*(\mathbf{z}^*)$ is a predictive measure for the imaginary data $\mathbf{z}^*$. It can be defined from beliefs as to how a training sample would behave, that is from subjective knowledge about the problem. For choices of m^* see Pérez and Berger (2002). The training sample is assumed independent of Z and often Z_i and Z_i^* are assumed independent identically distributed random variables on a common sample space. The expected-posterior prior can be seen as a hierarchical prior with the first- stage prior being $\pi^N(\boldsymbol{\theta}|\mathbf{z}^*)$ and the second stage being $m^*(\mathbf{z}^*)$ with $\mathbf{z}^*$ viewed as an hyperparameter. One possibility to deal with expected-posterior priors, proposed by Pérez and Berger (2002), is then to introduce $\mathbf{z}^*$ as a latent variable and consider joint generation of $\boldsymbol{\theta}$ and $\mathbf{z}^*$ from the density

$$\pi(\boldsymbol{\theta},\mathbf{z}^*|\mathbf{z}) = \pi(\boldsymbol{\theta}|\mathbf{z},\mathbf{z}^*)\,m^*(\mathbf{z}^*)$$

that is proportional to[8]

$$f(\mathbf{z}|\boldsymbol{\theta})\,\pi^N(\boldsymbol{\theta}|\mathbf{z}^*)\,m^*(\mathbf{z}^*) = f(\mathbf{z}|\boldsymbol{\theta})\,f(\mathbf{z}^*|\boldsymbol{\theta})\,\pi^N(\boldsymbol{\theta})\,m^*(\mathbf{z}^*)/m^N(\mathbf{z}^*)$$

where $m^N(\mathbf{z}^*) = \int f(\mathbf{z}^*|\boldsymbol{\theta})\,\pi^N(\boldsymbol{\theta})\,d\boldsymbol{\theta}$ is the marginal density of $\mathbf{z}^*$ under $\pi^N(\boldsymbol{\theta})$. Given the observations $\mathbf{z}$ and starting values for $\boldsymbol{\theta}$ and $\mathbf{z}^*$, the following Gibbs sampler will generate dependent random variables from this target distribution

- generate $\boldsymbol{\theta}^{(l)}$ from $\pi(\boldsymbol{\theta}|\mathbf{z}, \mathbf{z}^{*(l-1)})$;
- generate $\mathbf{z}^{*(l)}$ from $f'(\mathbf{z}^*|\boldsymbol{\theta}^{(l)}) \propto m^*(\mathbf{z}^*)\,f(\mathbf{z}^*|\boldsymbol{\theta}^{(l)})/m^N(\mathbf{z}^*)$;
- repeat until convergence.

A Metropolis- Hastings algorithm may be required to generate $\mathbf{z}^*$.

This method can also be useful, for example, to introduce the knowledge of the geologist: it is easier to give values for the expected results than to define prior distributions for the unknown parameters.

3.3.3 Prediction of $\mathbf{Z}_0$

As we want to highlight the importance of accounting for parameters uncertainty we will use the conditional distribution $f(\mathbf{z}_0|\mathbf{z}, \lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\nu})$ that was given in (3.22) for predictions. However, observe that given the prior distributions that were considered it is possible to obtain $f(\mathbf{z}_0|\mathbf{z}, \lambda, \boldsymbol{\nu})$. As $(\boldsymbol{\beta}|\mathbf{z}, \lambda, \alpha, \boldsymbol{\nu})$ has a gaussian distribution, it follows that

$$(g_\lambda(\mathbf{z}_0)|\mathbf{z}, \lambda, \alpha, \boldsymbol{\nu}) \sim \mathcal{N}_k(M_{\lambda,\nu}, \alpha^{-1}\,\Sigma_{\lambda,\nu})$$

where

$$M_{\lambda,\nu} = L_\nu K_\nu^{-1} g_\lambda(\mathbf{z}) + (F_0 - L_\nu K_\nu^{-1} F)\mathbf{b}_c$$

and

$$\Sigma_{\lambda,\nu} = (N_\nu - L_\nu K_\nu^{-1} L_\nu^T) + (F_0 - L_\nu K_\nu^{-1} F)\,S_c\,(F_0 - L_\nu K_\nu^{-1} F)^T.$$

We then have that

$$f(\mathbf{z}_0|\mathbf{z}, \lambda, \alpha, \boldsymbol{\nu}) \propto \alpha^{\frac{k}{2}}\,|\Sigma_{\lambda,\nu}|^{-\frac{1}{2}}\,\exp\left(-\frac{\alpha}{2}||g_\lambda(\mathbf{z}_0) - M_{\lambda,\nu}||^2_{\Sigma^{-1}_{\lambda,\nu}}\right)\,J(\lambda, \mathbf{z}_0).$$

Note that for $\lambda = 0$, for simplicity let $k = 1$, the optimal predictor, that is[9]

$$E[\mathbf{z}_0|\mathbf{z}, \alpha, \boldsymbol{\nu}] = \exp\left(M_{\lambda=0,\nu} + \frac{1}{2\alpha}\Sigma_{\lambda=0,\nu}\right)$$

[8] Note that

$$\pi(\boldsymbol{\theta}, \mathbf{z}^*|\mathbf{z}) = \frac{f(\mathbf{z}|\boldsymbol{\theta})\,\pi^N(\boldsymbol{\theta}|\mathbf{z}^*)\,m^*(\mathbf{z}^*)}{m(\mathbf{z})} \text{ and } \pi(\boldsymbol{\theta}|\mathbf{z}, \mathbf{z}^*) = \frac{f(\mathbf{z}|\boldsymbol{\theta})\,\pi^N(\boldsymbol{\theta}|\mathbf{z}^*)}{m(\mathbf{z})}$$

[9] If Z is a log- normal random variable, that is $Z = \exp(Y)$ with $Y \sim \mathcal{N}(m_Y, \sigma_Y^2)$ then $E[Z] = \exp(m_Y + \sigma_Y^2/2)$ and $\text{Var}[Z] = E[Z]^2(\exp(\sigma_Y^2) - 1)$

and the corresponding predictive variance, that is

$$\text{Var}[\mathbf{z}_0|\mathbf{z},\alpha,\boldsymbol{\nu}] = E[\mathbf{z}_0|\mathbf{z},\alpha,\boldsymbol{\nu}]^2 \left(\exp\left(\frac{1}{\alpha}\Sigma_{\lambda=0,\nu}\right) - 1\right)$$

are easily found. If $\boldsymbol{\beta}$ was known, the predictor

$$E[\mathbf{z}_0|\mathbf{z},\beta,\alpha,\boldsymbol{\nu}] = \exp\left(M_{\lambda=0,\beta,\nu} + \frac{1}{2\alpha}\Sigma_{\lambda=0,\beta,\nu}\right)$$

would correspond to the log- normal kriging predictor which in this case would be optimal under the mean squared prediction error. Note that here also the predictor depends on α. Thus the predictor will be sensitive to both changes in α and $\boldsymbol{\nu}$. The Bayesian approach permits to avoid the problems encountered with the optimality of the log- normal kriging predictor for $\boldsymbol{\beta}$ unknown (see Chilès and Delfiner, 1999; pp 190-192). Note that the mean squared error is not a useful measure for asymmetric distributions. The median may be preferred. However, the Bayesian framework provides the entire predictive distribution.

As in the previous section we can integrate $f(\mathbf{z}_0|\mathbf{z},\lambda,\alpha,\boldsymbol{\nu})$ with respect to the precision α. We have that

$$(g_\lambda(\mathbf{z}_0)|\mathbf{z},\lambda,\boldsymbol{\nu}) \sim t_\eta(M_{\lambda,\nu}, s_1^2\,\Sigma_{\lambda,\nu})$$

which is a t density with η degrees of freedom, with location parameter $M_{\lambda,\nu}$ and scale parameter $s_1^2\Sigma_{\lambda,\nu}$. Therefore

$$f(\mathbf{z}_0|\mathbf{z},\lambda,\boldsymbol{\nu}) \propto |2a_{22}\,\Sigma_{\lambda,\nu}|^{-\frac{1}{2}}\,[1 + (g_\lambda(\mathbf{z}_0) - M_{\lambda,\nu})^T(2a_{22}\,\Sigma_{\lambda,\nu})^{-1}(g_\lambda(\mathbf{z}_0) - M_{\lambda,\nu})]^{-\frac{\eta+k}{2}}\,J(\lambda,\mathbf{z}_0).$$

The integration of λ and $\boldsymbol{\nu}$ cannot be done analytically.

Example

This example is introduced to show the effect of taking into account the uncertainty on the transformation parameter. Simulations are carried out to obtain a 1 dimensional data set, **z**, of 30 observations from a log-normal random variable $Z = \exp(Y)$, where Y is a gaussian random variable of mean 1.5 and variance 1. A nugget effect of 0.4 and an exponential covariance with sill 0.6 and scale parameter 30m are defined. The data are simulated on a grid of 290m with a mesh of 10m. The mean of the data is 4.78 and the variance is 13.14. If the log- transformation is considered, the mean of the transformed data is 1.31 and the variance is 0.72. A constant mean β and an exponential covariance model of precision α and scale parameter ν are defined. Note that the model does not take the nugget effect into account.

The noninformative Box and Cox prior distribution is taken. In particular, for ν a discrete uniform prior is defined between 10 and 150. For λ a discrete uniform prior is defined between -1 and 1. The algorithm used to sample from the parameters posterior distribution consists of the following steps

1. choose an initial value for λ in the range specified by the prior.
2. Update ν
   - compute the approximate discrete posterior distribution in the support chosen for ν, that is $\tilde{\pi}(\nu|\mathbf{z},\lambda)$;
   - choose a new value, ν', from $\tilde{\pi}(\nu|\mathbf{z},\lambda)$.

3. Update λ

   - compute the approximate discrete posterior distribution in the support chosen for λ, that is $\tilde{\pi}(\lambda|\mathbf{z}, \nu)$;
   - choose a new value, λ', from $\tilde{\pi}(\lambda|\mathbf{z}, \nu)$.

Steps 2-3 were repeated until convergence was reached. Then we sampled $(\alpha|\mathbf{z}, \lambda, \nu)$ and $(\beta|\mathbf{z}, \lambda, \alpha, \nu)$ every kth $(= 50)$ realisation of the chain. The resulting posterior distributions are presented in Figure 3.26. The mean values of the posterior distributions are 0.05 for λ, 1.38 for β, 0.35 for α and 44.25 for ν. The log-transformation, that corresponds to λ equal to 0, is strongly supported by the data. However, the uncertainty on λ is important and, as it can be seen in Figure 3.27, the histograms of the transformed data for two values of λ that have a high posterior probability, that is -0.1 and 0.2, are quite different.

It is difficult to define a prior distribution that assumes the parameters are dependent. The Box and Cox prior is arbitrarily defined. For this we considered a prior distribution that assumes the parameters are a priori independent. The non informative prior is taken for β and α. The priors defined previously for λ and ν are used. The dashed line in Figure 3.26 represents the posterior distributions obtained with this independent prior. The mean values of the posterior distributions are 0.10 for λ, 1.44 for β, 0.32 for α and 46.50 for ν. The posterior distributions obtained either with the Box and Cox prior or the independent prior are very similar. Figure 3.28 compares the correlations between the transformation parameter and the other parameters using the samples from the posterior distributions for the Box and Cox prior and the independent prior. No significative difference is remarked.

The independent prior is finally kept. Figure 3.29 shows the predictive distributions for $x_0 = 55$m and $x_0 = 350$m. The Bayesian and the *plug- in* predictive distributions are compared. The importance of accounting for the uncertainty on the model parameters is highlighted. In particular the two distributions greatly differ for the isolated point $x_0 = 350$m. Note that the Bayesian distribution has heavier tails. Figure 3.30 compares the Bayesian and the *plug- in* predictive distributions for $x_0 = 55$m. In particular, Figure 3.30(a) highlights the importance of accounting for the uncertainty on the transformation parameter λ. The continuous line represents the Bayesian predictive distribution, the dashed line represents the predictive distribution with $\lambda = 0$ and the covariance parameters unknown and the dotted line represents the predictive distribution with $\lambda = 0$ and β, α and ν fixed at their posterior mean. This last curve corresponds to log-normal kriging. The Bayesian predictive distribution has heavier tails as it accounts for the uncertainty on the transformation. Figure 3.30(b) presents the effects of misspecifying the transformation parameter. The continuous line represents the Bayesian predictive distribution, the dashed line is obtained with $\lambda = -0.5$ and the covariance parameters unknown and the dotted line is obtained with $\lambda = -0.5$ and the covariance parameters fixed at their posterior mean. If λ is not correctly specified, not only the variance of the predictions but also the predictions may be erroneous. In particular the distributions are different for low values whereas they behave similarly for high values. Note that the *plug- in* distributions have tails heavier than the Bayesian distribution. This is due to the fact that the transformation parameter is not correctly specified.

This example highlighted the effect of taking into account the uncertainty on the model parameters. This mostly influences the prediction variance.

The uncertainty on the parameters is not important if the transformation parameter is defined correctly and interest lies in the prediction. On the contrary, taking into account the uncertainty on the parameters has a strong impact if the variance and the tails of the predictive distribution are of interest. This is particularly

true when isolated points are concerned. Moreover, this model permits to check, for example, the hypothesis of log- normality.

3.3.4 Application to the test case

The results for the lower part of the orebody are now presented. For the non informative independent prior, the following marginal prior distributions were defined

- a discrete prior for ν on the interval $(10, 100)$. More weight is given to values in the interval $(20, 40)$, so that the mean is 30 and
- a uniform discrete prior for λ on the interval $(-1, 1)$.

The algorithm used to sample from the parameters posterior distribution was described in the previous example. Figure 3.31 presents the samples of $(\lambda, \boldsymbol{\beta}, \alpha, \nu)$ given $\mathbf{z}$ and the evolution, as a function of the number of iterations, of the Monte Carlo estimates of these parameters. The chains seem stationary. Figure 3.32 presents the correlation between the parameters: as expected, there is a strong dependence between the parameters α and ν. Although Figure 3.25 showed a strong dependence between $\boldsymbol{\beta}$ and λ and between α and λ, here no correlation is highlighted. This is due to the relatively small range over which the transformation parameter λ varies. Figure 3.33 shows the posterior distribution for the model parameters. The posterior mean and variance of λ are 0.13 and 0.002, respectively. The uncertainty on λ is small and the histograms obtained with $\lambda = 0.05$ and $\lambda = 0.18$, for example, are very similar. However, this analysis provides additional insight into the lognormal quest. It reveals that the posterior mean of λ is not 0, but it is closer to 0 than to 1. It is closer to being lognormal than normal. However, no strong evidence of lognormality is observed: the posterior distribution of λ suggests other choices than lognormal. The posterior mean of $\boldsymbol{\beta}$ is $(-1.250, -0.014, -8.24 \times 10^{-5})^T$, of the precision α is 0.31 and of the correlation parameter ν is 31.98. Note that the Monte Carlo estimate of the precision α is much smaller than the value obtained with the gaussian model. It must be noted that a physical interpretation of the estimated parameters is difficult.

The results for the non informative independent prior are now compared to the results for the prior defined through the training sample and for the informative prior. Figure 3.34 presents the posterior distributions obtained with a training sample to define the conditional prior $\pi(\boldsymbol{\beta}, \alpha, \nu|\lambda)$. The predictive measure m^* was defined with $\boldsymbol{\beta} = (-0.5, 0, 0)^T$, $\alpha = 1$ and $\nu = 30$. The prior samples $\mathbf{z}^*$ were simulated on a regular grid with coordinates in the upper part of the deposit. The size of this training set is 30. The posterior mean and variance of λ are 0.21 and 0.005, respectively. The variance is higher than with the non informative independent prior. The posterior mean of $\boldsymbol{\beta}$ is $(-0.850, -0.008, -5.19 \times 10^{-5})^T$, of the precision α is 0.58 and of the correlation parameter ν is 22.49. The posterior distribution of the drift parameters is less dispersed than with non informative independent prior. This is also the case for the correlation parameter ν. On the contrary, the posterior distribution of α has a higher variance.

Figure 3.35 shows the posterior distribution for the model parameters. As we expected these posterior distributions and the corresponding Monte Carlo estimates of the parameters are quite different than the ones obtained with the non informative independent prior. The prior distribution seems to have a much more important role here than in the gaussian case. This is due to the fact that in this case we actually used the upper data to compute the prior distribution. The posterior distributions have thus a smaller variance. The posterior mean and variance of λ are 0.23 and 0.001, respectively. The posterior mean of $\boldsymbol{\beta}$ is $(-0.156, 3.83 \times 10^{-3}, -7.97 \times 10^{-6})^T$, of the precision α is 0.86 and of the correlation parameter ν is 14.10. With the exception of α the posterior variances are smaller with the informative prior distribution

than with the non informative independent prior. The results obtained with the training sample lie in between those of the non informative independent and the informative priors. The precision tends to be higher with an informative prior distribution.

The posterior distribution is sensitive to the choice of the prior distribution. However, here we compared two extreme priors: a vague and an informative prior. If a prior that accounts for the dependence between the parameters needs to be defined, the expected- posterior prior should be used instead of the Box and Cox prior. However, this requires not only the choice of a prior for the training sample but also the spatial location of this training sample. The influence of these choices on the parameters posterior and the predictive distribution must be checked.

We continue to work with the results from the informative prior. To obtain the Bayesian predictive distribution, we sample from $f(Z(x_0)|\mathbf{z}, \lambda, \boldsymbol{\beta}, \alpha, \nu)$ for each $(\lambda, \boldsymbol{\beta}, \alpha, \nu)$ obtained previously. For this we compute the approximate discrete predictive distribution in the support of $Z(x_0)$. Figure 3.36 presents the predictive distributions at the 4 selected locations presented in Figure 3.19. Three predictive distributions are compared: the continuous line represents the Bayesian predictive distribution that is with all parameters unknown, the large dashed line represents the predictive distribution with all parameters unknown but λ fixed at its posterior mean (the simulations were carried out again with $\lambda = 0.23$) and the dotted line represents the *plug-in* distribution with all parameters fixed at their posterior mean. As expected, this last distribution is the less dispersed. The difference between the Bayesian predictive distribution and the *plug-in* distribution is important. This is for taking into account the uncertainty on the transformation parameter. In general, the Bayesian predictive distribution tends to be more dispersed than the distribution with λ fixed, but it is not always the case as in Figure 3.36(b) and (d). This could be due to the fact that the Bayesian predictive distribution is a weighted mean of the *plug- in* distributions.

3.4 Conclusions

The full Bayesian approach allows to mitigate the consequences of misspecification of the correct transformation and hence obtain a more robust inference. However, Bayesian inference for the transformed Gaussian model demands the specification of a joint prior distribution for the model parameters. Because the interpretation of the other parameters depends on the value of the transformation parameter, it makes little sense to use independent priors. However, as we have seen in the Example, the results did not differ particularly whether the Box and Cox prior or an independent prior was used. It is difficult in practice to elicit dependent priors. In particular, the parameters posterior distribution seems sensitive to the choice of the prior. Moreover, the approach may be problematic if we are interested in a physical interpretation of the fitted parameters. In any case, we recommend it as a first tool to analyse non- gaussianity in geostatistical data. For example, it can help deciding whether the log- transformation is appropriate or not for the data under study.

A critical point for the use of the Box- Cox family of transformations in Bayesian analysis is that as the parameter λ changes the range of the transformed data changes.

Two issues for the implementation of the Bayesian model were not considered. One is the selection of the covariance function, and how much this choice influences the predicting performance of the model. The second refers to the selection of the family of transformations.

As we have seen under the gaussian assumptions the resulting prediction methodology has a very close relationship to classical geostatistical kriging methods, but the treatment of unknown parameters departs markedly from the classical geostatistical approach. The classical approach uses curve- fitting methods to match empirical and theoretical variograms, whereas the Bayesian approach emphasises the use of the likelihood function for parameter estimation. The Bayesian approach has the attractive property that uncertainty in the values of model parameters is recognised leading to a more honest assessment of prediction error. It is true that it is difficult to define the priors. This is particularly important when few data are available and the prior's weight has in this case a strong influence in the results.

The application of the Bayesian approach to the case study to obtain the reserves is described in the next chapter. We will consider the anamorphosis function used for conditional simulations to transform the data in normally distributed values. The gaussian model will then be defined for these transformed data. This will permit to compare the Bayesian approach and conditional simulations highlighting the importance of accounting for the uncertainty on the covariance parameters. It is a more general approach than that proposed by De Oliveira *et al.* (1997).

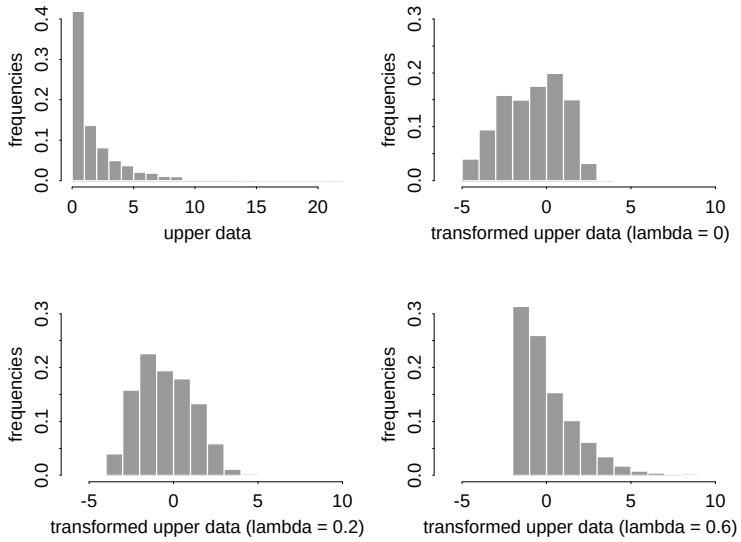


Figure 3.24: *Histogram of the upper data and of these data transformed for 3 different values of* λ*. The mean and variance of these transformed data are: -0.83 and 3.35, respectively, for* $\lambda = 0$*, -0.49 and 2.48, respectively, for* $\lambda = 0.2$*, and finally 0.02 and 2.66, respectively, for* $\lambda = 0.6$

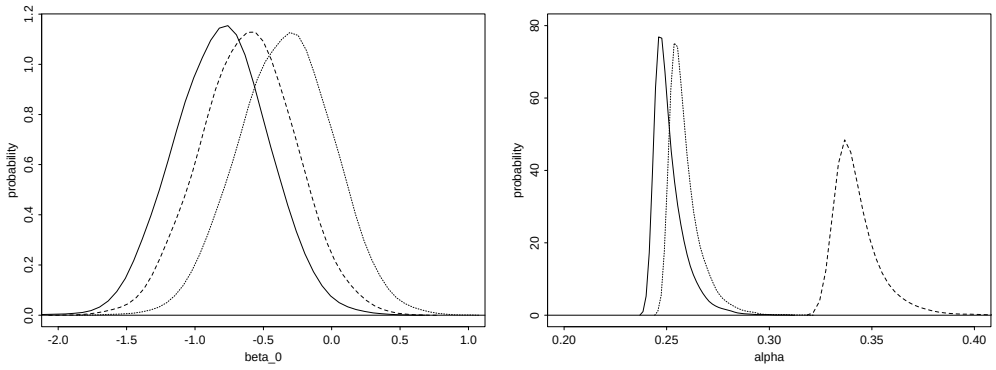


Figure 3.25: *Prior distributions: on the left the prior* $(\beta_0|\mathbf{z}^*, \lambda, \alpha = 1, \nu = 30)$ *is presented and on the right the prior* $(\alpha|\mathbf{z}^*, \lambda, \nu = 30)$ *is presented. Three different values of* λ *are considered. The continuous line is obtained with* $\lambda = 0$*, the dashed line with* $\lambda = 0.2$ *and the dotted line with* $\lambda = 0.6$*. Note that the mean of* β_0 *depends on the transformation parameter, while its variance just depends on the location of the data and not on their value. For* α *both the mean and variance depend on* λ

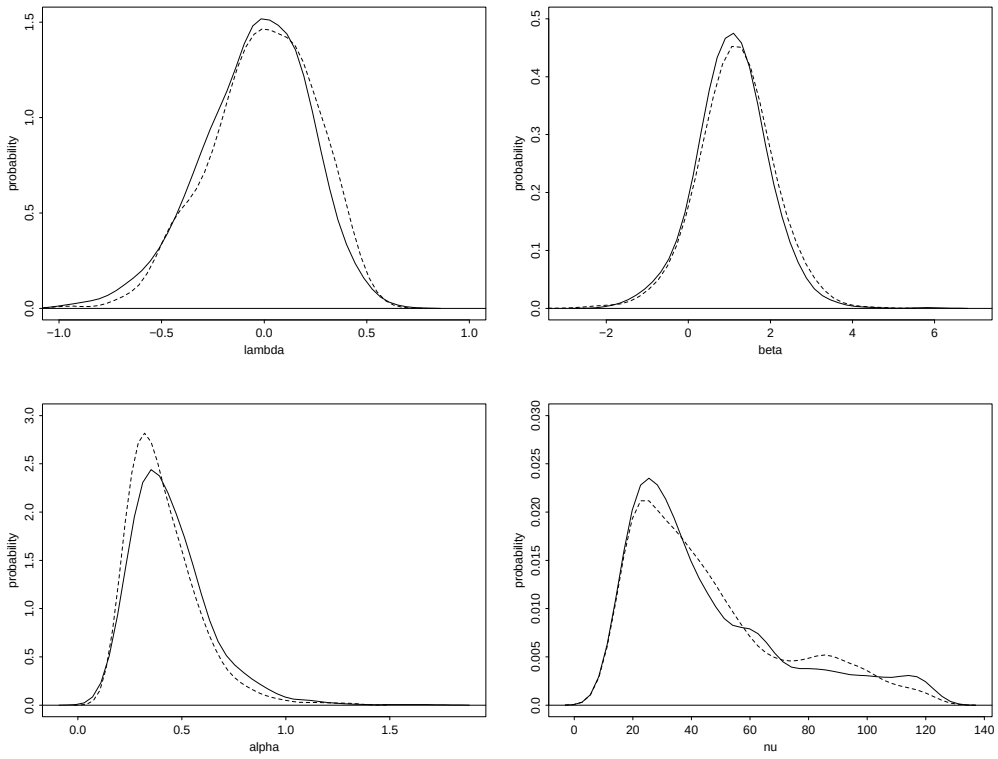


Figure 3.26: *Parameters posterior distributions. The continuous line is obtained with the Box and Cox prior. The dashed line is obtained with the independent prior*

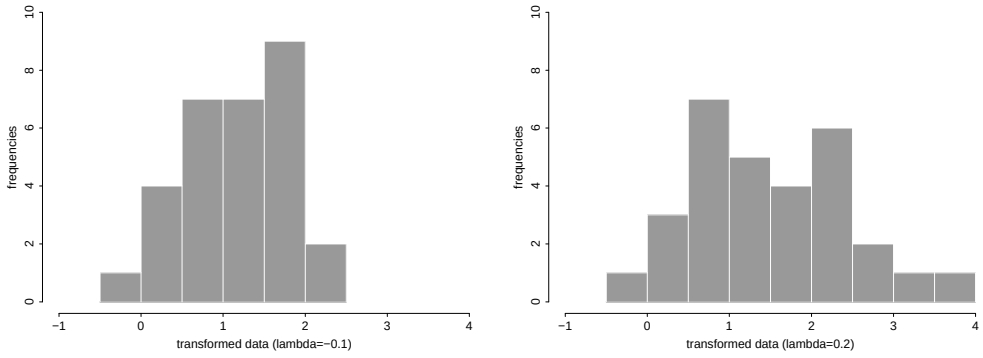


Figure 3.27: *Histogram of the transformed data for two different values of* λ*:* $\lambda = -0.1$ *(on the left) and* $\lambda = 0.2$ *(on the right). The mean and variance of the transformed data with* $\lambda = -0.1$ *are 1.21 and 0.40, respectively. The mean and variance of the transformed data with* $\lambda = 0.2$ *are 1.57 and 0.91, respectively*

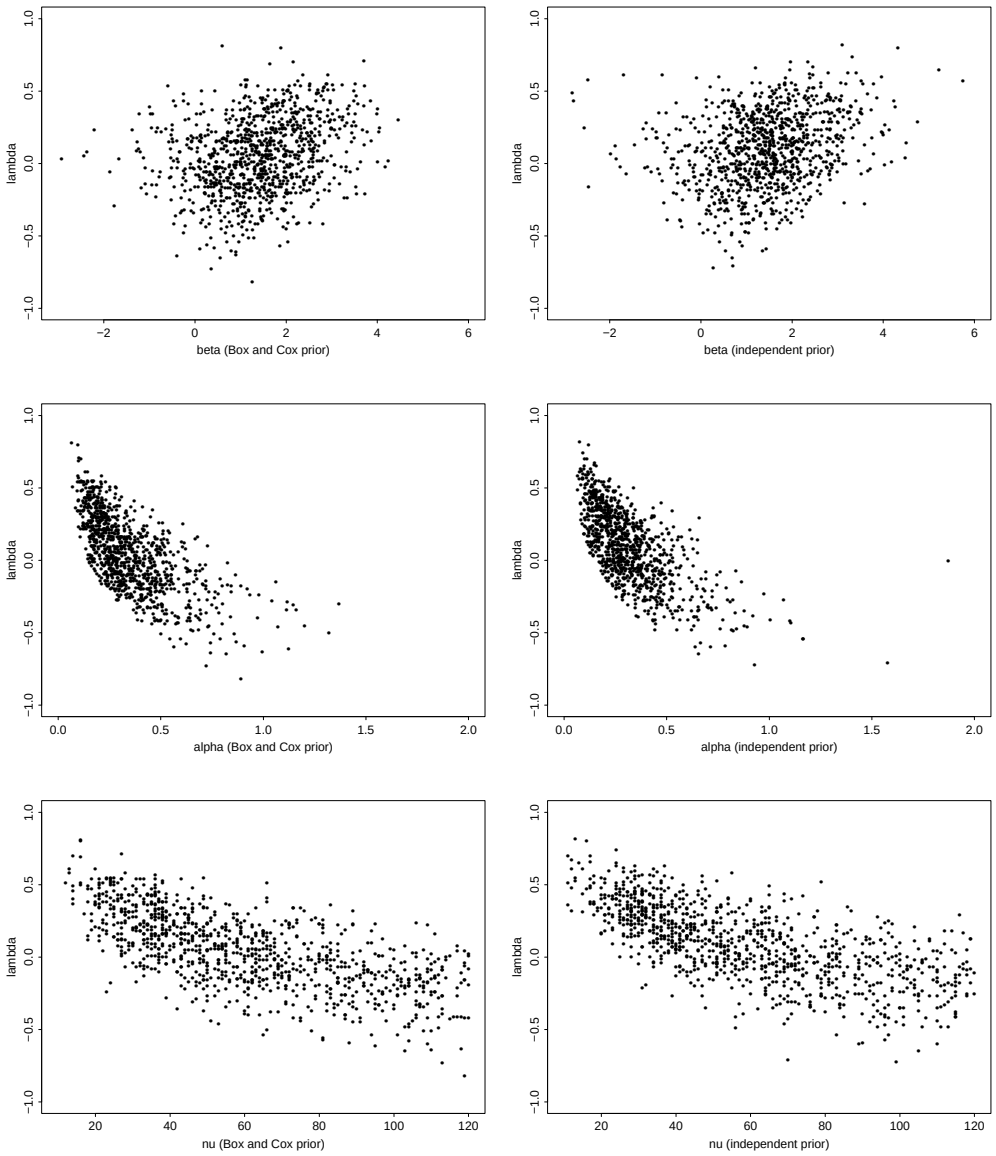


Figure 3.28: *Correlation between the parameters. The results for the Box and Cox prior (on the left) and for the independent prior (on the right) are compared: there is no marked difference. Note the strong correlation between* α *(and thus* ν*) and* λ

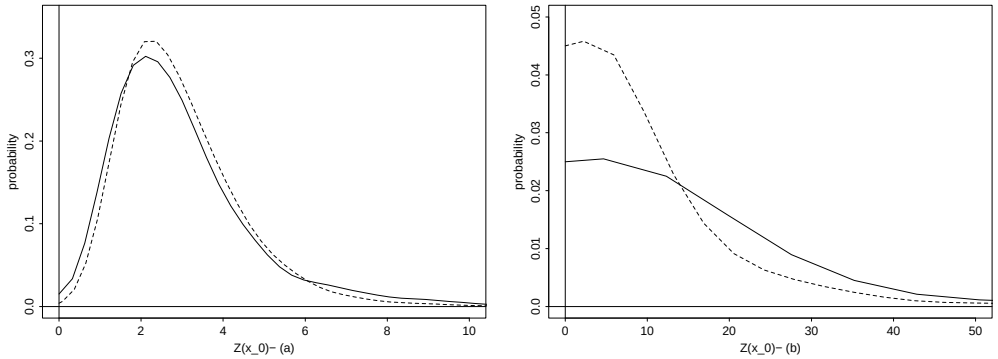


Figure 3.29: *Predictive distributions at (a) $x_0 = 55m$ and (b) $x_0 = 350m$. The effect of taking into account the uncertainty on the transformation and covariance parameters is highlighted. The continuous curve is obtained with all the parameters unknown, the dashed curve is obtained with λ, α and ν fixed at their posterior mean*

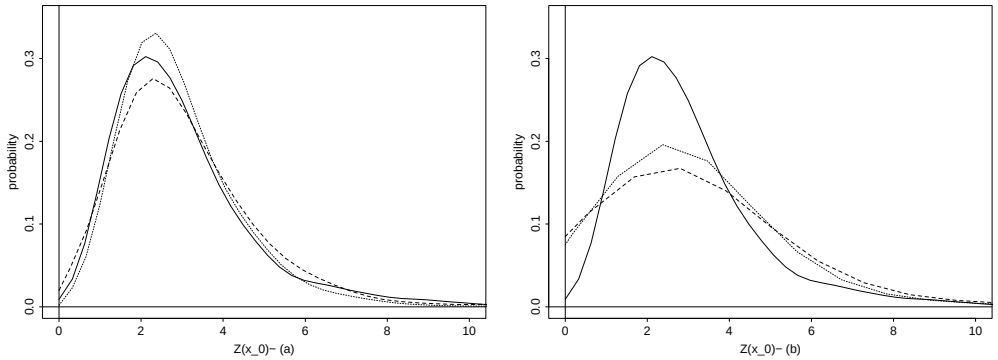


Figure 3.30: *Predictive distributions at $x_0 = 55m$. In (a) the continuous curve is obtained with all the parameters unknown, the dashed curve is obtained with $\lambda = 0$ and β, α and ν unknown, the dotted curve is obtained with $\lambda = 0$ and β, α and ν fixed at their posterior mean. In (b) the continuous curve is obtained with all the parameters unknown, the dashed curve is obtained with $\lambda = -0.5$ and β, α and ν unknown, the dotted curve is obtained with $\lambda = -0.5$ and α and ν fixed at their posterior mean*

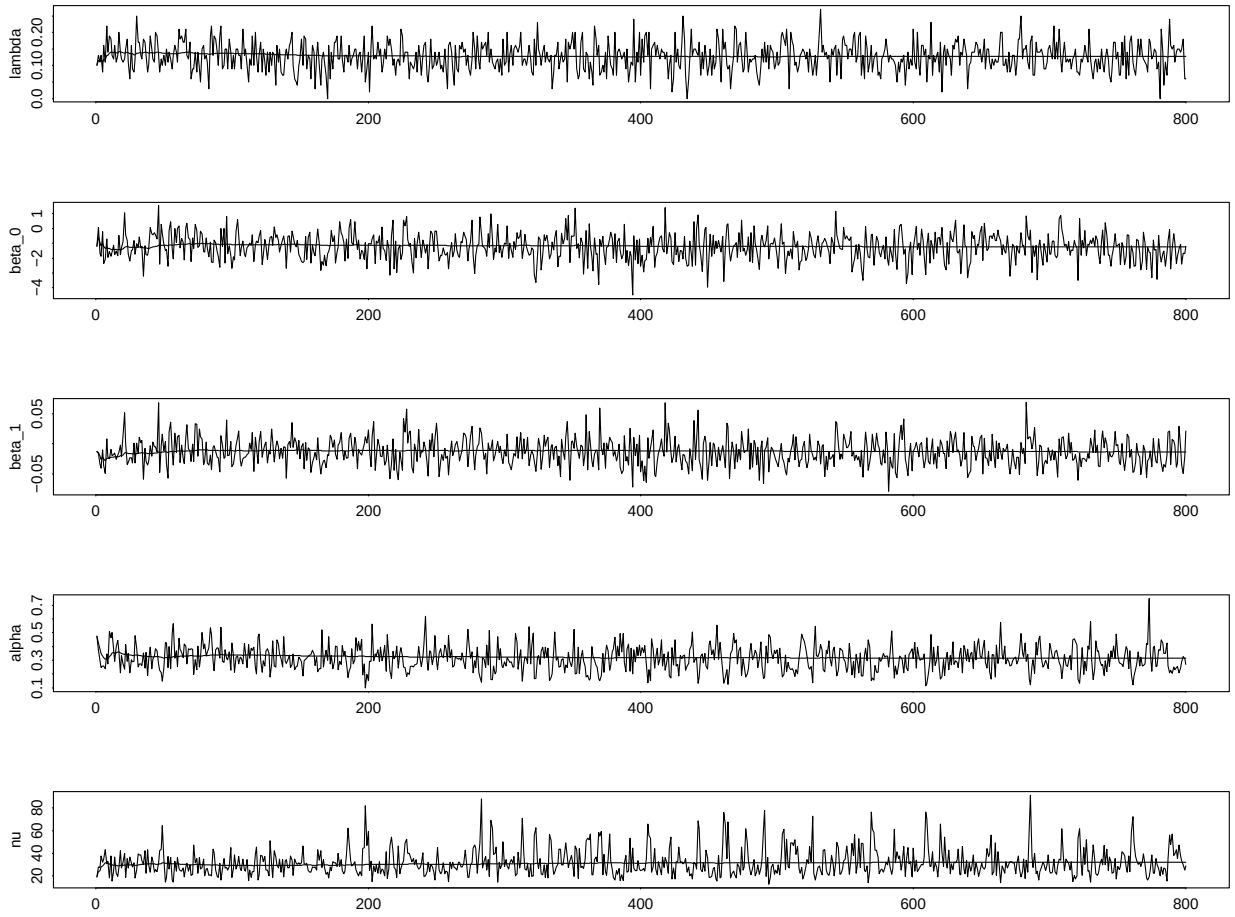


Figure 3.31: *Time series plot of the MCMC output every 50th iteration of the model parameters given* $\mathbf{z}$ *and the evolution of their estimates as a function of the number of iterations. The chains seem stationary. The Monte Carlo estimate of the posterior mean of* α *is 0.31 and of* ν *is 31.98*

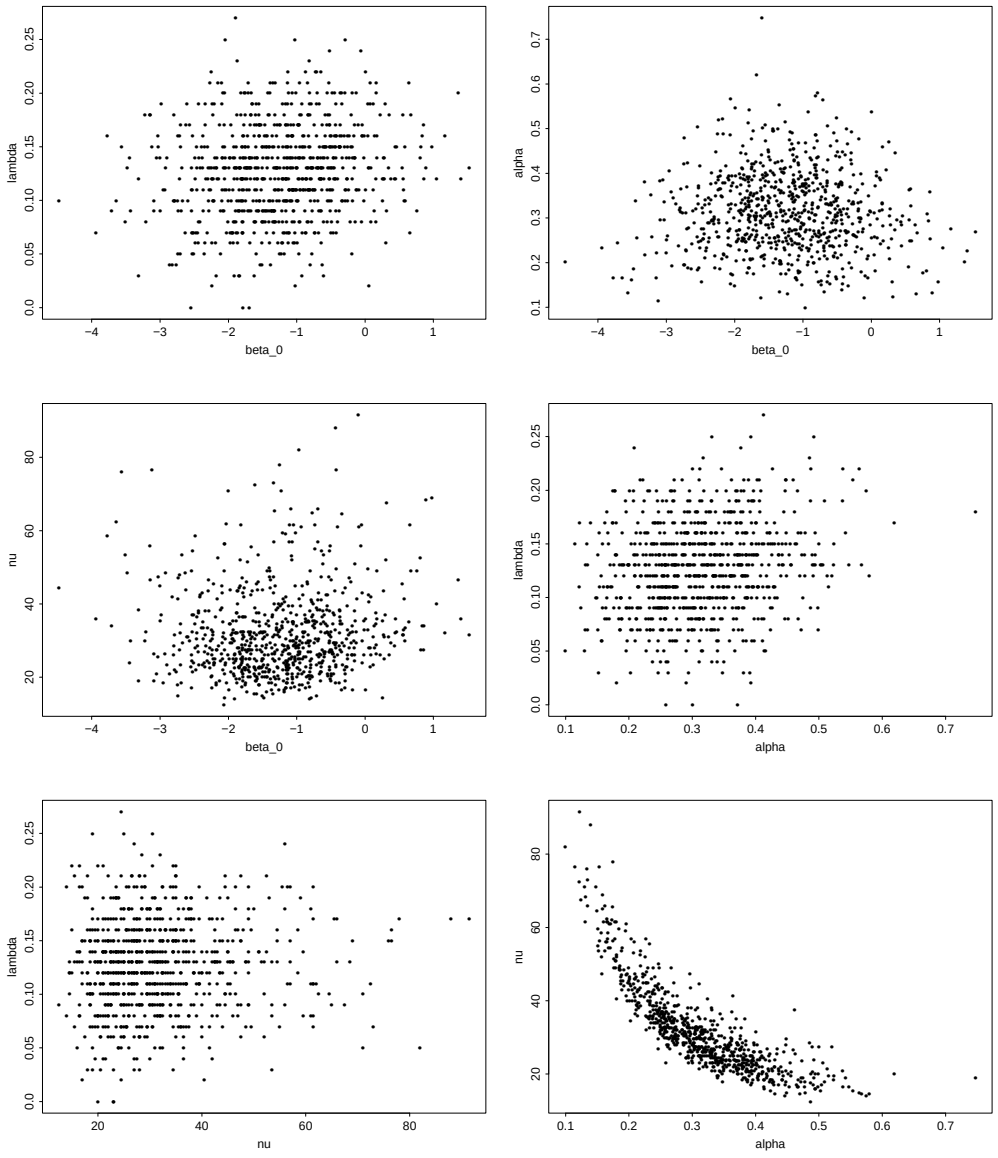


Figure 3.32: *Correlation between the parameters using samples from the posterior distribution. As expected, there is a strong dependence between α and ν. No correlation is highlighted between the transformation parameter λ and the other parameters*

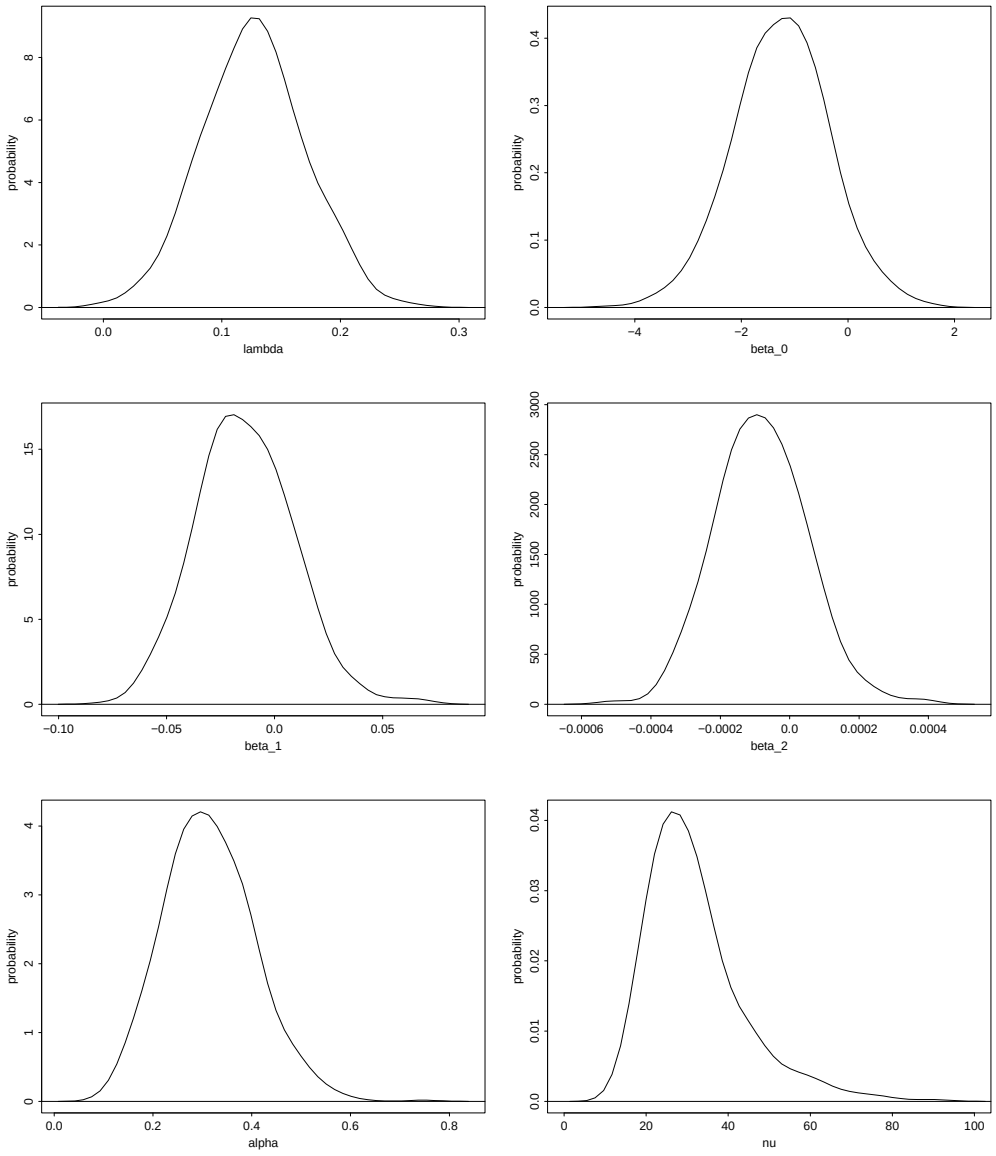


Figure 3.33: *Posterior distribution for the model parameters given the non informative independent prior. The Monte Carlo estimate of the posterior mean of α is 0.31 and of ν is 31.98*

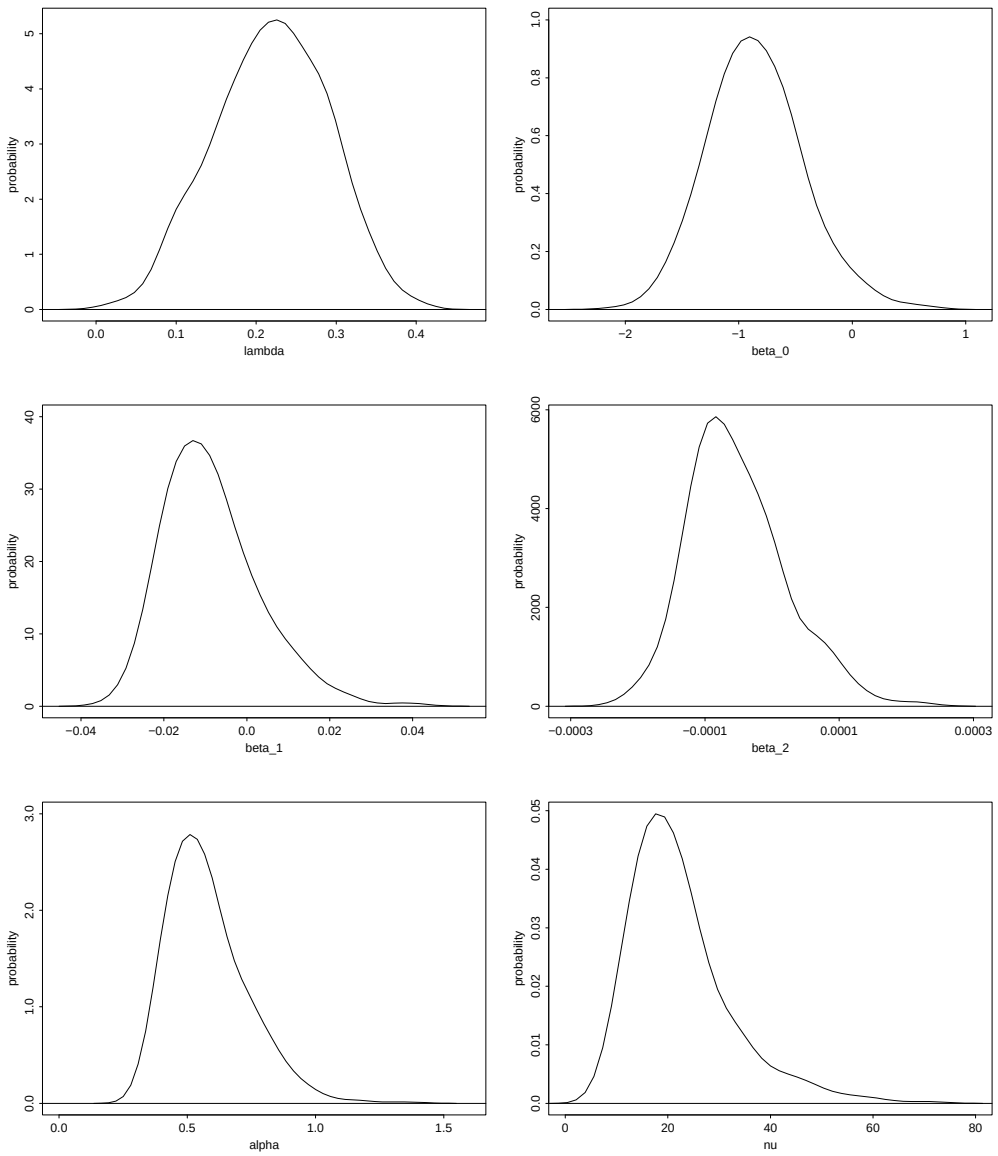


Figure 3.34: *Posterior distribution for the model parameters obtained with a training sample. The Monte Carlo estimate of the posterior mean of α is 0.58 and of ν is 22.49*

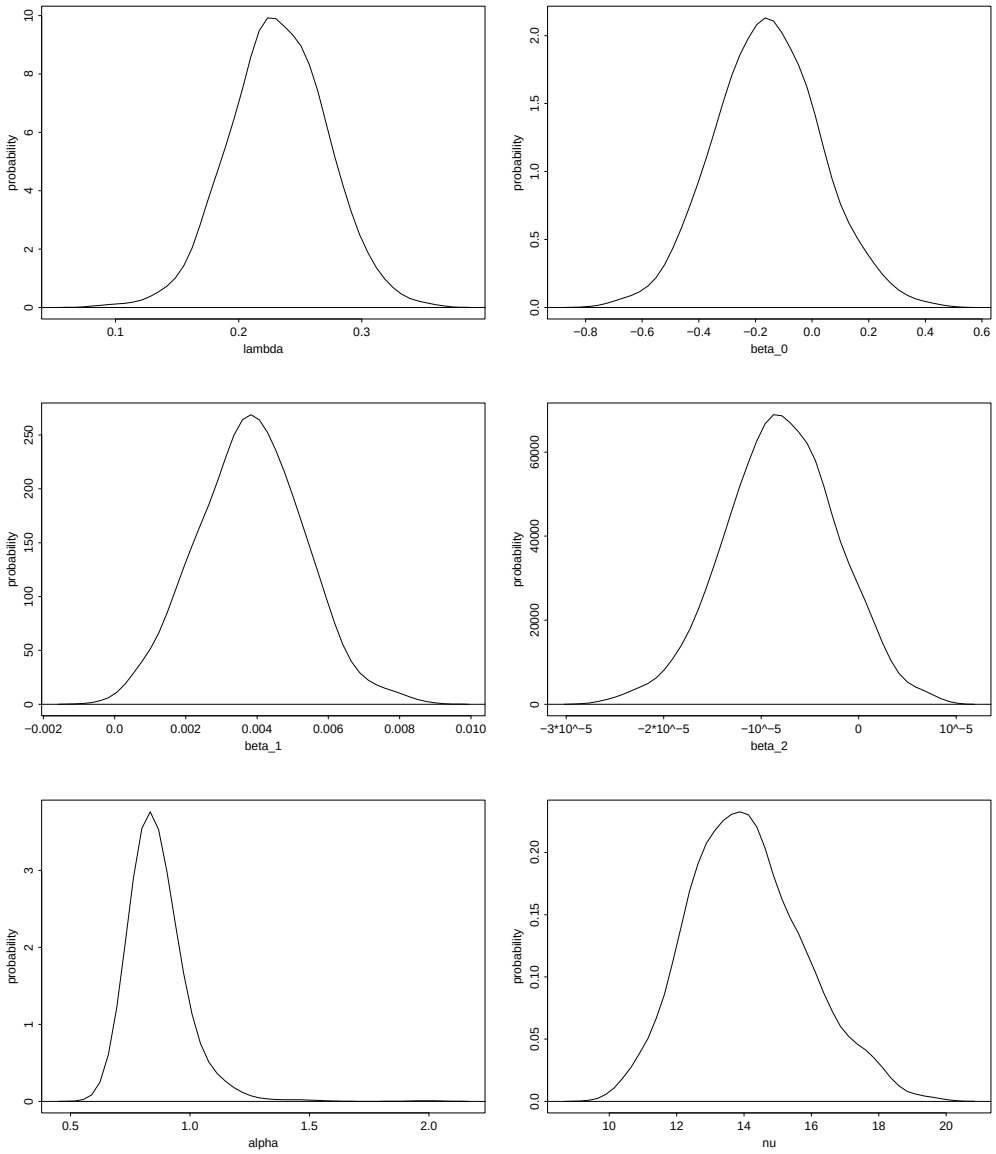


Figure 3.35: *Posterior distribution for the model parameters given the informative prior distribution. The Monte Carlo estimate of the posterior mean of α is 0.86 and of ν is 14.10*

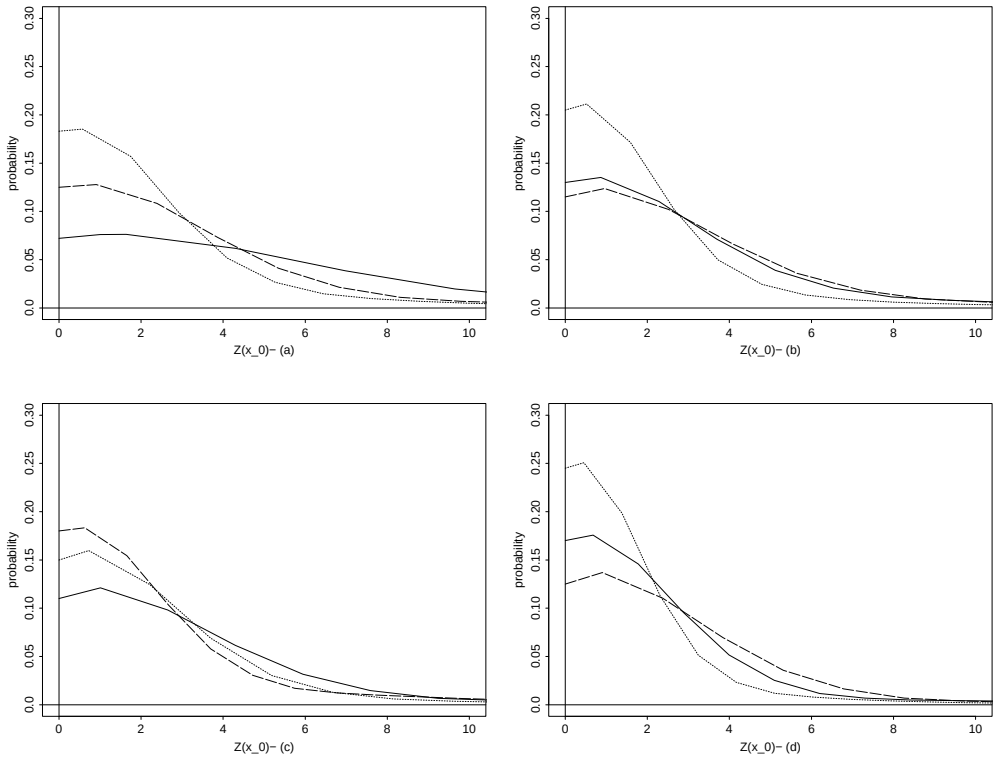


Figure 3.36: *Predictive distributions at (a)* $x_0 = (17100, 23400, -2.5)$*, (b)* $x_0 = (17000, 23600, -75)$*, (c)* $x_0 = (17075, 23150, -50)$ *and (d)* $x_0 = (17125, 23250, -150)$*. The continuous curve is obtained with uncertainty on all parameters taken into account. The large dashed curve is obtained with uncertainty on all parameters taken into account except for* λ*. The dotted curve is obtained with all parameters fixed. The Bayesian predictive distribution is more dispersed in (a) and (c). The difference between the Bayesian and* plug-in *distributions is more important here than in the gaussian case* (Figure 3.20)*. This is due to the uncertainty on the parameter* λ

Chapter 4

Recoverable reserves

This chapter is devoted to the calculation of the reserves and the choice of an exploitation hypothesis to define the quantity of ore and waste exploited each year. Section 1 describes the Bayesian approach that is used to obtain the reserves. Section 2 gives the reserves and compares them to the results obtained with the *plug-in* approach, that is conditional simulations, given in Chapter 2. Section 3 describes the exploitation hypothesis. Section 4 finally presents a first evaluation of the three development options using the valorisation formula that was defined in Chapter 1 and the NPV approach. An example that highlights the value of managerial flexibility is also given.

4.1 The model

The data, after transformation, are assumed to be realizations of the same gaussian random function Y with constant mean m and covariance K. Finally, we decided to transform the data using the same anamorphosis function defined for multiple simulations. Thus for both conditional simulations and the Bayesian approach the multigaussian assumption of Y is made. Once the simulations are carried out, the data are back- transformed.

To simplify the calculations we decided to fix *a priori* 15 covariance models and compute their inverse once and for all. They are assumed equally likely. They are composed of a spherical[1] and two exponential components. The spherical component has a short range and is included as an alternative to the exponential to model short scale variability. It exhibits a linear behavior near the origin. Figure 4.1 highlights the behaviors of the spherical and exponential models near the origin. As the minimum distance between data is of 5m, no information is available to guide us in defining this part of the model. Thus the 15 covariance models differ in the sill of this spherical component of the model and in the range and sill of the exponential components. In particular, taking into account the uncertainty on the parameters of the second exponential

[1]The spherical model has the general form

$$K(h) = \begin{cases} \alpha^{-1} \left(1 - \frac{3}{2}\frac{h}{\nu} + \frac{1}{2}\frac{h^3}{\nu^3}\right) & 0 \leq h \leq \nu \\ 0 & h > \nu \end{cases}$$

The corresponding variogram reaches its sill at $h = \nu$. The nugget- effect model can be considered as a particular case of the spherical covariance function with an infi nitely small range. There is an important difference: the nugget effect describes a discontinuous phenomenon while the spherical covariance represents a phenomenon that is continuous, but not differentiable

component is interesting as they may describe the relationship between the upper and lower data. These models are presented in Table 4.1. The model defined in Table 2.5 is also included (model 14 in Table 4.1). As we have seen in the previous chapter the uncertainty on the variance parameter does not have a large influence on the predictive distribution. Thus the total variance is fixed at 1 for all the models. It is true that certain of these models are very similar: models 1 and 2 or models 8 and 9, for example. The models chosen fit graphically the experimental variogram (Figure 2.9). Although it is not a rigorous Bayesian approach, this simplified approach still permits to let some flexibility in the definition of the variogram model.

The prior distribution for m is taken noninformative. No prior information was available to be introduced in the estimation process. The Bayesian approach is considered to allow some uncertainty on the covariance parameters. Let K_i denote the ith model for the covariance. To be more general and account for different short scale behaviours we could have used the Matérn class of correlation functions.

As we have seen in the previous chapter with a noninformative prior for m we can get an analytical expression for the posterior probability for model K_i, $\pi(K_i|\mathbf{y})$. The algorithm for simulations is then

- simulate from $\pi(K_i|\mathbf{y}) \propto f(\mathbf{y}|K_i)\,\pi(K_i)$;
- simulate from $\pi(m|\mathbf{y}, K_i) \propto f(\mathbf{y}|m, K_i)\,\pi(m)$;
- simulate from $(Y_0|m, K_i, \mathbf{y}) \sim \mathcal{N}(M, \Sigma)$, where M, that was given in (3.17), is a function of $\mathbf{y}$ and of both m and K_i and Σ, that was given in (3.18), is a function of K_i.

Figure 4.2 presents the histogram of the simulated covariance models. Note that although the maximum a posteriori is model 5, models 1, 2, 6 and 14 have a high posterior probability and are very similar. Figure 4.3 presents the simulated values of the mean m and the evolution of its Monte Carlo estimate. The Monte Carlo estimate of the mean is -0.34.

Given the covariance models and the values of m we need to sample from $(Y_0|m, K_i, \mathbf{y})$. Given the dimension of $\mathbf{y}$, to carry out the simulations of Y_0 we proceed exactly as for conditional simulations that is a neighborhood is defined. The same grid defined for kriging and conditional simulations is considered. Taking into account the uncertainty on the mean and covariance parameters of the anamorphosed grades, thus not imposing they are values from a standard gaussian distribution which is the usual assumption, is a way to introduce some uncertainty on the anamorphosis function that is important mostly when a zone with few data is simulated.

The data are finally back- transformed to obtain the realisations of Z. That is classic conditional simulations are carried out. Here for each conditional simulation a different mean and covariance model is defined (simple kriging is used to condition instead of ordinary kriging). This model is more general than the gaussian transformed model by De Oliveira *et al.* (1997).

Figure 4.4 compares the *plug- in* and Bayesian predictions and prediction variances for the small pit. The Bayesian predictions do not differ too much from the plug- in predictions. However, there are differences in the prediction variances. Note that the plug- in variances tend to slightly over- estimate the variance of the predictive distribution. As we have seen in the previous chapter, the larger the nugget effect (here the spherical component) the larger the prediction variance (of course only for non- isolated points). The Bayesian predictive distribution is a weighted mean of distributions with a nugget effect and without a nugget effect.

	Sill	Range 1 (m)	Range 2 (m)	Range 3 (m)
Spherical	0.1	5	5	5
Exponential 1	0.7	95	95	60
Exponential 2	0.2	350	350	100
Spherical	0.1	5	5	5
Exponential 1	0.7	110	110	50
Exponential 2	0.2	350	350	100
Spherical	0.1	5	5	5
Exponential 1	0.8	110	110	60
Exponential 2	0.1	350	350	100
Spherical	0.1	5	5	5
Exponential 1	0.9	150	150	60
Exponential 1	0.9	120	120	50
Exponential 2	0.1	350	350	100
Spherical	0.2	5	5	5
Exponential 1	0.7	120	120	50
Exponential 2	0.1	350	350	100
Spherical	0.2	5	5	5
Exponential 1	0.6	120	120	60
Exponential 2	0.2	350	350	100
Spherical	0.1	5	5	5
Exponential 1	0.6	85	85	60
Exponential 2	0.3	350	350	100
Spherical	0.1	5	5	5
Exponential 1	0.6	80	80	50
Exponential 2	0.3	350	350	100
Exponential 1	0.8	80	80	50
Exponential 2	0.2	350	350	100
Spherical	0.3	5	5	5
Exponential 1	0.5	150	150	80
Exponential 2	0.2	350	350	100
Exponential 1	0.7	80	80	50
Exponential 2	0.3	350	350	100
Spherical	0.3	5	5	5
Exponential 1	0.6	150	150	80
Exponential 2	0.1	350	350	100
Spherical	0.1	5	5	5
Exponential 1	0.7	100	100	60
Exponential 2	0.2	350	350	100
Exponential 1	0.6	70	70	70
Exponential 2	0.4	350	350	100

Table 4.1: *Variogram parameters of the 15 fitted models (the main directions of the orebody are obtained through a rotation of (-165, 45, 150) that is indicated in the geological convention, relative to east)*

Additional information

The model parameters are updated when the additional information $\mathbf{z}_I$, that is the rich, average or poor data set, is available. Using Bayes' Theorem we have

$$f(\mathbf{z}_0|\mathbf{z}, \mathbf{z}_I, \boldsymbol{\theta}) \propto f(\mathbf{z}_0|\mathbf{z}, \boldsymbol{\theta})\, f(\mathbf{z}_I|\mathbf{z}_0, \mathbf{z}, \boldsymbol{\theta}).$$

This means that $f(\mathbf{z}_0|\mathbf{z},\boldsymbol{\theta})$ can be seen as a prior for $\mathbf{z}_0$ and when $\mathbf{z}_I$ is available it is updated. Thus to generate samples from $f(\mathbf{z}_0|\mathbf{z},\mathbf{z}_I,\boldsymbol{\theta})$ we could use $f(\mathbf{z}_0|\mathbf{z},\boldsymbol{\theta})$ as the transition probability in the Metropolis-Hastings algorithm and accept $\mathbf{z}_0'$ with probability

$$a(\mathbf{z}_0,\mathbf{z}_0') = \min\left\{1, \frac{f(\mathbf{z}_I|\mathbf{z}_0',\mathbf{z},\boldsymbol{\theta})}{f(\mathbf{z}_I|\mathbf{z}_0,\mathbf{z},\boldsymbol{\theta})}\right\}.$$

This has the advantage of not having to carry out the simulations given the observations $(\mathbf{z},\mathbf{z}_I)$. However the acceptance rate may be very small if the two distributions have a quite different support. This is our case, as the additional data, rich and poor, modify the beliefs about the grade distribution. For this reason when the data set $\mathbf{z}_I$ is considered the computations are repeated given $(\mathbf{z},\mathbf{z}_I)$.

4.2 Recoverable reserves: Bayesian versus *plug-in*

Table 4.2 presents the number of blocks with average grade above cutoff. We observe that

- for the large pit and the lower part of large pit: the expected number of blocks above cutoff is slightly higher given the rich and average additional information. This can also be seen in Figure 4.5. The standard deviation of the number of blocks above cutoff is smaller given the rich additional information than given just the initial data: the rich extra information helps better defining the project.

- for the small pit: given the number of simulated blocks we can say that the expected number of blocks above cutoff is more or less the same given the initial data or given the initial data + extra information. In particular, note that the standard deviation is much smaller given the rich data set.

These results and the ones obtained in Table 2.8 with the *plug-in* approach can be compared. The expected tonnages obtained with the Bayesian approach or the *plug-in* approach are very close, while the dispersions differ greatly. For the lower part of the deposit the results obtained with the Bayesian approach are less dispersed (except for the average additional information). This can also be seen in Figure 4.5. On the contrary for the whole large pit and the small pit they are more dispersed.

Table 4.3 presents the recovered average grade. We note that

- for the large pit and lower part of large pit: the uncertainty on the recovered average grade is larger given the rich additional information than given the initial data. Whereas it is smaller given the average and poor extra data. The additional information has a large impact on the expected recovered average grade. This is also shown in Figure 4.6.

- for the small pit: the expected recovered average grade and its standard deviation are more or less the same given the initial or the initial + extra information.

These results and the ones obtained in Table 2.9 can be compared. As it was noted earlier the Bayesian and the *plug-in* predictive distributions differ particularly in the tails and this is highlighted taking the values above a cutoff. Given the initial data the results for the recovered average grade for the large and mostly for the small pit are very similar. It is reassuring as the upper part of the deposit is densely sampled and the uncertainty is small. The Bayesian approach gives slightly less dispersed values. The results for the lower part of the large pit obtained with the Bayesian approach are also less dispersed. Looking at Figure 4.6 we can see that the *plug-in* approach gives a larger weight to the high values.

Initial 90 drill-holes							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	17782	24118	**21397**	1328	19843	21378	22896
Lower part	3678	5993	**4743**	554	4001	4633	5601
Small pit	13645	18155	**16117**	1009	14792	16157	17436

Initial 90 drill-holes + rich fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	18921	24753	**21442**	1077	20107	21506	22781
Lower part	4186	5767	**4945**	390	4464	4951	5547
Small pit	14295	18842	**15942**	844	14882	15843	16891

Initial 90 drill-holes + average fictive data							
	Min	Max	Mean	Std dev	10 %	50 %	90 %
Large pit	18621	24673	**21504**	1456	19736	21391	23239
Lower part	3670	7001	**4944**	587	4264	4898	5663
Small pit	13981	18215	**16049**	1094	14614	15941	17645

Initial 90 drill-holes + poor fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	18319	24033	**21232**	1319	19665	21061	22905
Lower part	3624	5611	**4737**	498	3974	4784	5373
Small pit	13814	18332	**15940**	1039	14619	15859	17148

Table 4.2: *Simulations results: tonnage for the whole large pit, lower part of large pit and small pit. The size of the blocks is 10m $\times$ 25m $\times$ 10m*

Finally the recoverable reserves can be computed for both approaches. They are presented in Figures 4.7 and 4.8 for the small and large pit, respectively. As expected, the results for the small pit are more or less the same whether they are obtained given the initial data or the initial + extra data. On the contrary, the results for the large pit are quite different whether the additional information is obtained or not. Moreover, the distributions, and in particular their tails, differ whether the Bayesian or the plug- in approach is used.

Consequently, and as we will see in Section 5.4, the evaluation of the large pit will differ whether the Bayesian approach or conditional simulations is used to calculate the reserves. In particular, the difference is not expected to be important for the poor data set as the reserves distributions are close as it can be seen in Figure 4.8.

Thus although enough data seemed available to choose the variogram parameters, the Bayesian approach still permits to account for their misspecification.

Initial 90 drill-holes							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.84	3.43	**3.08**	0.12	2.94	3.06	3.26
Lower part	2.86	4.19	**3.41**	0.30	3.01	3.38	3.80
Small pit	2.79	3.30	**3.00**	0.11	2.85	2.99	3.12

Initial 90 drill-holes + rich fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.97	3.79	**3.24**	0.16	3.05	3.24	3.40
Lower part	3.35	5.73	**3.92**	0.41	3.49	3.88	4.31
Small pit	2.85	3.39	**3.06**	0.13	2.90	3.03	3.23

Initial 90 drill-holes + average fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.88	3.44	**3.12**	0.12	2.97	3.12	3.28
Lower part	3.15	4.12	**3.50**	0.27	3.21	3.43	3.85
Small pit	2.74	3.32	**3.05**	0.12	2.93	3.04	3.17

Initial 90 drill-holes + poor fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.75	3.35	**3.04**	0.13	2.88	3.04	3.19
Lower part	2.90	3.85	**3.27**	0.22	3.03	3.24	3.56
Small pit	2.74	3.35	**2.99**	0.14	2.85	2.96	3.19

Table 4.3: *Simulations results: recovered average grade above a cutoff of 1g/t for the whole large pit, lower part of large pit and small pit. The size of the blocks is 10m $\times$ 25m $\times$ 10m. The recovered average grade has a higher variance given the additional information than given the initial data (the coefficients of variation for the lower part of large pit are: 0.088, 0.105, 0.077, 0.067 given the initial data, the rich, average and poor information, respectively)*

4.2.1 A drift?

To let the model be more flexible a drift could be defined instead. This assumption will have a larger impact on the lower part of the deposit as it is sparsely sampled. As we have seen in Figure 2.3 the upper part of the deposit has a higher mean grade than the lower part. A drift, that is a function of the depth, that takes

this into account could be considered. Also given the rich, average and poor extra data, the introduction of a drift that looks at the different behaviour of the lower part with respect to the upper part (Figure 2.11) may be considered. We thought of a quadratic drift, in the place of a simple linear drift, as the top of the upper zone of the deposit has a low mean grade. Figure 4.9 presents the estimated quadratic drift[2]. The vertical line is the estimated constant mean. The presence of a drift seems plausible as the quadratic term is clearly non zero. Additional information may help better defining the drift. Figure 4.10 presents the estimated drift given the three sets for additional information. Note that the lowest part of the deposit is considered rich and thus interesting only given the rich extra data.

Figure 4.11 shows the interest for the Bayesian approach. The two drifts presented in the figure are estimated with a different covariance model. Both models could have been chosen graphically.

Note that in the model it is the drift for the gaussian data that must be defined. The average grade of the gaussian data for each depth and the estimated quadratic drift are presented in Figure 4.12.

Figure 4.13 presents the simulated values of the mean $\boldsymbol{\beta}$ and the evolution of the Monte Carlo estimates. The histograms of these samples are presented in Figure 4.14. Note that the posterior distributions support the presence of a drift. To get a sample of 50 the last values are kept. The Monte Carlo estimate of $\boldsymbol{\beta}$ is $(-0.32, 8.28 \times 10^{-4}, -2.55 \times 10^{-6})^T$. The variance is 0.12, 5.12×10^{-7} and 3.75×10^{-12} for β_0, β_1 and β_2, respectively.

Gaussian conditional simulations were then carried out. Finally the values were back- transformed. Tables 4.4 and 4.5 present the number of blocks with average grade above cutoff and the recovered average grade, respectively. It can be noted that the results for the number of blocks above cutoff differ greatly from the results obtained with a constant mean. If the drift was assumed for the original data this could be due to the assumed drift function: from Figure 4.9 we observed that in the top and lower parts of the deposit the drift is much smaller than the constant mean. Although in the central part the drift is much higher than the constant mean, the constant mean is around 1g/t and thus close to the cutoff. Thus the expected number of blocks with average grade above cutoff would be expected to be smaller for the results obtained with the drift. Figures 4.15 and 4.16 compare two simulations obtained either assuming a constant mean or a quadratic drift for two different depths in the top and lower parts of the deposit. It can be noted that in these two zones the number of blocks with grade above cutoff is much smaller for the simulations obtained assuming the presence of a drift. This is also shown in Figures 4.17 and 4.18. The results for the recovered average grade, although those obtained with the drift are more variable, can be compared. This is because in the central zone of the deposit for blocks above cutoff the predicted grade is much higher when the drift is allowed for.

As we have seen it is extremely important to check the stationarity hypothesis and define correctly the drift function. Remark that the quadratic drift highlights the limits of the deposits. However, it gives a very rich central zone. The most delicate zone is 100m down as it is scarcely sampled. In particular the zone from 0m to 100m is poor and sparsely sampled. Although the Bayesian results support the presence of a drift, we decided to keep a constant mean as it is the simplest model. If *a priori* a constant mean seemed

[2]The drift coeffi cients $\boldsymbol{\beta}$ are estimated through minimum least squares, that is

$$\hat{\boldsymbol{\beta}} = (F^T K^{-1} F)^{-1} (F^T K^{-1} \mathbf{z}),$$

where F is the matrix of the known covariates and K is the covariance matrix

realistic, a prior distribution with a large weight for β_0 could have been defined. The hypothesis of a drift or a constant mean plays an important role in the zones with few data. For this decision the opinion of an expert, as the geologist, for example, may be crucial.

Initial 90 drill-holes							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	3424	21885	**11365**	4842	4727	11222	17786
Lower part	1364	6105	**3538**	1266	1965	3614	5166
Small pit	1875	15224	**7540**	3584	2555	7161	12872

Initial 90 drill-holes + rich fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	1802	29679	**11018**	5190	6346	10063	15983
Lower part	436	6085	**2593**	1127	1571	2384	4084
Small pit	1303	22708	**8165**	4027	4751	7413	11690

Initial 90 drill-holes + average fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	3466	23303	**11474**	4255	6300	10952	16612
Lower part	954	5393	**2620**	942	1677	2556	3697
Small pit	2129	17491	**8605**	3328	4416	8081	13014

Initial 90 drill-holes + poor fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	4304	21236	**11350**	3855	6514	10757	15612
Lower part	615	4334	**2314**	917	1130	2237	3607
Small pit	3290	16321	**8803**	3044	4935	8788	12595

Table 4.4: *Simulations results with the quadratic vertical drift: tonnage for the whole large pit, lower part of large pit and small pit. The size of the blocks is 10m $\times$ 25m $\times$ 10m*

4.3 Exploitation hypothesis

Different exploitation hypotheses could have been considered. We assumed a constant tonnage, that is a constant ore tonnage and a constant waste tonnage, and average grade per year. The lifetime of the deposit is supposed fixed. Thus once the simulations are carried out, the blocks with a recovered average grade above cutoff are selected and the ore tonnage is computed. The annual ore tonnage is then given by the total

Initial 90 drill-holes							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.66	3.78	**3.06**	0.25	2.78	3.03	3.42
Lower part	3.11	4.80	**3.73**	0.39	3.31	3.63	4.22
Small pit	2.30	3.33	**2.85**	0.21	2.62	2.85	3.09

Initial 90 drill-holes + rich fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.59	4.53	**3.31**	0.39	2.85	3.30	3.78
Lower part	3.00	5.39	**3.84**	0.48	3.38	3.78	4.41
Small pit	2.37	4.35	**3.21**	0.40	2.67	3.26	3.66

Initial 90 drill-holes + average fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.54	3.63	**2.98**	0.21	2.77	2.94	3.25
Lower part	2.72	4.08	**3.15**	0.26	2.89	3.12	3.46
Small pit	2.49	3.51	**2.92**	0.21	2.70	2.87	3.22

Initial 90 drill-holes + poor fictive data							
	Min	Max	**Mean**	Std dev	10 %	50 %	90 %
Large pit	2.64	3.42	**2.90**	0.16	2.71	2.88	3.09
Lower part	2.81	3.74	**3.13**	0.19	2.93	3.07	3.36
Small pit	2.59	3.45	**2.90**	0.17	2.71	2.87	3.15

Table 4.5: *Simulations results with the quadratic vertical drift: recovered average grade above a cutoff of 1g/t for the whole large pit, lower part of large pit and small pit. The size of the blocks is 10m* $\times$ *25m* $\times$ *10m*

ore tonnage divided by the number of years. The annual average grade is the average grade above cutoff. This implies that blocks above cutoff can be mined irrespectively of where they are located. This is far from what is possible in reality. It is perhaps too simplistic given the size of the deposit and the possible presence of a vertical drift. But it allows to calculate approximate ore and waste tonnages and grades quickly avoiding having to schedule the mining sequence.

If the vertical drift was considered, a more realistic exploitation hypothesis would have to take it into account. A constant tonnage per year as in the first hypothesis and thus a fixed number of years would be defined. But the average grade per year would not be constant. It would vary as the hypothesis of free selection is not considered and according to the drift. This would have to be done for each simulation.

Whether the drift function is considered or not the results may be very different for the project evaluation. If we defined a vertical drift, the lowest part of the deposit would not be exploited (Figure 4.9) reducing the costs and duration of the project.

4.4 Development options

For the moment consider the three following development options

1. develop the small pit immediately, with a sure profit;
2. develop the large pit immediately, with the chance of a larger profit;
3. first carry out additional drilling which would reduce the uncertainty on the grades in the lower part of the orebody, but would be costly and would delay the start of the project. Then choose whether to develop the small or large pit.

The objective is to choose the optimal development option, that is develop the deposit immediately or after having obtained additional drilling, and the optimal project, that is develop the small or large pit.

The value of the two projects: a first approximation

Suppose that the only source of uncertainty in the evaluation of this mining project comes from the unknown reserves. With the notation introduced in Chapter 1, let U_1 be *develop the large pit* and U_2 be *develop the small pit*. The optimal project maximizes the expected profit $E[V_U]$, that is the conventional income net of fixed costs. It is given by

$$\arg\max_U E[V_U].$$

In this case we now have all the information necessary to evaluate the 3 development options and choose the optimal project. Let

- $S = \$\,320$ per ounce troy of gold (that is $\$\,320/31.103$ per g of gold) be the expected gold price;
- $FC = \$\,100$ million be the fixed costs;
- $VC_{U1} = \$\,5.5$ and $VC_{U2} = \$\,4.5$ be the cost per ton of waste for project U_1 and for project U_2, respectively;
- $IC = \$\,10$ million be the cost of additional drilling.

A cost is associated to the extraction of waste. This cost is higher for U_1 as it is harder to extract the waste in the lower part of the deposit. Given the recovered average grade and number of blocks with grade above cutoff the reserves B and quantity of waste W are easily computed[3] (to obtain the respective tonnages just multiply by the volume of the block, 2500 m^3, and by the ore density, 2 t/m^3). The Monte Carlo approximation of the expected profit of project U is given by

$$E[V_U] \approx \frac{1}{L}\sum_{l=1}^{L}(SB_l - VC_U W_l) - FC - IC$$

[3] In W the waste that has to be removed to access the mineralised zone (see Figure 2.1) is included. It was estimated by ordinary kriging using all the data

where L is the number of Monte Carlo simulations that are carried out (that is $L = 50$). Table 4.6 presents these Monte Carlo estimates obtained with the Bayesian approach for the two projects. The optimal project is U_1, that is develop the large pit. However, the two project values are very close and the risk of developing the large pit is high: the uncertainty on the reserves is larger for the large pit than for the small pit. As it can be seen in Figure 4.19, the probability of both a negative or small profit (< \$ 200 million) and of a high profit (> \$ 800 million) is higher for U_1 than for U_2. Figure 4.20 compares the 2 project values for each of the 50 simulations: U_1 is preferred to U_2 in 50% of the simulations. Observe that the expected value of the positive differences is more or less the same as that for the negative differences. This can be seen in Figure 4.21 which also compares the 2 project values.

Available data	**Project value** Large pit (U_1)	Small pit (U_2)
Bayesian approach		
Initial data	**448**	431
Rich new data set	**612**	447
Average new data set	**494**	455
Poor new data set	363	**397**
plug-in approach		
Initial data	**428**	404
Rich new data set	**657**	476
Average new data set	**515**	416
Poor new data set	377	**395**

Table 4.6: *Value of the two projects in \$ million. The results for both the Bayesian and* plug-in *approach are presented. The optimal projects given the initial and initial + additional data are highlighted*

Given the supplementary observations the distance between the two project values is larger, that is the additional data reduce the risk of taking the wrong decision. But is it enough to justify the cost of additional data? The value of additional information is

new value of project - old value of project.

As the cost of getting more information is incorporated in the new value, it is worthwhile obtaining it if this difference is positive. The project value with average additional information is slightly higher than with just the initial data. Figure 4.19 compares the two project values given the average additional information. Project U_1 is still optimal. Note that the average additional information does not help choosing the optimal project: the two histograms are more or less the same as the ones obtained given just the initial data. Figure 4.19 also compares the two project values given the poor additional information. The optimal project shifts to U_2, that is develop immediately the small pit. In this case the additional information permits to avoid incurring in a loss as the expected project value is not as great as expected. The optimal project remains U_1 given the rich scenario for additional data. From Figure 4.21 observe that the rich additional information best discriminates between the 2 projects: given this information the development of the large pit is strictly preferred to the development of the small pit. The average and poor extra data do not help choosing the optimal project.

We can compute the expected project value integrating with respect to the additional information although

it is an approximation as we chose and fixed 3 scenarios. Suppose that 0.25, 0.50 and 0.25 are the probabilities of the rich, average and poor scenarios, respectively. We assume they represent 25 %, 50 % and 25 %, respectively, of the possible extra data sets. The expected value of additional information is then $ 51.25 million. It augments the value of the project by 11.44 %. Thus it is worthwhile to obtain it.

Table 4.6 also presents the results obtained with conditional simulations. Note that the optimal projects are the same as the ones obtained with the Bayesian approach. The project values are different with the exception of the poor data set which gave the closest Bayesian and conditional simulations reserves distributions (Figure 4.8). The expected value of additional information is $ 92.50 million that is it augments the value of the project by 21.61 %. Thus the value of the projects and the expected value of additional information obtained with the Bayesian or *plug-in* approach differ.

The comparison of the project values computed given just the initial data and given the initial + additional data permits us to check whether the cost of additional information is justified by a better definition of the project or not. To value additional information the change in the risk of undertaking the optimal project due to the extra-holes should also be taken into account. We need to define a loss function and a corresponding risk function. These will be specified in the next chapter.

Note that in order to obtain additional information the project must be delayed. Moreover, to evaluate this project and choose the best development option the uncertainty on gold price and the time factor must be taken into account.

For comparison purposes, we computed the project values for the *drift* reserves. They are negative. However, note that the results cannot be compared as assuming the presence of a drift the large and small pit must be re- defined as well as the exploitation hypothesis.

...The NPV approach

Present value discounting recognizes time preference. The general equation for the NPV is

$$\text{NPV} = \frac{\text{CF}_t}{(1+r)^t}$$

where CF_t is the cash flow realized at a future time t and r is the discount rate which represents the decision maker's attitude toward time value of money.

The expected price is still used, but the discounting is now considered with rate $r = 2\%$. We will discuss in the next chapter the choice of the discount rate. The life of the mine is fixed at $T = 7$ years if the large pit is developed, and at $T = 5$ years if the small pit is developed. In addition if the extra information is obtained, we must wait for 1 year to develop the pit. The Monte Carlo estimate of the NPV is

$$\frac{1}{L}\sum_{l=1}^{L}\left\{\sum_{t=0}^{T-1}\frac{1}{(1+r)^t}\left(SB_l - VC_U W_l\right)\right\} - FC - IC$$

where B and W are now the annual reserves and waste. Table 4.7 presents the project values calculated with the NPV approach for the Bayesian reserves. Note the reduction of the 2 expected project values due to the discounting. The optimal project given the initial information is U_1. Note that the 2 project values are very close. As it can be seen in Figure 4.22 the NPV of project U_2 is less dispersed and thus project U_2 should be

chosen. Given the rich and the average additional information the optimal project is still U_1. The difference between the 2 projects is now higher. However, Figure 4.22 shows that the average extra data does not help discriminating the 2 projects. Given the poor additional information the optimal project is U_2. The expected value of additional information is $ 39 million that is it augments the value of the project by 9.35 %. Thus it is worthwhile to obtain it.

Available data	**Project value** Large pit (U_1)	Small pit (U_2)
Bayesian approach		
Initial data	**417**	410
Rich new data set	**558**	415
Average new data set	**449**	423
Poor new data set	328	**368**
plug-in approach		
Initial data	**398**	384
Rich new data set	**599**	443
Average new data set	**468**	386
Poor new data set	340	**366**

Table 4.7: *Expected NPV of the two projects in $ million. The results for both the Bayesian and* plug-in *approach are presented. The optimal projects given the initial and initial + additional data are highlighted*

Table 4.7 also presents the results obtained with conditional simulations. Note that the optimal projects are the same as the ones obtained with the Bayesian approach. The expected value of additional information is $ 77.25 million that is it augments the value of the project by 19.41 %. Thus the value of the projects and the expected value of additional information obtained with the Bayesian or *plug-in* approach differ.

The uncertainty on the gold price has to be taken into account. For this a stochastic process could be defined. However, the NPV approach fails to capture all sources of value associated with the irreversible investment opportunity. It does not value future investment and operating paths that are optional. It cannot give value to the option for active or strategic management in response to evolving price and reserves information. If the price drops, for example, management can decide to abandon the project temporarily or permanently. Without taking into consideration these options the investment will then be undervalued. The added value comes from managerial flexibility.

...Towards real options

Real options to value the case study project is the subject of the next chapter. Here we introduce an example to highlight the value of flexibility.

Suppose a manager embed an option to invest in a project in 1 year by a cost of $ 10 million. The NPV is of $ -50 million if the project is undertaken today and in future the same value is expected. The manager decides to acquire the option to investment. Why? The traditional DCF approach would not recommend to embed an option to invest which is expected to be negative. But the investment is an option, not an obligation. Rational managers will not exercise the option at $t = 1$ year in case of bad news and will exercise only in case of good news. As the option will be exercised only if the NPV > 0, the NPV of the project is now $

90 million. This is the real option value. Figure 4.23 compares the histograms of the static NPV and of the NPV with flexibility.

In the case study, as in the previous example, management can decide to abandon the project if the reserves are found to be too low. Real options will be applied to our project in the next chapter. From now on we will use the reserves obtained with the Bayesian approach.

4.5 Conclusions and perspectives

The Bayesian approach is clearly more interesting when few data are available and the uncertainty regarding the quantities of interest is large. If prior information, that is external but related information, is available, the Bayesian approach permits to combine it with the data in the prediction process. Although the data set we work on is large, we have seen that accounting for the uncertainty on the covariance parameters is important, particularly in the lower part of the deposit.

In the place of the Box- Cox family of power transformations, the anamorphosis function was considered. The uncertainty on the covariance parameters was taken into account. Alternatively, we could have used the lognormal model, for example, if it was appropriate, which corresponds to $\lambda = 0$ in the Box- Cox family of transformations. This model did not seem a realistic model for the data in our test case. Letting the mean and covariance parameters unknown in the Bayesian approach permits to leave some uncertainty on this transformation. And most importantly, compared to log- normal kriging the Bayesian method allows to find the optimal predictor. Thus we could have supposed that

$$\log(Z(x)) = \sigma Y(x) + \mu$$

where $Y(x)$ is a standard gaussian random variable,

$$\mu = E[\log(Z(x))] = \boldsymbol{\beta}^T f(x) \quad \text{and} \quad \sigma^2 = \text{Var}[\log(Z(x))] = \alpha^{-1}.$$

This model is interesting as there is *permanence* of log- normality (Chilès & Delfiner, 1999; p 433): when $Z(x)$ has a log- normal distribution, $Z(v)$ is still log- normal (as long as v is small).

We assumed that the exploited ore tonnage per year is constant. If the drift function was defined and the average grade was exploited according to the drift the evaluation of the 2 projects would be very different of the one we obtained.

A first evaluation of the two projects, development of the large pit and development of the small pit, was presented and the difference in the results computed with either the Bayesian method or conditional simulations was highlighted. Note that, although the optimal project is the same, the results obtained with the Bayesian and *plug- in* approaches and the expected values of additional information are quite different. We have seen that the reserves distributions principally differ in the tails leading to different expected values for the 2 projects. Thus accounting for the reserves uncertainty is important.

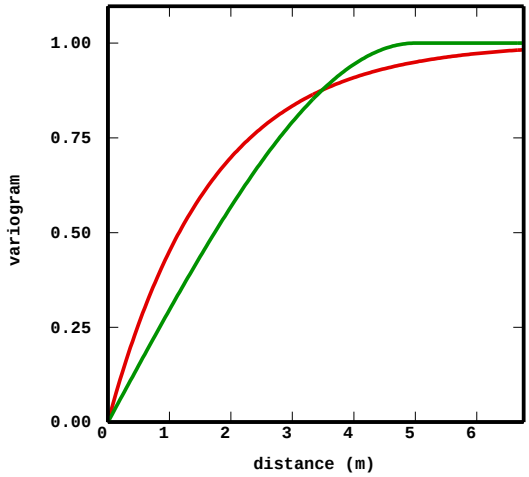


Figure 4.1: *Behavior of the spherical (green) and exponential (red) models near the origin. They have the same sill and range. Both have linear behavior at the origin*

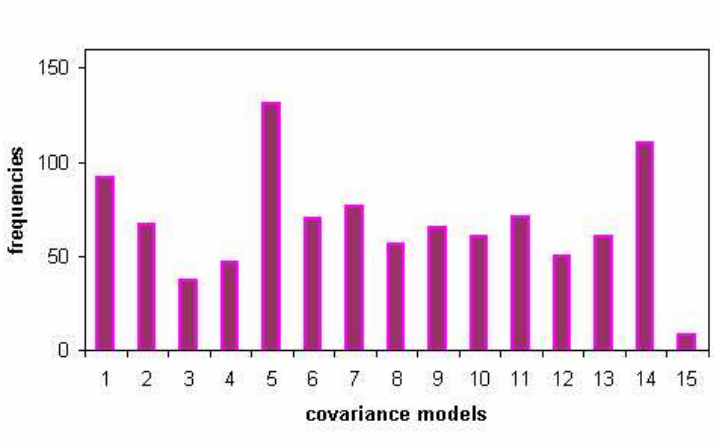


Figure 4.2: *Histogram of the simulated covariance models K_i. Models 1, 5 and 14 have the highest posterior probability. Model 14 corresponds to the* plug-in *model. Model 1 is very similar to model 14. Model 5 does not have a nugget effect*

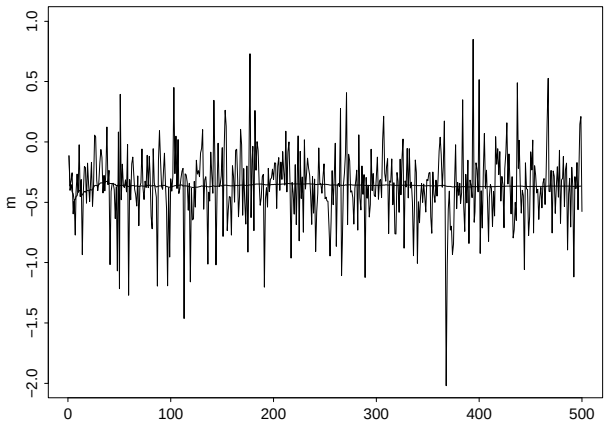


Figure 4.3: *A Monte Carlo simulation of the mean parameter β. The figure also presents the evolution of its Monte Carlo estimate as a function of the number of iterations. To get a sample of 50 the last 50 values are kept. The Monte Carlo estimate of the mean is -0.34*

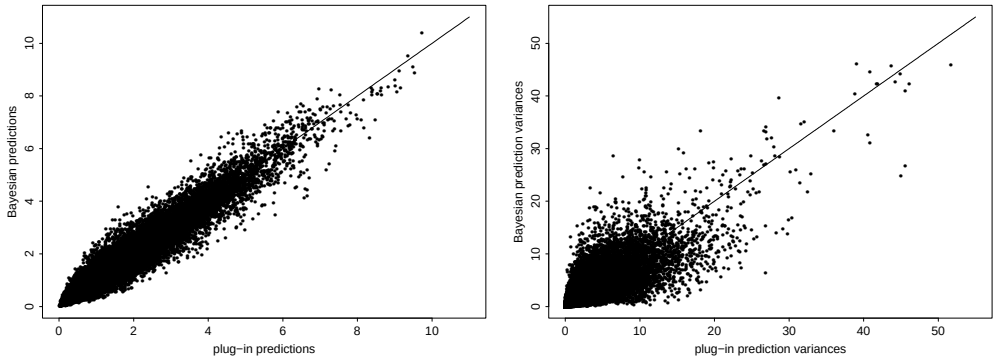


Figure 4.4: *Comparison of the plug- in and Bayesian predictions (left) and prediction variances (right) for the small pit*

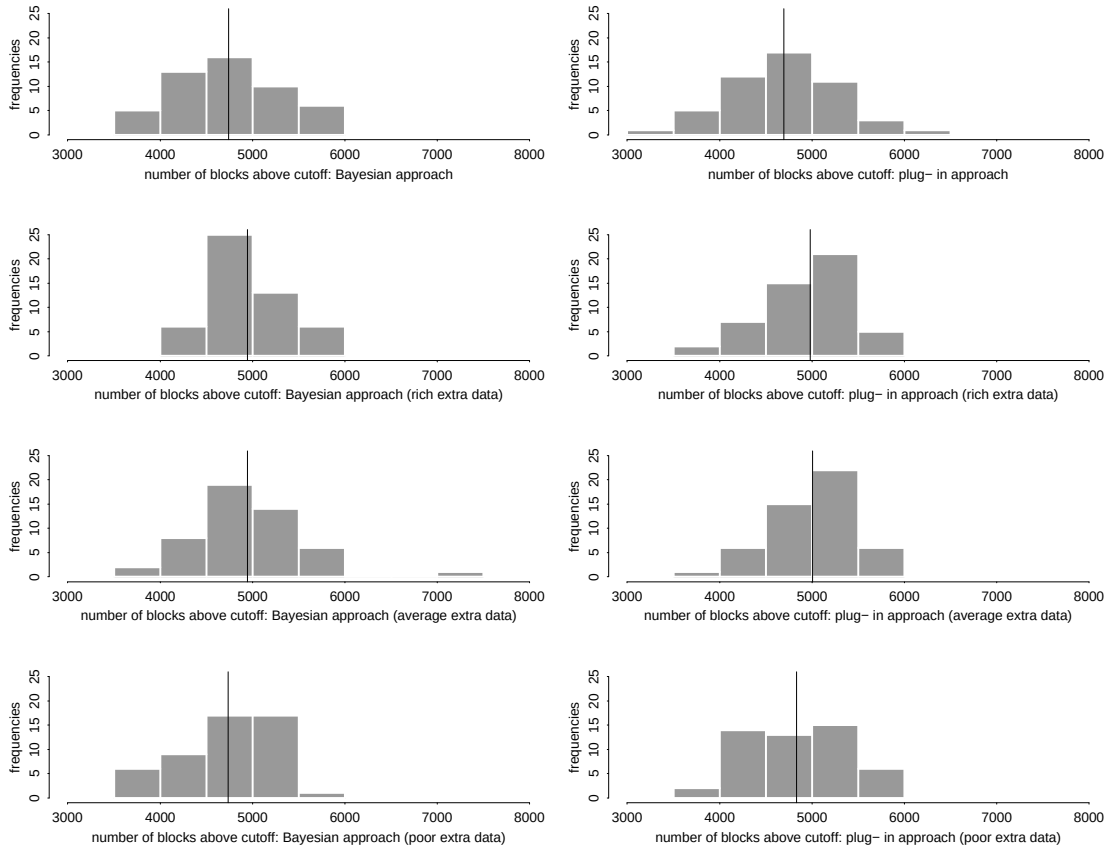


Figure 4.5: *Lower part of large pit: histograms of the number of blocks above cutoff. The histograms on the left are obtained with the Bayesian approach. The histograms on the right are obtained with the conditional simulations approach. The vertical line represents the mean. The expected number of blocks above cutoff is more or less the same whether the Bayesian or plug- in approach is used. Given the initial data the* plug-in *distribution gives a slightly larger weight to the tails*

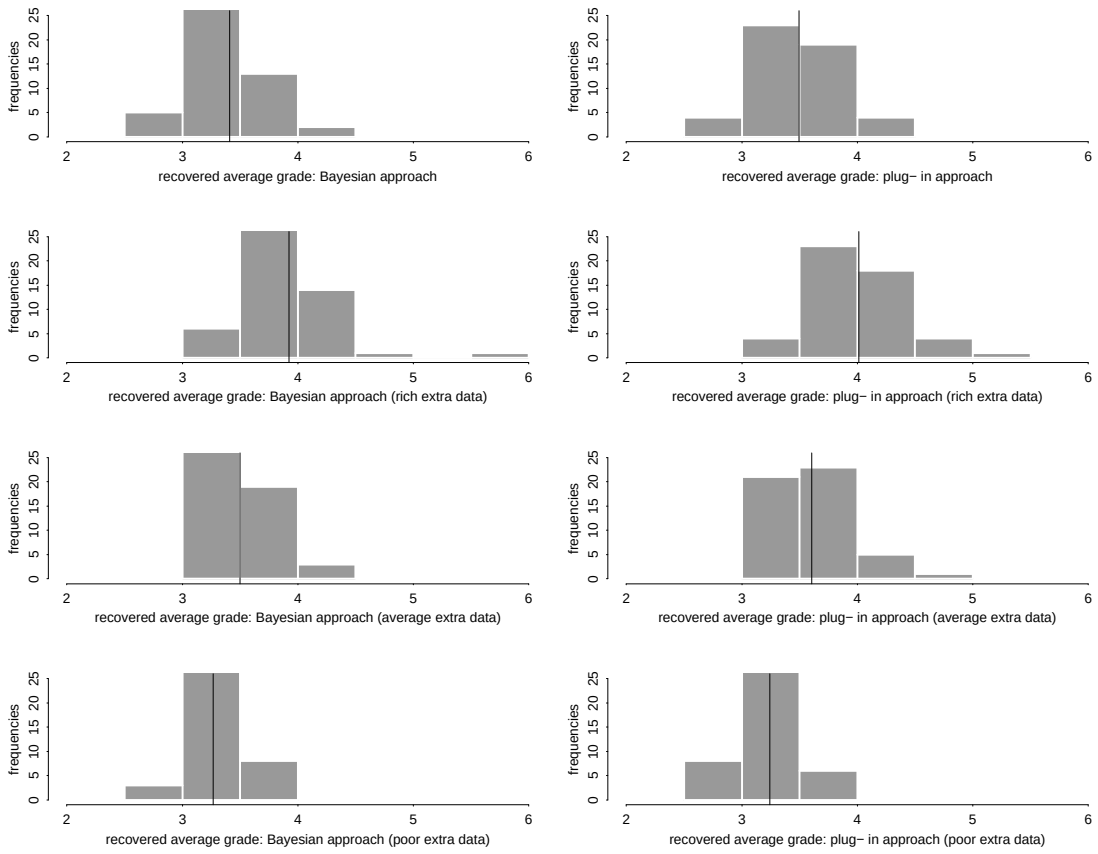


Figure 4.6: *Lower part of large pit: histograms of the recovered average grade (g/t) above cutoff. The histograms on the left are obtained with the Bayesian approach. The histograms on the right are obtained with the conditional simulations approach. The vertical line represents the mean. Given the rich additional information the* plug-in *distribution gives smaller weight to the low values. Given the average additional information the* plug-in *distribution gives larger weight to the high values*

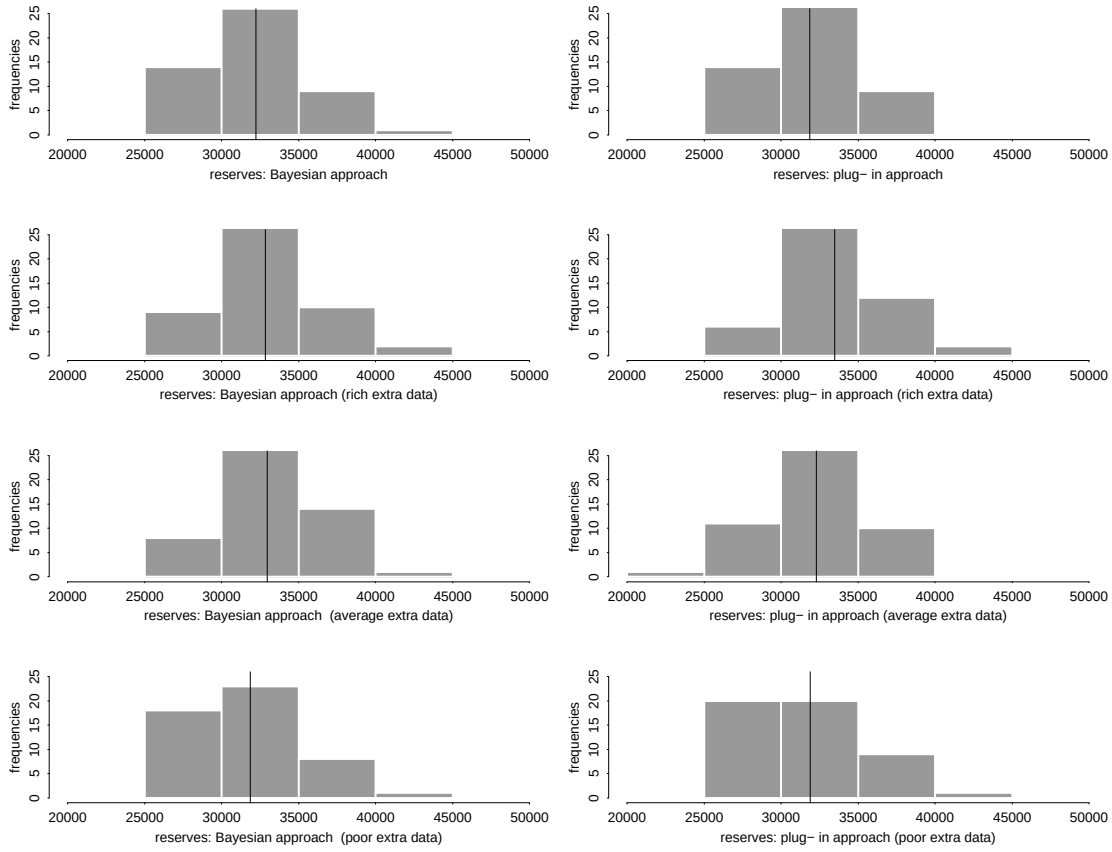


Figure 4.7: *Small pit: histograms of the reserves (g/t). The histograms on the left are obtained with the Bayesian approach. The histograms on the right are obtained with the conditional simulations approach. The vertical line represents the mean. As expected, the two histograms are very similar. The expected reserves obtained given only the initial data or given the extra data is more or less the same. The expected reserves obtained with the plug- in or the Bayesian approach is more or less the same. An exception are the rich extra data*

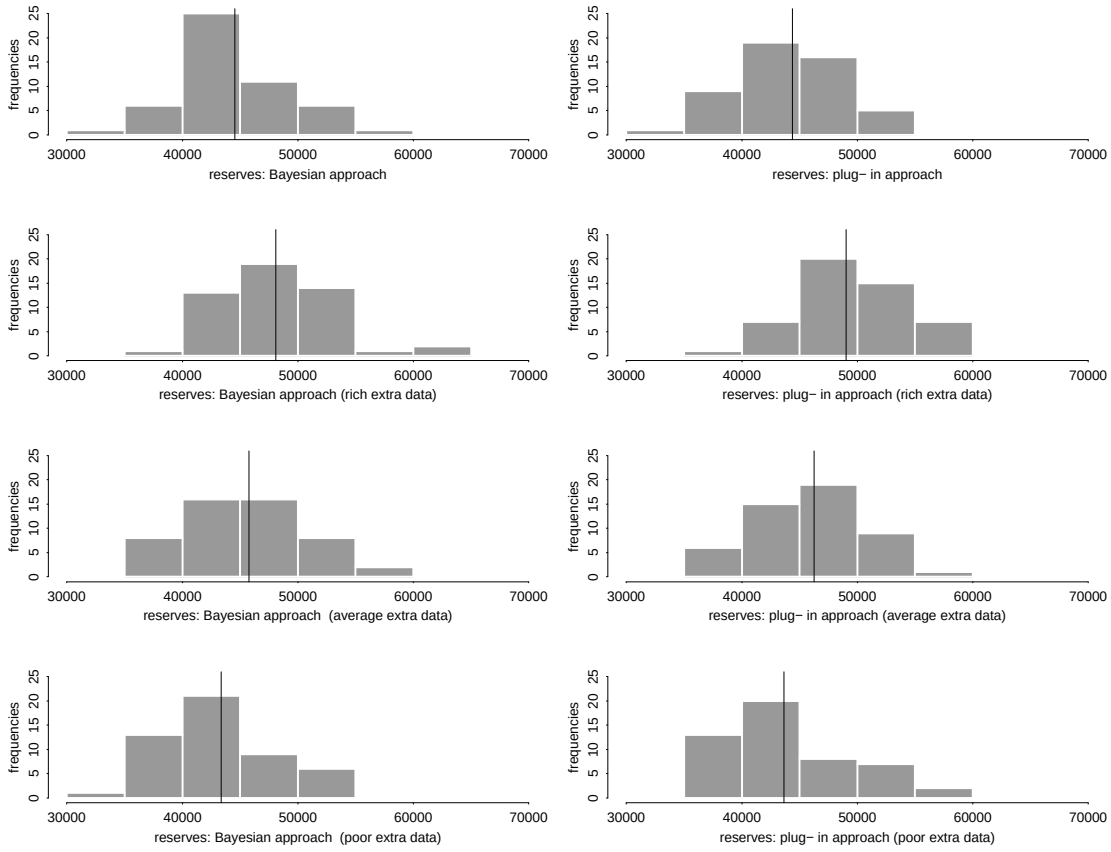


Figure 4.8: *Large pit: histograms of the reserves (g/t). The histograms on the left are obtained with the Bayesian approach. The histograms on the right are obtained with the conditional simulations approach. The vertical line represents the mean. Note that the expected values are more or less the same whether the Bayesian or plug- in approach is used. The histograms differ. In particular, note that they differ in the tails giving different weight to large or small values. An exception are the poor extra data: the 2 histograms are very similar*

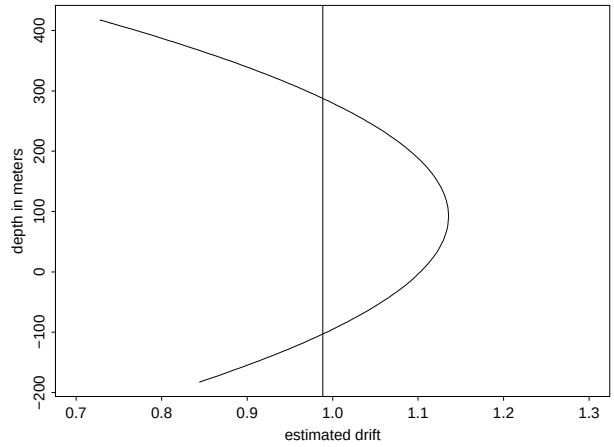


Figure 4.9: *Estimated quadratic drift, that is $\beta_0 + \beta_1 x_3 + \beta_2 x_3^2$ where x_3 is the vertical coordinate and $\beta_0, \beta_1, \beta_2$ are the drift coefficients. The vertical line is the estimated constant mean. The presence of a drift seems plausible. It helps highlighting that the zone between 50m and 300m is the richest (see Figure 2.3) and that also the lower part from 0m to -150m is interesting. On the contrary, the top zone from 300m up and the lowest zone from -150m down do not seem interesting*

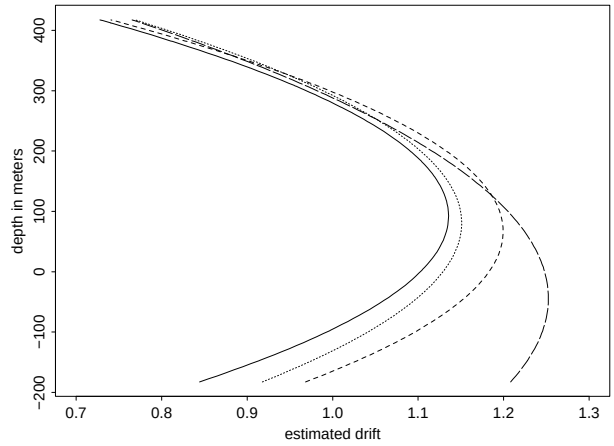


Figure 4.10: *Estimated quadratic drifts given the initial data (continuous curve), the initial data + rich additional information (large dashed curve), the initial data + average additional information (dashed curve) and the initial data + poor additional information (dotted curve). Note that the rich extra data are the richest for the lower zone not for the upper zone*

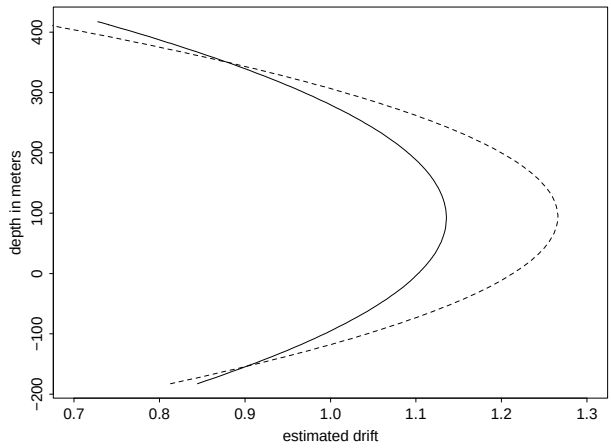


Figure 4.11: *Estimated quadratic drifts given the initial data. The continuous curve is obtained with a covariance model without nugget effect. The dashed curve is obtained with a covariance model with a nugget effect. Observe that without a nugget effect we force the continuity of the phenomenon*

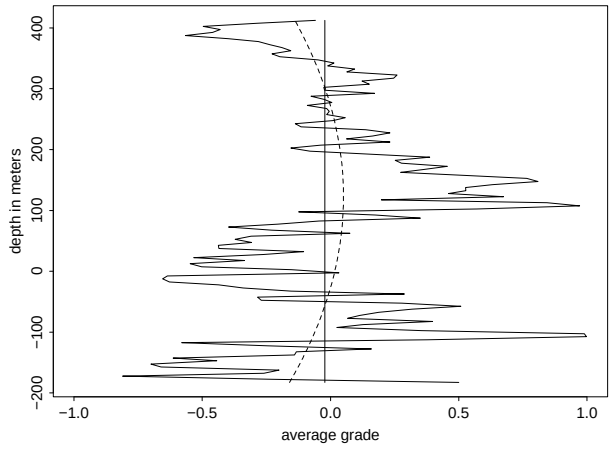


Figure 4.12: *The average grade of the gaussian variable is computed for each depth. The estimated quadratic drift and constant mean are also presented*

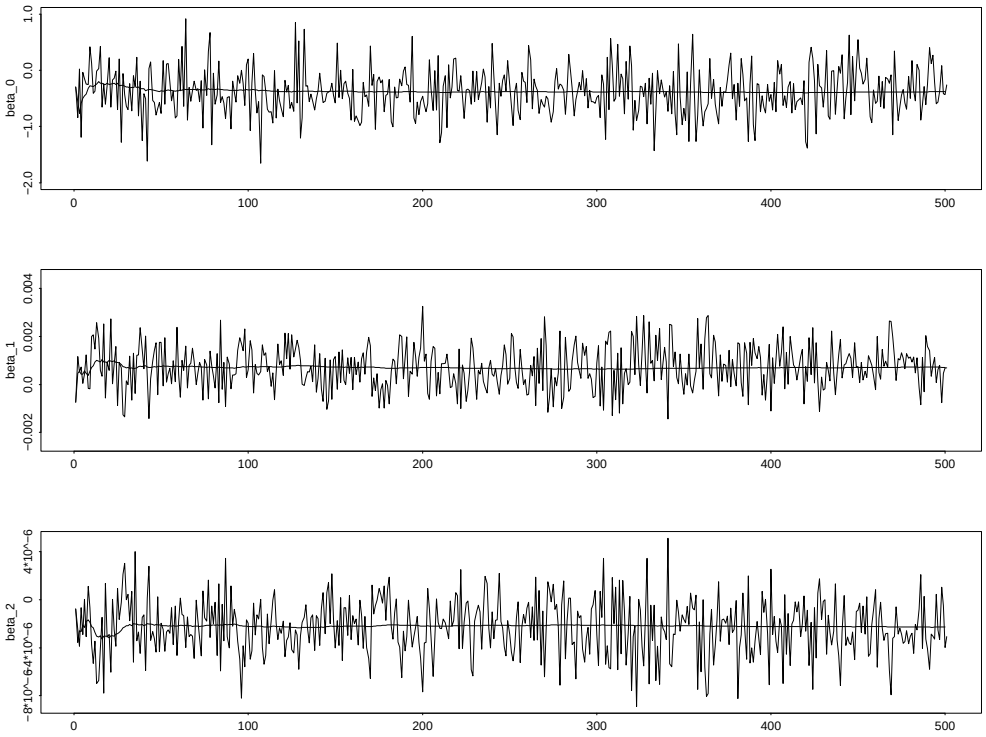


Figure 4.13: *A Monte Carlo simulation of the mean parameter β. The figure presents the evolution of the Monte Carlo estimates as a function of the number of iterations. The Monte Carlo estimate of β is* $(-0.38, 7.13 \times 10^{-4}, -2.27 \times 10^{-6})$

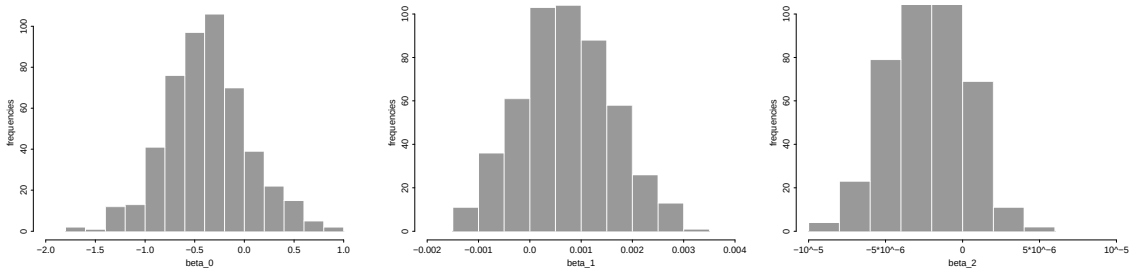


Figure 4.14: *Histograms of the samples of β. To get a sample of 50 the last 50 samples are kept. The Monte Carlo estimate of β is* $(-0.32, 8.28 \times 10^{-4}, -2.55 \times 10^{-6})$. *The coefficients of variation are 1.08, 0.86 and 0.76 for β_0, β_1 and β_2, respectively*

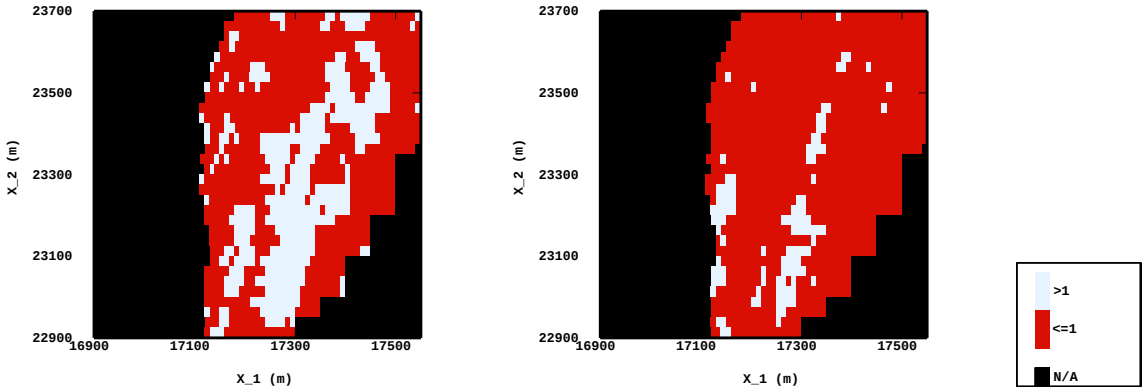


Figure 4.15: *Comparison of 2 simulations obtained with a constant mean (on the left) and with a quadratic drift (on the right): horizontal section with depth fixed at 305m*

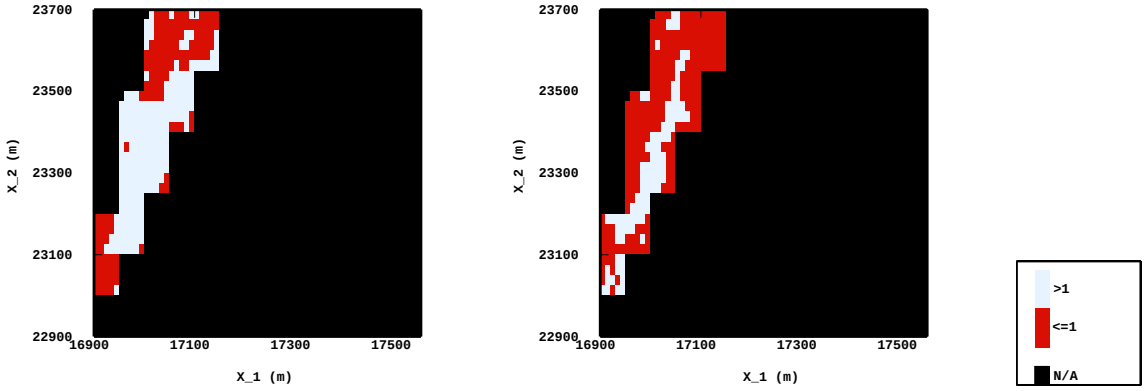


Figure 4.16: *Comparison of 2 simulations obtained with a constant mean (on the left) and with a quadratic drift (on the right): horizontal section with depth fixed at -175m*

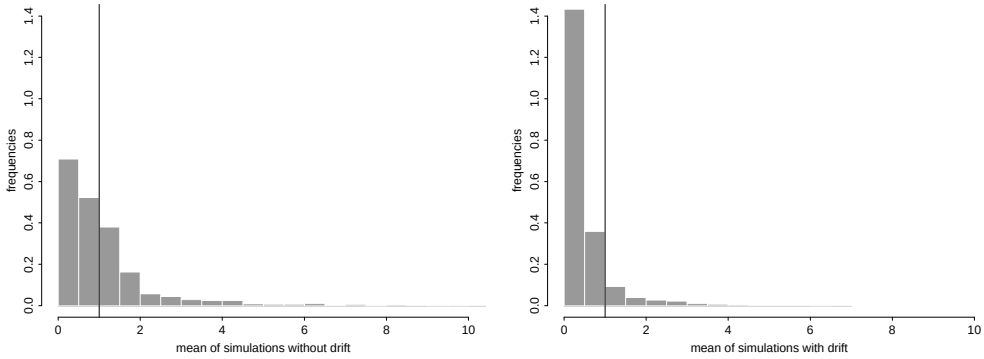


Figure 4.17: *Histograms of the mean of the 50 simulations obtained with a constant mean (on the left) and with a quadratic drift (on the right) for the depth fixed at 305m. The vertical line represents the cutoff grade*

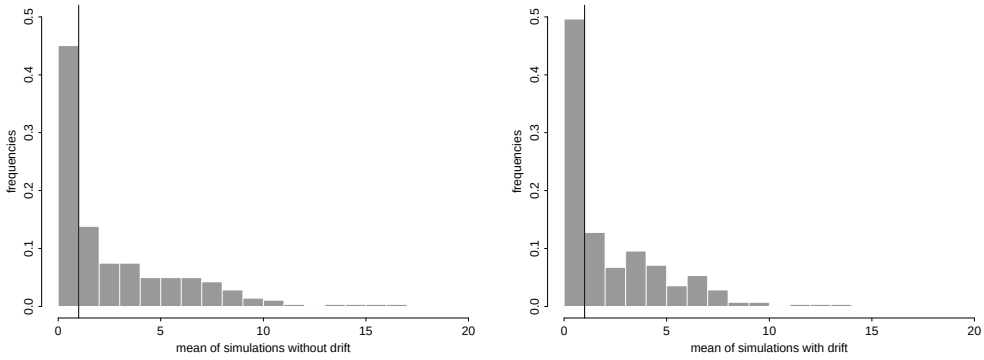


Figure 4.18: *Histograms of the mean of the 50 simulations obtained with a constant mean (on the left) and with a quadratic drift (on the right) for the depth fixed at -175m. The vertical line represents the cutoff grade*

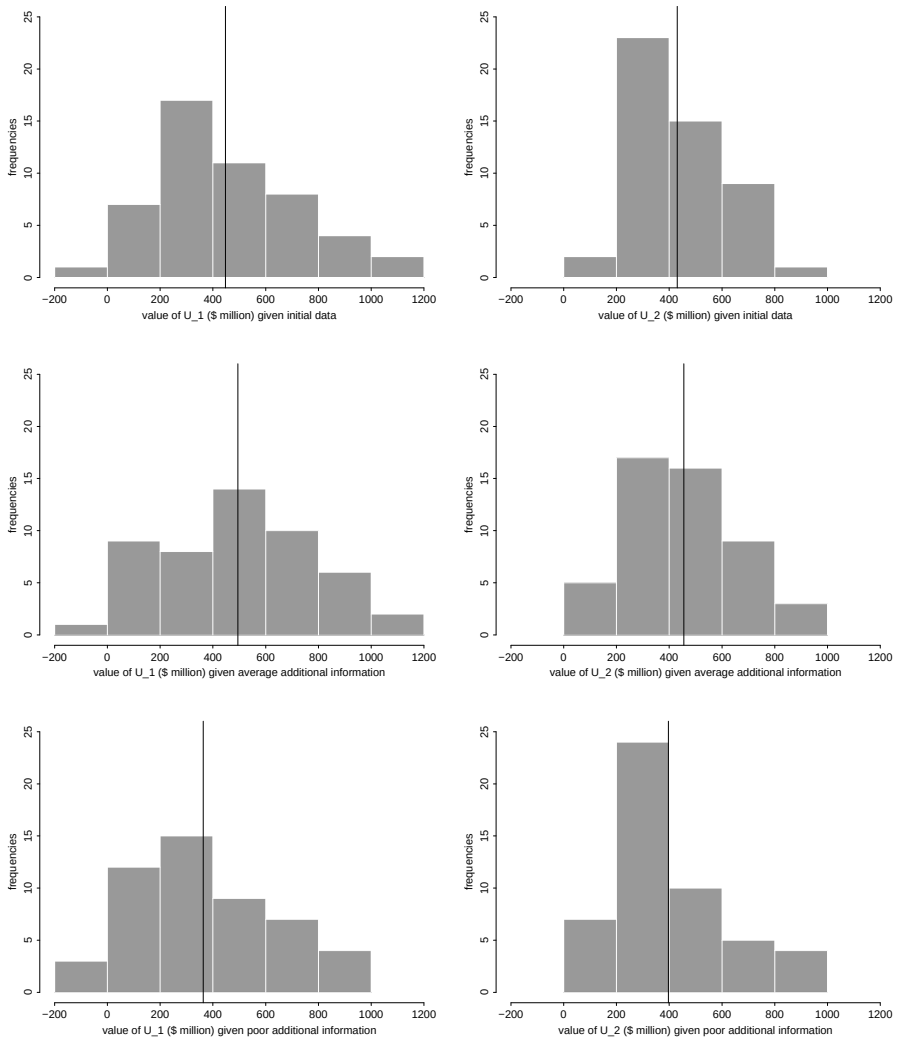


Figure 4.19: *Histograms of the Monte Carlo project values for U_1 (on the left) and U_2 (on the right) given the initial, the average and poor data sets. The vertical line represents the expected project value. As expected, the results for U_2 are less dispersed. The probability of a high profit (> $ 800 million) is higher for U_1. However, the probability of a negative or small profit (< $ 200 million) is also higher for U_1. Given the average additional information the two projects are very close. Although the expected profit of U_1 is higher, project U_2 could be preferred as it is less risked. However, the probability of a high profit for U_1 is still higher than for U_2. This information does not help choosing the optimal project. Given the poor information project U_1 is not interesting as the probability of a high profit (> $ 800 million) is the same as that for project U_2. The expected profit for U_2 is now higher*

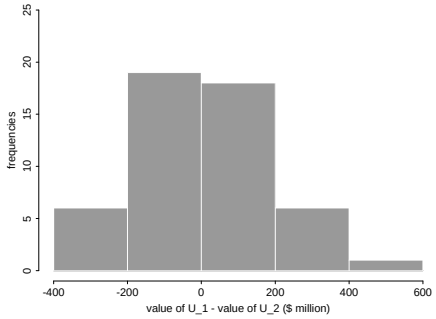


Figure 4.20: *Histogram of the differences of the 2 project values. Note that for 25 of the 50 simulations project U_1 is preferred to project U_2, that is it has an higher expected value than U_2. Thus the risk of taking the wrong decision is high*

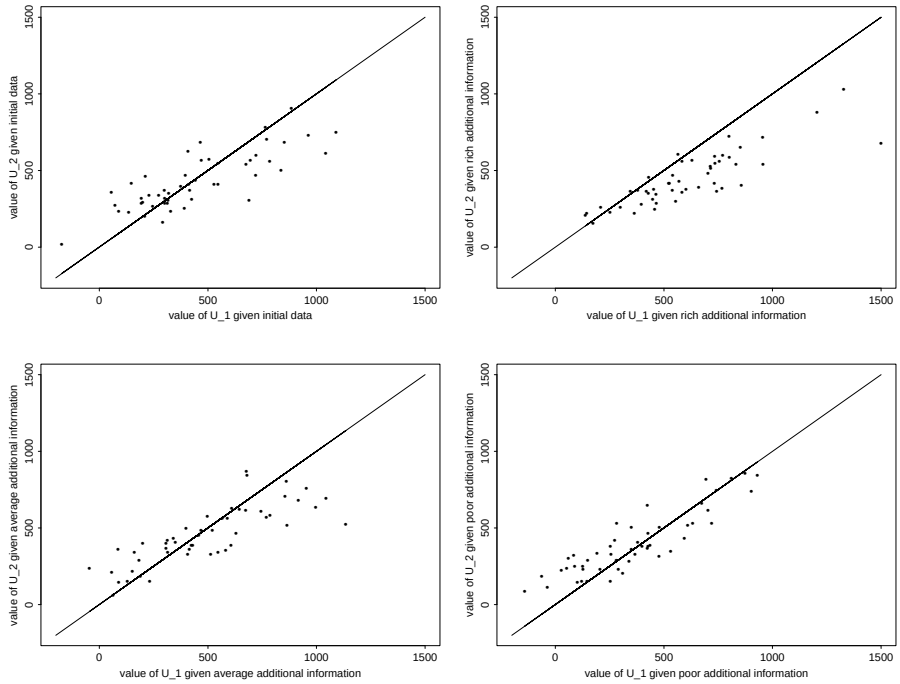


Figure 4.21: *Comparison of the 2 project values ($ million) given the initial data and given the initial + rich, average and poor extra data. Given the rich additional information, project U_1 can be strictly preferred to U_2. The average additional information does not help choosing the optimal project. The expected profit of U_1 is still higher than that of U_2, but the risk of taking the wrong decision is still high. Given the poor extra data the two projects are even closer*

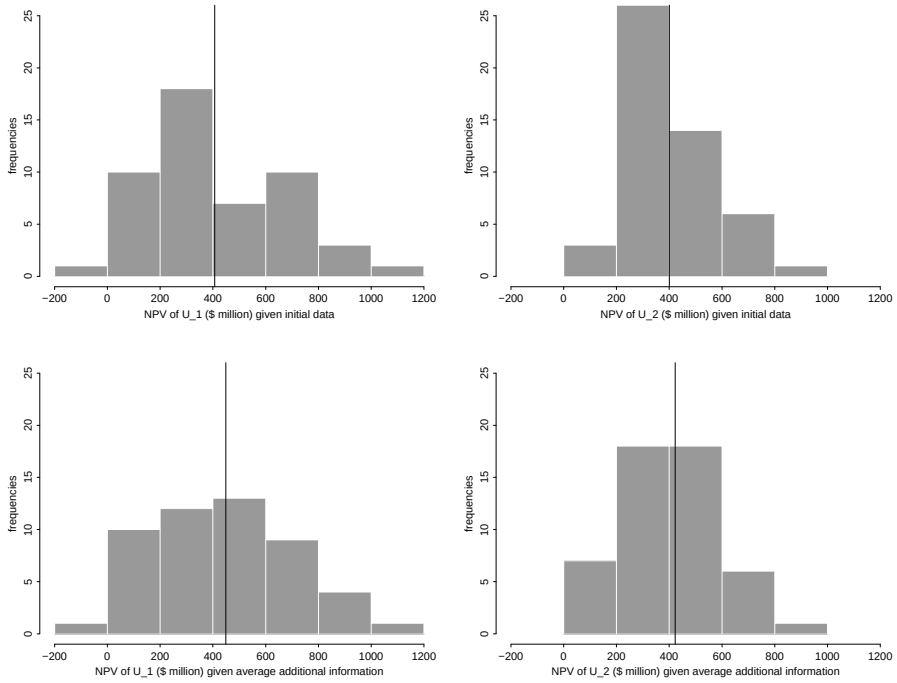


Figure 4.22: *Histograms of the Monte Carlo project values for U_1 (on the left) and U_2 (on the right) given the initial data (above) and the initial + average extra data (below). The vertical line represents the expected project value. The results for U_2 are less dispersed. The probability of small NPV (< $ 200 million) is much higher for U_1 than for U_2. Whereas the probability of high NPV (> $ 800 million) is slightly higher for U_1 than U_2. The average additional information does not help choosing the optimal project*

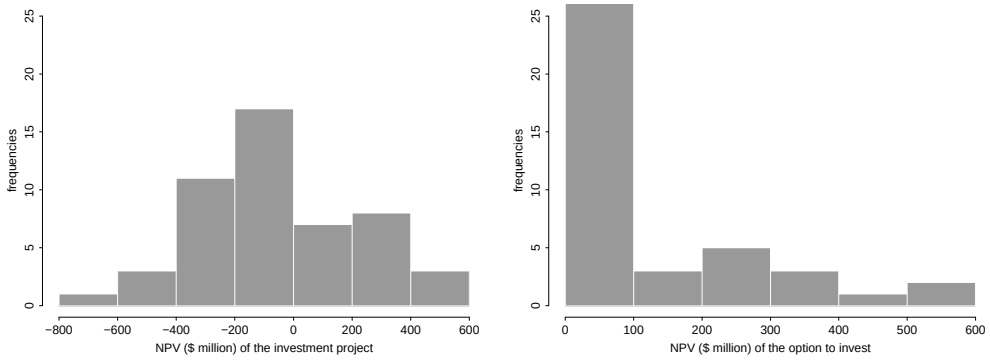


Figure 4.23: *Comparison of the static NPV and of the NPV with flexibility*

Chapter 5

Real options

5.1 Introduction

In the 1970s and the 1980s, developments in the valuation of capital investment opportunities based on option pricing revolutionized capital budgeting. Managerial flexibility to adapt and revise future decisions in order to capitalize on favorable future opportunities or to limit losses has proven vital to long-term corporate success in an uncertain and changing market place such as in mineral and energy markets.

Nowadays it is widely recognised that traditional discounted cash flow approaches fail to capture the impact of two important factors: uncertain costs and prices, and managerial flexibility. Decision trees and Monte Carlo simulations were developed to estimate the probability distribution of future cash flows, but they have difficulties to incorporate future decision possibilities or contingencies in project value (Trigeorgis, 1996).

The similarity between real and financial decision making was first recognized by Brennan and Schwartz (1985) and McDonald and Siegel (1985). They extended the financial option theories of Black and Scholes (1973) and Merton (1973) to evaluate irreversible real investments, such as those in mining. Investment opportunities are viewed as options. This led to the development of real options analysis.

The motivation for using an option-based approach to capital budgeting arises from its potential to conceptualize and quantify the flexibility component value that arises from active management and strategic interactions. This value is seen as a collection of *real options* embedded in investment opportunities having as underlying asset the gross project value of discounted expected cash flows. It must be noted that the real options approach is not a substitute for traditional discounted cash flow analysis, but an enhancement of discounted cash flow principles that allows to properly value managerial flexibility.

Trigeorgis (1996) identified four basic sources of flexibility. These are

- *initiation flexibility* (defer options): this refers to the possibility of waiting until more information has become available. The possibility to decide when, or if, to start can be of great value to a mine or oil development project. The major gain from a postponement comes from the benefit of getting new information. One example is the possibility to wait and observe the gold price if the current price happens to be low. By being able to wait, the project can utilise information that arrives and take advantage of better knowledge. This necessarily adds value to the project. For example, consider a

mineral property that is uneconomic if developed immediately, as its NPV is negative. Given that there is some probability that prices will rise in the future, this project is negative only because of the assumption that development and operations must begin immediately. The managerial option to delay development until prices rise, gives some properties a positive expected present value. " This is why we see unattractive undeveloped projects trading in the market at positive values, and then sitting dormant for years " (Davis, 1996). Initiation flexibility, also called investment timing flexibility, is similar to having an American call option on the value of the project. The investment cost is the exercise price of the option, while the last possible time to start the project is the option's expiration date. Instead of using a positive NPV as criterion for project initiation, the decision rule is now to defer the start until the price has reached a given level.

Operating flexibility includes

- *termination flexibility* (abandonment options): this offers the possibility to make the investment in stages, deciding at each stage, based on the newest information, whether to proceed further or whether to stop. In fact, it may not be optimal to continue operations if events, either internal or external, with a negative impact on the project occur. The value of being able to terminate can then be substantial, in particular for projects with a long expected remaining lifetime. As the development project evolves the value of the termination flexibility declines. Again the similarity to financial options must be noted where the value of the option decreases as the time to expiration (the maturity date) approaches. For the development project the remaining life time corresponds to the time to expiration.

- *Contraction/ expansion flexibility* or *stop/ start flexibility*: this represents the possibility to adjust the scale of the investment depending on whether market conditions turn out favorably or not. The possibility of starting and stopping the project during its operation can be of significant value. Suppose, in the case of a mine, that management optimally shuts the mine during periods of sufficiently low prices and that management optimally reopens the mine when prices have risen to a sufficiently profitable level. The value of the project will then be higher than the value obtained through the *flat price* DCF technique. However, it must be noted that the cost of starting and stopping the project can be high, making the flexibility less valuable. Managers prefer keeping mines open even if the actual ore price is low.

- *Capacity flexibility* (switching options): this allows changing to a different technology. In mining, it is, for example, equivalent to varying cutoff rates. Increasing the cutoff grade raises the average grade of ore produced but shortens the mine's life.

It must be observed that the value of different flexibility types is not in general additive (Lund, 1999). This work focuses on a special case of the first option (defer) in which the company has the possibility to wait and obtain additional information on key parameter values. This option has been analysed among others by Lund (1999), Galli *et al.* (2001) and Dias (2002).

The chapter is structured as follows. In Section 2 we review traditional discounted cash flow (DCF) and decision trees, explaining their advantages and disadvantages. Section 3 is devoted to real options. In Section 4 we present contingent claims analysis and stochastic dynamic programming, the two equivalent procedures to evaluate real options. In Section 5 we review applications that have used contingent claims analysis in mining and applications that have used dynamic programming. Section 6 describes information flexibility. In Section 7 the model used to value flexibility is presented and applied to evaluate the gold mining project.

5.2 Discounted cash flow methods and decision trees

The fundamental idea of the discounted cash flow approach is that the value of a project is defined as " the future expected cash flows discounted at a rate that reflects the riskiness of the cash flow " (Smith and Nau, 1995). Major problems with the DCF methods concern its ability to estimate future cash flows, the possibility to properly value management flexibility and the determination of the appropriate discount rate. Traditional discounted cash-flow approaches make implicit assumptions concerning an *expected scenario* of cash flows and presume management's passive commitment to a certain static *operating strategy*. In order to implement DCF analysis, we need estimates of expected future cash flows and a discount rate deemed appropriate to their risk. Discounted cash flow only uses information that is known at the time of the appraisal. This approach neglects the stochastic nature of prices (instead of the distribution of future prices their expected value is considered) and of possible managerial responses to price variations leading to suboptimal investment decisions. Traditional DCF approaches to the appraisal of capital investments projects, the net present value rule for example, do not properly capture management's flexibility to adapt and revise later decisions in response to unexpected market developments. In the real world which is characterized by change, uncertainty and competitive interactions, cash flows differ from what management expected at the outset. As new information arrives and uncertainty about market conditions and future cash flows is gradually resolved, management may have valuable flexibility to alter its initial operating strategy in order to capitalize on favorable future opportunities or to react so as to mitigate losses. For example, management may be able to defer, expand, contract, abandon, or alter a project at various stages of its useful operating life. This means that management may truncate the distributions of risk on the downside and augment the upside distribution of the chances for success (Trigeorgis, 1996). Thus, when applied to systems operating in an uncertain environment, the DCF procedures fail to recognize that effective management of the risks enhances the value of the system.

The value of active management is better captured by using a decision tree. Flexibility is modelled through decision nodes allowing future managerial decisions to be made after some uncertainty has been resolved and more information has been obtained, before proceeding to the next stage. However, some problems are quite cumbersome to solve as decision trees. Dynamic programming can then be used to evaluate flexible decision models. Moreover, the presence of flexibility embedded in future decision nodes changes the payoff structure and the risk characteristics of an actively managed asset in a way that invalidates the use of a constant discount rate. Unfortunately, decision tree analysis does not provide any recommendation concerning the appropriate discount rate.

5.3 Real options

Managerial flexibility is likened to financial options. Its quantification was first solved by Brennan and Schwartz (1985) using option pricing techniques. If price uncertainty is assumed to be the only source of uncertainty, an investment opportunity is analogous to an American call option, not yet exercised, on a common stock: it gives the right- but not the obligation- to acquire the gross present value of expected cash flows by making an irreversible investment on or before the date the opportunity ceases to be available. Using the analogy with options on financial assets, such investment flexibility is called a *real option*. Real options have value when investment involves an irreversible cost in an uncertain environment. The beneficial asymmetry between the right and the obligation to invest under these conditions is what generates the option's value. In order to exercise a real option, one must pay the exercise price. Table 5.1 compares a call option on stock to a real option on project.

Call option on stock	**Real option on project**
Current value of stock	Gross present value of expected cash flows
Exercise price	Investment cost
Time to expiration	Time until opportunity disappears
Stock value uncertainty	Project value uncertainty
Riskless interest rate	Riskless interest rate

Table 5.1: *Comparison between a call option on a stock and a real option on a project*

Moreover, as with options on financial securities, management, after an initial investment, can gather more information about project progress, market characteristics and competitive reactions and, based on this information, change its course of action. The value of this managerial flexibility enhances the project value by improving its upside potential while limiting downside losses relative to the initial expectations of a passive management. Brennan and Schwartz found that, for mineral assets, the additional value created through optimally executed managerial flexibility can be *priced* just as American stock options are priced. The manager owns managerial options that add value to the asset, that is

$$\text{Expanded value} = \text{DCF value} + \text{option premium}$$

where the expanded value is in fact the asset's option price (Trigeorgis, 1996). The magnitude of the nonnegative option premium derived from managerial flexibility, the option value, depends on the characteristics of the underlying asset.

The real options approach recognizes the value of flexibility in the context of uncertainty, especially when system operators can manage this uncertainty. As Davis (1996) observes, in financial option pricing, the value of an option depends critically on several parameters. First, the more volatile the underlying asset, the more valuable the option written on that asset. By analogy, the more volatile the price of the mineral being produced, the more valuable the option to manage the mineral asset. If future prices were certain, the option premium associated with managerial flexibility would become zero. Hence, managerial flexibility is worth more for mineral assets with high price volatility. Second, " financial options that are at or near the money have more value than those well out of the money and unlikely to be exercised prior to expiration[1]. From this, the more marginal the mineral asset, where the operating options are at or near the money, the more valuable managerial flexibility ". Finally, longer time to expiration increases an option's value. Analogously, managerial flexibility is worth more when the time frame of that flexibility is extended.

The real options method of valuation provides a useful framework for strategic decision making. One key insight generated by this approach to investment is that " higher uncertainty in the payoffs of the investment increases the value of managerial flexibility, or the value of real option " (Dixit and Pindyck, 1994). As Huchzermeier and Loch (2001) observe, " the intuition is clear- with higher payoff uncertainty, flexibility

[1] An in the money option is one that would lead to a positive cash fbw to the holder if it were exercised immediately. An at the money option would lead to zero cash fbw if it were exercised immediately, and an out of the money option would lead to a negative cash fbw if it were exercised immediately

has a higher payoff potential of enhancing the upside while limiting the downside ". An important managerial implication is that the more uncertain the project payoff is, the more effort should be made to delay irreversible commitments and maintain the flexibility to change the course of action. Moreover, the tools from finance theory including stochastic calculus and dynamic programming can be used.

Although there are many similarities between real and financial options, there are also important differences. Several implications flow from the fact that real options deal with physical projects

- most real-life projects are equivalent to a sequence of options whose values may interact (see Trigeorgis (1996) for more details on compound options). For example, acquiring a property gives the owner the option to explore, obtaining information from the exploration phase gives him the option to develop, and developing gives him the option to extract. Moreover, after a mine is fully operational, the owner has the option to mothball or to abandon. These are in general valued separately. An early exception is the work of Brennan and Schwartz (1985), who determine the combined value of the options to shut down (and restart) a mine and to abandon it for salvage.

- Managers make decisions about whether to acquire a real option, for example to exploit a mine, only once.

- Whereas the analysis of options on commodities and stocks can be based on years of data on the volatility of these assets, the analysis of real options may have little historical data to draw upon. Fortunately, as de Neufville (2001) states, managers of technological systems, for example, do not require great accuracy because they typically only need to make choices, not precise judgements. In making a choice, one only needs to know the relative value of alternatives, not their precise value. To decide whether to undertake a R&D process that will lead to a real option on the launch of a new product, for example, managers only need to know if the value of the option is greater than the cost to acquire it. Options analysis is thus quite different for managers than for financial analysts who have to decide on a precise price to pay for options, as they trade them day after day.

- In addition, a real option analysis must take account of both technical and market considerations. In particular, a specific mine project, for example, is not traded in the market. The use of *similar* mines is an approximation. By contrast, financial options are simply paper trades.

Real options is particularly interesting for R&D activities and in general activities that proceed through a linear developmental process. For example, each of the steps in the discovery and exploitation of natural resources can be thought of as an option on the next phase. Thus a lease on the mineral rights to an area is an option on subsequent exploration, which is in turn an option on eventual exploitation. Real options analysis may be used to value such projects and to determine the optimal policies for developing, managing and abandoning them.

Applications to natural resources

Many applications focus on valuation problems in the context of the oil industry and mining facilities (for example, Paddock *et al.*, 1988; Stensland and Tjøstheim, 1989; Gibson and Schwartz, 1990, 1993; Smith and McCardle, 1998, 1999; Cortazar and Casassus, 1998; Lund, 1999; Slade, 2000; Moel and Tufano, 2000, 2002). The evaluation of mining and oil projects is made particularly difficult by the high degree of uncertainty in the prices and the output quantities. In the early literature, researchers often adopted simplifying assumptions: the reserves were supposed perfectly homogeneous and known, the cost were supposed known, and they usually modelled price as a geometric Brownian motion (non stationary model). More

recently, researchers have begun to consider a mean reverting model for the price (Gibson and Schwartz, 1990; Slade, 2000; Moel and Tufano, 2000). But as Slade (2000) observes, different models for prices are estimated and compared, but no tests of stationarity are performed. As far as the uncertainty on reserves is concerned, researchers have just started to take it into consideration. Moel and Tufano (2002) study the opening and closing of 285 developed North American gold mines (period 1988-1997). The authors confirm that the probability that a mine is open is related to both market factors (level and volatility of gold price and interest rate) and mine specific factors (its fixed costs, variable costs and reserves).

The possibility to revise the initial predictions on the reserves, as new information is available, is rarely considered. Exceptions are the works of Kemna (1993), Chorn and Carr (1997), Chorn and Croft (1998) and, as was said earlier, Lund (1999), Galli *et al.* (2001) and Dias (2002). This will be seen in more detail in Section 6.

In practice, decision makers do not often apply real options. Moyen *et al.* (1996) conducted a survey to determine how mining companies evaluate projects. The authors interviewed vice presidents in charge of corporate development and treasurers of Canadian mining companies. They found out that

- all firms use some form of DCF calculation to evaluate projects. The base calculation is often supplemented with sensitivity analysis for key parameters such as price.
- Most firms use a long-run commodity price, that is they replace the random variable with its expected value. In addition, there is substantial agreement concerning this price.
- Most firms make adjustments to their DCF calculations, the most important being an increase in the discount rate to reflect risk. The most common rate is 15%.
- Very few decision makers had heard of real options theory and had used it.

This is also the case of petroleum companies that use the DCF approach as a finance tool to value undeveloped reserves (Connell, 2002). However, several global technological companies are beginning to use real options to re-frame the way they think about technology management, innovation and system development (de Neufville, 2001). This is recent, dating only from the mid 1990's.

Following Dixit and Pindyck (1994) we recognise two equivalent methods to evaluate real options and obtain optimal investment rules: contingent claims analysis and stochastic dynamic programming. Both take from decision tree analysis the use of decision nodes in modelling flexibility. But in contrast to the conventional decision analysis that works with a predetermined set of possible decisions, the real options approach seeks to identify new possible paths (i.e. insert additional decision nodes into the tree to reflect the options) to change the decision tree by adding in flexibility. In addition, contingent claims analysis uses NPV's notion of a comparable security to properly price risk.

5.4 Two approaches to evaluate real options

There are two equivalent procedures for evaluating options and risky projects where managerial flexibility plays an important role: one is decision analytic, based on stochastic dynamic programming, and the other is option pricing theory (or contingent claims analysis), based on the no-arbitrage theory[2] of financial markets.

[2]The no-arbitrage principle states that: two investments with the same payoffs at all times and in all states must have the same value

The two methods are identical in modeling flexibility, but they differ in valuing risky cash flows.

5.4.1 Contingent claims analysis

Contingent claims analysis (CCA) builds on ideas from financial economics. The firm that owns the right to an investment opportunity, owns an asset that has a value. The idea behind CCA is to value an option not on his own, but as part of a riskless portfolio. If, for example, the underlying security is a stock, it is possible to take a *long position* in the stock option and a *short position* in the underlying asset (the stock). Both positions are affected by the same source of uncertainty: the stock price. Moreover, the capital gains associated with one investment are completely offset by the losses associated with the other. The rate of return on the portfolio is thus riskless and should equal the risk-free rate (see Appendix D.2 for more details on Options and Options valuation). If this asset is not traded directly on the market, it can be related to other assets that are traded. A combination or a portfolio of traded assets that will exactly replicate the pattern of returns from the investment project, at every future date and in every future uncertain eventuality is thus needed. The value of the investment project must equal the total value of that portfolio. The fundamental assumption underlying this approach is that the value of a nontraded project is " the price the project would have if it were traded " (Smith and Nau, 1995). The required rate of return[3] on the asset is derived as an implication of the overall equilibrium in capital markets. Only the riskless rate is taken to be exogenous. Thus this approach offers a better treatment of the discount rate. But in order for a portfolio that is long in the mine and short in futures contracts to be riskless, there must be a deterministic relationship between commodity spot and futures price. This relationship is determined by the convenience yield[4] that is associated with owning the commodity (Moyen *et al.*, 1996).

The usual assumptions (Black and Scholes, 1973; Merton, 1973) are that trading and decision making take place in continuous time and that the underlying sources of uncertainty follow Brownian motions.

The similarities between financial and real options arise because the ability to control or manage a cash flow stream represents an option, and, more importantly, equivalent martingale pricing techniques are appropriate to both. The major difference is that the rights to controllable cash flows typically cannot be reduced to claims on traded assets: they are contingent on state variables that are not traded. For example, costs include human capital that is not bought and sold (Slade, 2000). This makes the determination of the equivalent martingale measure or risk adjustment more problematic than in the case for financial options. To overcome this difficulty Slade uses the method that was developed by Brennan and Schwartz (1982), Cox *et al.* (1985) and Gibson and Schwartz (1990) which consists of adjusting the drift of each stochastic process by an amount that reflects a process specific risk premium, where the risk premium is obtained from an equilibrium model of financial markets. Future expected cash flows can then be discounted using the risk-free rate. Therefore, unlike conventional real-option models of traded natural resource assets, it is not necessary to model the convenience yield that is associated with holding each asset. An alternative procedure is to adjust the discount rate to reflect risk. But then the risk that is associated with different sources of uncertainty is discounted by the same factor (Slade, 2000).

The weakness of this approach is its lack of generality. The risks of research projects, for example, are

[3] The risk- adjusted discount rate translates expected values of future cash flows at a specified risk into market values today

[4] "Convenience yield is the flow of services that accrues to the owner of a physical commodity but not to the owner of a contract for future delivery of the commodity " (Brennan and Schwartz, 1985). The role played by convenience yield in real option pricing is similar to the role played by dividends in financial option theory

project-specific and cannot be replicated by securities traded in financial markets (Dixit and Pindyck, 1994; Smith and Nau, 1995). In particular, technical risks (failure of an experiment, ...) are not correlated with any asset traded in the financial market. For example, when valuing an oil property, it is reasonable to assume that price risks can be hedged by trading oil futures contracts, but it seems unreasonable to assume that reservoir specific uncertainties can be hedged by trading securities (Smith and McCardle, 1998).

In other words, options pricing is difficult to apply directly. This explains the result of a recent survey which showed that real options pricing has not found wide usage in practice because managers find it too technical and based on unrealistic assumptions (Loch and Bode-Greuel, 2001).

5.4.1.1 Modeling the value of a real option

The papers by Brennan and Schwartz (1985) and McDonald and Siegel (1986) have set the assumptions that have been the starting point for much of the work on real options.

McDonald and Siegel (1986) assume that the project value obeys a GBM. Consider a derivative asset whose value, F, is a function of the value of a single underlying asset, V, and time, t; $F = F(V, t)$. In real option analysis, F might be the value of an American call option to irreversibly invest in a project that would currently be worth V if installed, or the value of an American put option to abandon a project currently worth V, where V is the expected present value of net income from the project. The goal of real option pricing is to determine the option value F given a stochastic process for V.

A standard approach to valuing these types of options is to assume that the value of the project fluctuates according to an Ito process[5],

$$dV^j = \alpha_V^j(V,t)\,dt + \sigma_V^j(V,t)\,dz, \quad j = I, A, \tag{5.1}$$

where $\alpha_V^j(V,t)$ is the expected instantaneous rate of drift in the value of the project, $\sigma_V^j(V,t)$ is the instantaneous volatility of the value of the project, and dz is the increment of a standard Wiener process. The index j indicates that these drift and volatility functions depend on whether the project underlies the invesment option ($j = I$) or abandonment option ($j = A$). The most popular form of this process is geometric Brownian motion,

$$dV^j = \alpha_V^j V\,dt + \sigma_V^j V\,dz, \tag{5.2}$$

where α_V^j and σ_V^j are constant. Constructing a riskless hedge portfolio (as in Appendix D.2 to value a call), applying Ito's Lemma and the no-arbitrage condition to the function $F(V,t)$ creates the partial differential equation

$$\frac{\partial F}{\partial t} + V\,(r - \delta_V^j)\,\frac{\partial F}{\partial V} + \frac{1}{2}\sigma_V^{2j}\,V^2\,\frac{\partial^2 F}{\partial V^2} + D^j = rF, \tag{5.3}$$

where r is the risk-free rate of interest, D^j (in general, $D^I = 0$ and $D^A > 0$) is the instantaneous net cash flow per unit time paid to the holder of the call or put option, and δ_V^j is the dividend yield (or *rate of return shortfall*), that represents the extent to which the percentage change in project value, α_V^j, falls short of the percentage return required on an investment of this risk class, $\hat{\alpha}_V^j$, $\delta_V^j = \hat{\alpha}_V^j - \alpha_V^j$ (Davis, 1998). The parameter $\hat{\alpha}_V^j$ translates expected values of future cash flows of a specific risk class into current market values. This process is used, for example, by Majd and Pindyck (1987), Paddock *et al.* (1988), Dixit and Pindyck (1994), and Trigeorgis (1996).

[5] see Appendix D.1

Standard numerical techniques, such as tree/ lattice, finite difference or Monte Carlo simulation methods, are used to solve Equation 5.3 for F, taking into account boundary conditions for the specific option being valued. The solution technique requires values for σ_V^j and δ_V^j. These parameters are generally unobservable in the market. In addition, an historical time series of project market values is seldom available as the project is rarely a traded asset or has a *twin* asset. Thus, in most applications *ad hoc* methods are used: σ_V^j is taken to be equal to the average percentage standard deviation of *stock market equity* (McDonald and Siegel, 1986; Majd and Pindyck, 1987; Dixit and Pindyck, 1994); δ_V^j is set equal to the convenience yield associated with the project's output good[6], or to an arbitrary value and the sensitivity of the option calculation to the value used is tested (Majd and Pindyck, 1987), or to zero (Trigeorgis, 1990). Davis (1998) demonstrates that σ_V^j can be estimated from the volatility of the unit price of the project's output, which can be calculated from published historic price series, and formalizes the calculation of δ_V^j linking it to observable financial characteristics of the project.

The approach developed by Brennan and Schwartz (1985) is different. They assume that the spot price follows a GBM. And the project value is a function of the spot price. This because the underlying asset can be hedged in the market. They introduced the convenience yield. For a non dividend paying stock the expected price gain is equal to the appropriate risk- adjusted discount rate. If it were not, investors would not be willing to hold the stock. This explains why the parameter δ_V appears in the valuation models for commodities. Most commodities are stored in some quantity. In order to explain storage of commodities, it is assumed that the storers have an advantage from the storage itself. The results were also found by McDonald and Siegel (1985). They used the Merton's formula instead of the Black and Scholes's for valuing an European call option where the stock pays a dividend yield (see Appendix D.2).

The papers which have been published in relation to the pricing of real options relating to natural resources fall in these two groups.

5.4.2 Stochastic dynamic programming

In the decision analysis approach, risky projects are valued by constructing a decision tree (or dynamic program) that describes the sequence of decisions and uncertainties surrounding the project. Technical (geological) uncertainties can be easily allowed for. The decision maker's beliefs about the project are captured by assessing probabilities for the uncertainties and preferences for the project cash flows are captured by using a risk-adjusted discount rate. Project values and optimal strategies are then determined through stochastic dynamic programming which is a very general tool for dynamic optimization.

The sequential decision problem can be formulated as a stochastic dynamic program with the following value function

$$V_t(i) = \max_{a_t} \left\{ R_t(i, a_t)\Delta t + \frac{1}{1+\rho\Delta t} E[V_{t+\Delta t}|a_t, i] \right\}, \quad \forall i, t$$

where

$V_t(i)$ is the maximum expected value in state i at time t,

[6] If the output of the project is a storable commodity, the convenience yield from storage is defined as the flow of benefits (less storage costs) that the marginal stored unit provides. These benefits can include an increased ability to smooth production, avoid stockouts and facilitate the scheduling of production and sales (Dixit and Pindyck, 1994, p.179). The dividend yield is then the convenience yield

$R_t(i, a_t)$ is the rate of the profit flow obtained by taking action a_t in state i,

ρ is the rate of return per unit of time.

The expectation is conditioned on the current state i and decision a_t. This leads to a decomposition of the optimisation problem into an immediate return function and a continuation value that describes the consequences of all subsequent decisions, starting with the position that results from the immediate decision. This implies that only the immediate control a_t must be chosen optimally at time t, since the remaining optimal strategy is subsumed in the continuation value. It must be noted that to compute $V_t(i)$ the access to the next stage value for all possible future states for each action is needed.

In the case of a finite planning horizon,

$$V_T(i) = \max_{a_T} \{R_T(i, a_T)\}, \ \ \forall i$$

and the very last decision can be found using standard static optimization methods. This solution then provides the valuation function appropriate to the penultimate decision. That in turn serves for the decision two stages from the end, and so on all the way to the initial decision (that is, standard backward recursion; for more details see Appendix E). In fact, in practice a discrete grid of values of t must be chosen to calculate the solution. In the case of an infinite horizon, there is no final value function from which to work backward. Instead the problem gets a recursive structure and brings the previous value function to be independent from t. Of course the current state i matters. This works provided the return R, the discount rate ρ and the transition probabilities are independent of t. The problem is then solved by an iterative procedure (see Appendix E).

Now, assuming, as for CCA, that the project value follows a GBM, $F(V, t)$ is found to vary according to the following equation

$$\frac{\partial F}{\partial t} + V\,\alpha_V^j\,\frac{\partial F}{\partial V} + \frac{1}{2}\sigma_V^{2j}\,V^2\,\frac{\partial^2 F}{\partial V^2} + D^j = \rho F$$

(see Dixit and Pindyck, 1994, p.117). This can be compared with (5.3). In CCA the exogenously specified discount rate ρ is replaced by the risk- free rate r and the growth rate α is replaced by $r - \delta_V^j$. In this case SDP can be used discounting at the riskless rate but assuming that the project value V follows a GBM with growth rate $r - \delta_V^j$. The two approaches give then the same results. Dixit and Pindyck (1994) state that this is an instance of equivalent risk- neutral valuation and the explicit hypothesis of a replicating portfolio is not needed.

The strength of this approach is its generality. It provides a framework, for example, for medical treatment decisions, decisions involving environmental risks and investment decisions. But in investment contexts, it rarely takes into account market opportunities to hedge project risks by trading securities even though these opportunities may have an impact on the project values and on the optimal investment startegy (Smith and McCardle, 1998). Since firms are not assumed to be able to hedge all risk, the risk free rate must be replaced by a rate of return that is adjusted to reflect risk. The drawback of this approach is thus that it does not address the question of the correct risk-adjusted discount rate. It requires an exogenously specified discount rate that reflects the decision-maker's risk attitude. It seems that one of the advantages of option theory over DCF analysis is lost. However, as we have seen, if a replicating portfolio can be defined, the difference between CCA and SDP is more apparent than real.

5.4.3 Comparing CCA and SDP

The two approaches are closely related and are commonly considered in the literature as alternatives. They can be applied to similar problems (see Dixit and Pindyck (1994) for several examples of investment problems solved by both CCA and SDP). As we have seen, the value function of SDP and the asset value in CCA satisfy very similar partial differential equations. The optimality equation of dynamic programming has an interpretation in terms of asset value and the willingness of investors to hold the asset (Dixit and Pindyck, 1994, p.105 and 120). Analogously, equivalent partial differential equations are obtained if, following Brennan and Schwartz (1985), it is the metal price that follows a GBM. In this case if only price is uncertain and commodity futures market exist the two approaches lead to identical results.

As we have seen, stochastic dynamic programming and contingent claims analysis make different assumptions about financial markets, and, in particular, the discount rates used to value future cash flows. No method can claim to be superior for all problems. However, specific problems normally favour one of the methods. CCA should be used where investments are dominated by market priced or public risks, and dynamic programming should be used where investments are dominated by private risks.

Smith and Nau (1995) define full decision tree analysis as traditional decision tree analysis including market opportunities to borrow and trade, and utility functions to capture time and risks preferences. They demonstrate that full decision tree analysis and contingent claims analysis lead to identical results for valuing risky projects when complete markets exist. In addition, they demonstrate that when complete markets do not exist (incomplete or partially complete markets) and option pricing methods can only produce bounds on project value, full decision tree analysis gives a solution between those bounds. The key to this approach is to recognize market opportunities to trade by including them in the decision analysis model. Smith and McCardle (1998) apply this *integrated* approach for oil exploration projects. They use option pricing for risks that can be replicated in the market[7] and dynamic programming for risks that cannot be priced. They insist that both of these risks are important and that management has a great deal of flexibility to adapt as these uncertainties are resolved. Though this extended approach utilises market information to value market risk, the analysis still requires the use of subjective preferences and beliefs.

5.5 Applications using contingent claims analysis

As was said earlier, many applications of real options analyze projects for which the underlying state variable is a commodity price. The reason for this is the existence of forward or future markets for many commodities allows a simple and direct calculation of the equivalent martingale measure. The equivalent martingale measure for commodity prices is obtained in the same way as in the Black- Scholes model for stock prices, except that the convenience yield on the commodity plays the role of the dividend yield on the stock.

Davis (1998) values the option to develop a precious metal reserve. The value of the developed mine is approximated by a GBM with constant drift and volatility. Davis notes that this assumption is unrealistic. Indeed, under a GBM the value of the project cannot become negative, whereas real projects can be worth less than zero. The author studies timing flexibility. He does not consider technical uncertainty. Assuming

[7] It is assumed that the decision maker may buy or sell as many shares of securities as desired at market prices without incurring any transaction costs. Similarly, the decision maker may borrow and lend in any desired amount at a risk-free interest rate r. This is modeled by assuming the existence of a risk-free security whose time t price is given by e^{-rt}

that the price of the project's output good follows a GBM (although recognizing that it is inconsistent with the hypothesis that the project's value follows a GBM), Davis presents a set of equations that can be used to estimate the volatility and dividend yield.

Slade (2000) applies CCA to value managerial flexibility in a copper mine project. She focuses on operating flexibility. The remaining reserves, unit costs and copper price are assumed to follow Ito processes. To model cost and reserve uncertainty that does not arise with respect to price, the author, following Brennan and Schwartz (1982) and Cox *et al.* (1985), subtracts a vector of asset-specific risk premia from the drifts of the three processes and discounts expected future profit at the risk-free rate. The risk premia are obtained from a capital asset pricing model (CAPM).

Moel and Tufano (2000) value the right to develop a copper deposit as a call option. They consider the developed mine as a claim on a traded underlying asset, where the traded underlying is a set of forward contracts on copper and zinc. The mine owner enters into forward contracts to sell the mine output in the future at fixed prices, the price risk is thus eliminated and the relevant discount rate is the risk-free rate (Brennan and Schwartz, 1985). To value the mine they use a Monte Carlo simulation model with a number of simplifying assumptions: only investment timing flexibility is considered (operating flexibility is ignored), the probability distribution of ore quantity is characterized by three discrete outcomes with a known probability of occurring, copper and zinc prices and their convenience yields follow a joint diffusion process (Gibson and Schwartz, 1990; Schwartz, 1997). The authors generate $10,000$ Monte Carlo simulations of price and convenience yield. These values are used to generate the forward prices for copper and zinc in future periods. These forward prices are then used to derive the cash flows in the three DCF scenarios. In the second part of the article, they describe the bidding rules that bidders were required to follow. In particular, each bid was required to specify the minimum amount that the bidder would spend on developing the property if he decided to go ahead. This is equivalent to specifying the exercise price of the development option. Moel and Tufano show that this could induce the successful bidder to make uneconomic investments.

Cortazar *et al.* (2001) develop a real options model for valuing natural resource exploration investments (copper) when there is joint price and geological- technical risk. Geological- technical risk factor can be for example the amount of mineral in a mine. Continuous time Brownian motions are used to model both uncertainty processes. In particular, the risk- neutral price is considered. And possible mine- types and their probabilities of occurrence are defined. Each possible mine is valued as a function of the metal price using the model by Brennan and Schwartz (1985). Finally using the conditional probabilities for each mine type the expected mine value is obtained. The model is relatively simple as it collapses price and geological technical uncertainty into a one- factor model. It is convenient because it allows for representing many geological- technical factors by their joint effect on mine value. Both timing and operating flexibility are valued.

Connell (2002) values an oil project allowing for price and volume uncertainties. Assuming that they are independent both are modelled by a GBM. The reserve distribution changes during exploration. As more seismic surveys are conducted and appraisal wells drilled, information acquired reduces uncertainty about the reserves. Reserve estimates also change as new technological and technical information is acquired. Two perpendicular binomial trees are used to approximate the joint bivariate lognormal distribution. Nodal risk- neutral probabilities are calculated for each independent binomial tree (the reservoir volume is assumed to have zero drift). Thus this model relies upon risk- neutral valuation. However, as the author observes, oil reserves are not traded in the market. Therefore, the creation of a replicating portfolio, as suggested in Trigeorgis (1996), is impossible. Connell (2002) assumes that cashflows are estimated *as if* they were

traded in the market[8].

Dickens and Lohrenz (1996) observe that careful consideration should be given to any strategic decision based on option pricing valuations. They valued an oil and gas asset using the Black-Scholes option pricing model and compared the results with traditional NPV valuations. The bottom line is that option valuation analysis leads to *accept* decisions more often- whether the decision is correct or not. The authors suggest that where the downside risk is limited by commodity prices that cannot drop below zero, the downside uncertainty of a mineral project is almost unlimited. Thus the option pricing techniques will tend to overvalue mineral assets since they ignore a substantial and real aspect of downside risk.

Although not explicitly stated the CCA papers make the replicating- portfolio assumption. However, no empirical evidence regarding the validity of the assumption is presented. A more realistic alternative, although equivalent, is to assume that cashflows are estimated *as if* they were traded in the market.

5.5.1 Applications using dynamic programming or decision trees

Smith and McCardle (1998) apply the integrated procedure developed by Smith and Nau (1995), to value and manage oil properties. In particular, they value termination flexibility. Production rates and oil prices are both uncertain and follow a geometric Brownian motion. The problem is separated into an investment problem that focuses exclusively on the project at hand, and a financing problem that focuses exclusively on securities, specifically how many oil futures contracts and shares of the risk-free security to buy. The decision maker's problem is to choose projects and trade securities so as to maximize his expected utility of consumption. The net cash flows are the sum of the cash flows generated by the oil property and those generated by trading. In order to apply this approach Smith and McCardle place some restrictions on the form of the decision maker's utility function and, as we have seen, on the structure of the securities market. In their case study three assumptions are made about the securities market: the security market is arbitrage-free; the security market is partially complete[9]; the security market is efficient[10]. Using this valuation procedure the decision maker's probabilities for the market uncertainties or his time preferences are not needed. The authors assume that there are two securities available: a risk- free security and an oil futures contract. These two securities are sufficient to give partially complete markets: "any project whose payoffs are a deterministic function of oil prices can be replicated by trading oil futures contracts and the risk- free security. This implies the existence of a unique risk- neutral oil price process" (see Appendix D.2). No assumptions are then needed about its drift rate. The basic idea is to use subjective beliefs and preferences to determine project values conditioned on the occurrence of a particular market state and then use risk-neutral valuation to evaluate these market- state contingent cash flows. Both uncertainties are allowed for as in Connell (2002). However, Connell assumes the market is complete.

[8]This is called *Market Asset Disclaimer*. It is convenient because it is a disclaimer rather than an assumption. For more details see the references in Connell (2002)

[9]The uncertainties in the model can be categorized as either market or private uncertainties. Market uncertainties are risks that can be perfectly hedged by trading securities. For example, if oil prices were the only uncertainty the oil project's cash fbws could be exactly replicated by buying and selling oil futures contracts and shares of the risk- free security (Smith and McCardle, 1998). Private uncertainties are risks that cannot be perfectly hedged by trading. For example, there are no securities whose payoffs are tied to the production rate of a specifi c project

[10]Given the current security prices, the decision maker believes that future prices are independent of the current private information

Lund (1999) considers a stochastic dynamic programming model for project evaluation under uncertainty. Both market risk and reservoir uncertainty are handled by the model, as well as different flexibility types. The types of flexibility presented are: initiation, termination, start/stop, information and capacity flexibility (that is, change the production capacity of the production unit or change the production capacity of the reservoir). A GBM is used to model oil price. Reservoir volume and well rate are assigned a discrete probability for high, average and low scenarios. The author considers the theoretical problems in extending option pricing theory to incorporate market risk which can be hedged, and technical uncertainty which can not be hedged. He argues that "without spanning assets it is not possible to replicate the uncertainty of project perfectly in the market. A risk- free portfolio that contains the project can then not be obtained and there is then no theory for determining the correct value of the discount rate". Having said that, he chooses the discount rate to use by saying that "the major oil companies can be conveniently conceived of as risk neutral, and it is therefore reasonable to apply a risk free rate in the model". Given the size of major oil companies, it is effectively reasonable to consider them to be risk neutral, but this does not necessarily apply to small companies such as junior mining companies. We will come back to this point in Section 7.4.

Loch and Bode-Greuel (2001) value a pharmaceutical R&D project using a decision tree as the major risks of research projects are typically project specific and cannot be replicated in external markets. A decision tree explicitly represents risks and decision points where flexibility is valuable. They recall that a decision tree is equivalent to option pricing for risks that can be priced on the financial markets (if trading of securities is explicitly included), and it can incorporate risks and flexibility that are not traded in financial markets via an explicit utility representation. The pharmaceutical group needed to choose between three research projects. The authors show that recognizing strategic options is the key to correctly evaluating risky projects. Timing and abandonment flexibilities are considered. The company, as a whole, is supposed to be risk neutral toward the projects, but individual managers are often supposed to be risk averse. Risk aversion is included in the analysis using an exponential utility function in all decision nodes with risky payoffs.

Huchzermeier and Loch (2001) consider a dynamic programming approach to evaluate operating flexibility in an R&D investment as it does not require asset replication which often does not apply in R&D projects, because risks are typically uncorrelated with financial markets. A risk-neutral attitude toward the project is assumed and discounting is done at the risk-free rate. The authors introduce the real option of *improving* the project, that is " the capability of an operational midcourse correction during the execution of the project ". This represents an additional source of value. In the model in addition to the market payoff, operational variables such as budget, product performance, market performance requirement and schedule are subject to uncertainty. Huchzermeier and Loch demonstrate that operational uncertainty may reduce the real option value. " If operational uncertainty is resolved before decisions are made and costs or revenues are incurred, flexibility can be applied to protect the project against a downside. In this case more uncertainty enhances the option value of managerial flexibility. However, if operational uncertainty is resolved after decisions are made or reduces the probability that flexibility is useful, more variability reduces the ability to respond and diminishes the option value of flexibility ". They indicate when it is most important to delay commitments.

In the SDP articles the data for the 2 uncertainties, market and technical, is based on subjective assessment.

To value the case study gold deposit we want to account for both price and reserves uncertainty. The SDP approach allows to easily visualize and describe the project highlighting management flexibility in response to price and reserves changes. However, as we have seen, it does not use market information and requires an exogenous discount rate. For this we believe that the approach by Smith and Nau (1995) and Smith and

McCardle (1998) which combines option pricing and dynamic programming must be considered. We will see that the approach is similar, under certain conditions, to that by Connell (2002) and to that by Cortazar *et al.* (2001).

5.6 Information flexibility

The investment in additional information before the development of mine deposits or petroleum reserves is an important alternative to the earlier development and the waiting for better market conditions. If there is no technical uncertainty, the choice is reduced to immediate investment and the wait and see policy.

Kemna (1993) works on a timing option for developing a new lease area in which the company is obliged to drill extra wells. He assumes extra wells do not provide additional information.

Chorn and Carr (1997) state that the purchase of information about a project has considerable value and can be treated as purchasing an option on the project. As with options on equities[11], if the information leads to the expectation of a positive investment outcome, the project should be funded. The authors consider an offshore gas field development and show how option pricing techniques can be used to value information surrounding a production capacity decision. Both gas price and recoverable volume uncertainties are considered in this analysis. Chorn and Carr model the evolution of reserve estimates over time using a stochastic process. The changes of the reserves due to additional information are modelled as a continuous process. As Galli *et al.* (2001) observe, although it may be appropriate for a field in production, it is not during the exploration phase when reserves change after each revision at discrete time intervals. In addition, the authors suppose that more information leads to a reduction of the variance of the estimated reserves. However, although more information leads to a better understanding of the reserves, it does not necessarily reduce the variance.

Chorn and Croft (1998) also consider a petroleum application. The authors compare three scenarios each with increasing levels of characterisation knowledge, and compare the value of those development approaches to one that represents perfect reservoir knowledge. They suppose that the uncertainty represented by well productivity can be approximated with a discrete probability distribution whose standard deviation decreases with increasing characterisation knowledge. The perfect knowledge is represented by all development wells producing at the expected rate. NPV values are then calculated and compared for the different scenarios. By increasing the knowledge level surrounding the reservoir, the uncertainty regarding well productivity and the potential for poor investment outcomes are reduced. The evolution in the probabilities should depend on the results of the new wells.

Lund (1999) considers the possibility to drill additional exploration wells. The author supposes the information received from the wells is binary and either indicates a low volume (the wells do not hit oil) or a high volume (the wells hit oil). This information is used to update the prior prediction of the reservoir volume. *A priori* the volume can take three values. Through Bayes' theorem the posterior probability is then computed. Information is found to have value in particular when the prior variance is large. Lund's results highlight that with a low degree of uncertainty on the volume it is optimal to accept the uncertainty without

[11] Equities are traditional stocks. If you own an equity, then you own stock in a company, and you own part of that company. The term stock is often used to include equities and other products which are traded in a manner similar to equities. The term equities is used to make clear that real stocks are considered

making any adaptations; for a higher uncertainty on the volume it is optimal to accept the uncertainty taking actions to limit any undesirable consequences; as the variances increases it is optimal to try to reduce the uncertainty.

Galli *et al.* (2001) value a satellite close to a large gas field in the North Sea. The development options considered include the possibility to drill one or two extra wells before deciding to go into production. The authors compute the value of the project conditional to the new information (for example, the height of the reservoir at the new well): this information would be available to management at that time.

Martzoukos and Trigeorgis (2001) introduce a model of learning with path- dependency and investigate the optimal timing of actions of information acquisition that result in reduction of uncertainty in order to enhance model option value. They suppose that the underlying asset value, related to the price, can change in unpredictable ways. And this change can be treated as a random variable with a known probability distribution. Learning is modeled as a jump of random size activated by the management, that is it is an event-driven process.

Dias (2002) studies the selection of the best alternative for the investment in information with different costs and different *revelation powers*, that is the expected percentage of variance reduction induced by the new information for the expectation of the uncertain parameter. Whereas Martzoukos and Trigeorgis focus on the timing of learning, Dias focuses on the distribution of expectations (jump size) after the investment in information and the selection of the best project to investment in information. We suppose that the investment in information can be done only once at the start of the project, that is before development. The information will help deciding which of the two projects is more interesting and this must be known at the start of the project. We did not follow the approach by Dias as we are interested in extreme cases for additional information (P10 and P90) and not just on the expected value of additional information. The approach we followed is more similar to that of Lund and Galli *et al.*. Different scenarios for the extra information are defined, then geostatistical simulations or Bayesian prediction are used to update the reserves.

5.6.1 Measure of the value of information

The primary goal of an investment in information is to reduce the uncertainty on the parameters and to reduce the risk of the decision if, as in our case, one of two projects has to be chosen. Note that additional information does not necessarily imply a reduction of the parameter variance. For example, in the test case, with the rich additional information the reserves are more variable. However, as we will see, the risk will be reduced as, even though the uncertainty is larger the information is richer.

We need to define a loss function. Let $\boldsymbol{\theta} \in \Theta$ denote the unknown technical parameters, for example the recoverable reserves. Let $V(\boldsymbol{\theta}, d)$ denote the expected NPV value of taking decision $d \in \mathcal{D}$ for each value of the reserves $\boldsymbol{\theta}$ (that is the expected value is taken with respect to the gold price). A loss function is defined on $\Theta \times \mathcal{D}$. It is convenient to define the following loss function

$$L(\boldsymbol{\theta}, d) = \max_{d'} V(\boldsymbol{\theta}, d') - V(\boldsymbol{\theta}, d)$$

which measures the amount of loss by not taking the optimal decision for every $\boldsymbol{\theta}$, that is it corresponds to the cost of uncertainty when decision d is taken and $\boldsymbol{\theta}$ is observed. It is a positive valued function as $\max_{d'} V(\boldsymbol{\theta}, d') \geq V(\boldsymbol{\theta}, d)$. It quantifies the decision maker's preferences. Let

$$R(d) = E_\theta[L(\boldsymbol{\theta}, d)] = E_\theta[\max_{d'} V(\boldsymbol{\theta}, d')] - E_\theta[V(\boldsymbol{\theta}, d)]$$

be the risk function. The optimal decision d^* is taken to minimize the risk[12], that is $d^* = \arg\min_d R(d)$. The underlying assumptions are

(i) $\sup_d V(\boldsymbol{\theta}, d) < \infty, \forall \boldsymbol{\theta} \in \Theta$;

(ii) $\arg\sup_d V(\boldsymbol{\theta}, d)$ admits a unique solution $\forall \boldsymbol{\theta} \in \Theta$.

Given a decision problem and a loss function L, it is possible to measure the value of an observation $\mathbf{z}_s$ in terms of reduction of risk induced by this information. Suppose that additional information is obtained at $t = 0$. The value of additional information $\mathbf{z}_s$ is given by the difference

$$g(\mathbf{z}_s) = \min_d R(d) - \min_d R(d(\mathbf{z}_s)|\mathbf{z}_s), \tag{5.4}$$

that is the difference between the optimal risk given the initial data and the optimal risk given the initial data + additional information. Observe that $g(\mathbf{z}_s) \geq 0$. This can be compared to the cost of additional information. The difference in (5.4) gives the *value of statistical information* of $\mathbf{z}_s$ (Mouchart in Droesbeke *et al.*; 2002). Note that if the data $\mathbf{z}_s$ yield the same optimal decision as the initial data, its value is zero. This presupposes that if $d = d(\mathbf{z}_s)$ then $R(d) = R(d|\mathbf{z}_s)$. The expected value of this quantity with respect to $\mathbf{z}_s$ can then be computed giving the *expected value of statistical information*, EVIS, of the experiment considered.

The expected gain due to perfect information on $\boldsymbol{\theta}$ is given by $R(d^*)$, which represents the expected cost of uncertainty of the initial distribution. It is called the *expected value of perfect information*, EVIP. The efficiency of an experiment can be measured by

$$\frac{\text{EVIS}}{\text{EVIP}}$$

which takes values in $[0, 1]$. The closer to 1 is this ratio, the higher is the value of information. An experiment with a cost greater than EVIS will never be considered.

Note that in our case the time factor has to be taken into account: the start of the project must be delayed to obtain additional information and the cost (gain) of delaying the project must be included in the model. We will consider the NPV values when optimal responses of management to negative NPV have already been taken into account, that is the NPV with operating flexibility. Thus we have two different risk functions and

$$g(\mathbf{z}_s) = \min_d R(d) - \min_d R_s(d(\mathbf{z}_s)|\mathbf{z}_s). \tag{5.5}$$

The available decisions d are to develop the large pit (U_1) or the small pit (U_2). Note that even if the optimal decision is the same with the new information as without it, the value of the information is not necessarily zero because $R \neq R_s$ (i.e. it may still reduce the risk). The difference in (5.5) can be positive or negative. We will call the value of information or *information flexibility* the reduction of risk induced by additional information. It will be equal to zero if it is optimal to develop the project immediately or without extra information. The expected value of (5.5), with respect to $\mathbf{z}_s$, can then be computed giving the expected value of information.

The difference in expected project values with and without additional information, that is

$$E_\theta[V_s(\boldsymbol{\theta}, d(\mathbf{z}_s))] - E_\theta[V(\boldsymbol{\theta}, d)],$$

[12]The minimisation of the expected loss is equivalent to the maximisation of V

gives a measure of the value of additional information. The two values of information differ as in (5.5) the risk of taking the wrong decision is considered not insisting only on the value of the project. On the contrary, the difference in project values highlights the fact that additional information permits to better define the project. The values of information will differ, but the conclusions, that is whether it is worthwhile or not to get the additional data, do not differ. Both measures will be considered.

The dynamic real option model could include a penalty factor for the lack of information, which causes sub-optimal development.

5.7 Application to the case study

The objective is to evaluate the mining project taking into consideration management flexibility to react to changes in the metal price and in the reserves. Thus two stochastic variables are considered: the gold price that is supposed to follow a GBM and the reserves that were obtained by Monte Carlo methods in Chapters 2 and 4. We suppose that the uncertainty regarding the reserves does not evolve over time. Additional drilling is obtained at the start of the project and the reserves distribution is then updated. Once the reserves distribution is fixed the approach is equivalent to the one used by Cortazar *et al.* (2001). If the uncertainty on the reserves was assumed to evolve with time the approach by Connell (2002) could be used. However, the author assumes that this uncertainty follows a GBM. As Connell observes it is a simplifying hypothesis: the variance of the GBM increases with time! An alternative model should be considered.

The exploration of the mine is supposed over. The economic assessment is made before tax.

In particular, as the exploration phase is over, additional information to reduce the uncertainty surrounding the reserves could be obtained either by additional drilling or by production. For simplicity and as we are interested in the value of extra holes, we suppose that no information is obtained from the production phase. To value information flexibility, that is to decide if it is worthwhile to carry out new drilling we must be able to quantify the value of this new information. In Chapter 2 simulations were carried out to define different possible scenarios. The additional information was supposed to be of type rich, average or poor. In Chapters 2 and 4 conditional simulations of the reserves were computed either with the *plug- in* or Bayesian approach. This gives possible mine- types and their probabilities of occurrence. If additional information is obtained these probabilities may be modified. In Chapters 2 and 4 conditional simulations of the reserves given additional drilling were computed either with the *plug- in* or Bayesian approach. Additional information may be obtained at the start of the project. Once the additional information is obtained, the uncertainty regarding the reserves does not evolve over time. Finally using the probabilities for each mine- type the expected mine value is obtained. The risk- neutral price process is used and combined with the subjective probabilities for the technical or private uncertainties.

Stochastic dynamic programming is considered to value the project. The model is an optimisation model consisting of discrete stages, where a sequential decision is made in a stochastic environment. The operator wants to maximise the expected net present value of the development project, and the model provides decision support by identifying the optimal development strategy. As we have seen SDP can be equivalent to CCA when the same hypotheses regarding the discounting are made.

5.7.1 The model

The model is a Markov decision process: the transition probability from the current state of the process to the next state only depends upon the current state. A Markov decision process is an optimisation model described by its state space, the stages, the action space, the transition probabilities and the reward function.

The state identifies the status of the system at each stage. At each stage the system considered is subject to control. The actions taken at one stage become effective at the next stage. The length of the decision epoch represents a time lag for the consequence of an action to materialise. All costs and incomes are related to the start of the decision epochs and are discounted at the risk-free rate. The transition from one state to the next is determined by the action taken and the transition probabilities. By taking an action a reward is accrued, and the objective is to optimise the discounted sum of all rewards. The rewards can be both positive and negative, corresponding to sales income and production costs.

Once the mine life is chosen, the problem has a finite horizon. Seven years is assumed to be the mine life. In particular, if the large pit is developed the life time is of 7 years, but if only the small pit is developed the life time is of 5 years. The orebody must be exploited within 8 years. Eight decision epochs are considered.

We now describe the state variables, and in particular the GBM, and its discrete approximation, for the price, the stages and decision variables, the transition probabilities and the equation of optimality.

State variables

The state variables that identify the status of the development project at each stage, are the recoverable reserves, the production per year, the gold price and the costs. They capture all relevant information available to the operator as background for his decisions.

Reserves. The reserves are defined by the recovered average grade and the tonnage of ore. The Bayesian results are used. The distribution and the updated distribution of the recovered average grade and the tonnage of ore were given in the previous chapter in Tables 4.2 and 4.3, respectively. Let B denote the reserves and W the tonnage of waste.

Blocks are exploited without any order, that is the selection of blocks above cutoff is free. It would have been more interesting or realistic to suppose that high grade blocks are exploited first or that blocks are exploited in order to reach high grade blocks first.

Production. For simplicity we assume that the production of ore per year is constant. It equals the total quantity of ore divided by the mine life.

Gold price. Two commonly assumed stochastic price processes are

- geometric Brownian motion and
- mean reverting.

Geometric Brownian motion is the most widely used model of price behavior. This model gives prices that fluctuate randomly around an exponential trend. Mean reversion is the term used to describe the fact that

commodity prices in general tend to oscillate around an estimated long- run, or mean, price[13]. Supply and demand responses to deviations from the long run price provide the mechanism by which mean reversion is likely to continue[14]. This is the case for basic commodities such as copper and oil (Slade, 2000; McCarthy and Monkhouse, 2003) but not for gold. For this we selected the **geometric Brownian motion**. It usually requires many years of data to determine with any degree of confidence whether a variable is indeed mean reverting. According to Dias (1997), it is difficult to statistically distinguish between a random walk and a mean-reverting process for a price horizon of up to 30 years of data. As both the geometric Brownian motion and the mean reverting process are acceptable it would be of interest to see how the choice of the stochastic process affects the value of flexibility. However, it must be noted that for 5 or 7 years, that is the time life of the small and large pit, respectively, the monthly price series (Figure 5.6) cannot be said to be stationary. This can be seen in Figure 5.7 which presents the variogram for the monthly price series.

We assume that metal is sold at the spot price. This is modelled as a geometric Brownian motion

$$dS_t = \alpha S_t dt + \sigma S_t dz_t$$

where dz_t is the increment of a standard Wiener process (of mean 0 and variance t), α is the expected return and σ is the standard deviation of the instantaneous, annualised rate of return. The price S_t is made of two components. First, there is an exponential trend that grows with a $\alpha\%$ rate. Second, there are random fluctuations around this trend. These variations increase over time because of higher prices. We can observe that $E[dS_t] = \alpha dt$ and $\text{Var}[dS_t] = \sigma^2 S_t^2 dt$, that is S_t is not stationary. Using Ito's Lemma the explicit formula for S_t is

$$S_t = S_0\, e^{(\alpha - \sigma^2/2)t + \sigma z_t},$$

where S_0 is assumed known (see Appendix D for more details). As the geometric Brownian motion is a continuous time process, its continuous time formulation must be transformed into a discrete formulation, before it can be applied in the SDP model.

Approximation of the geometric Brownian motion. We use the binomial model first proposed by Cox, Ross and Rubinstein (1979) to approximate the geometric Brownian model. Let S be the initial gold price. The assumptions of the binomial model are

(A1) the price over each period of time Δt can either move up to uS or go down to dS with $0 < d < u$.

(A2) The probability of a movement up or down is p and $1 - p$, respectively (Figure 5.1).

(A3) The expected return is

$$E[S_{t+\Delta t}] = S_t\, e^{\alpha \Delta t}. \tag{5.6}$$

[13]In other words, price increases tend to be followed by price declines as the price of the commodity reverts to its long- run level. This reversion to a long- run price is observed in historical price data and, based on futures prices, it is expected to continue in the future

[14]The key question for most commodities is the strength of mean reversion (see Appendix D for more details on mean reverting processes). The signifi cance of mean reversion in commodity prices is that it implies a saturation of price risk. If the commodity price risk is a signifi cant contributing factor to a company's systematic risk, the discount rate for the project should decline over time. But at lower levels of mean reversion, that is the expected deviations from the mean become greater (in the extreme case it becomes a GBM), the decline in the discount rate over time will be less pronounced

The parameters u, d and p are determined so that the mean and variance of the discrete binomial model are consistent in the limit with their continuous counterparts (Trigeorgis, 1996)[15]. In addition, an arbitrary third equation is imposed

$$u\,d = 1 \tag{5.8}$$

which reflects a symmetry between upward and downward movement of the price. The parameters u and d are given by (up to terms of higher order)

$$u = e^{\sigma\sqrt{\Delta t}} \quad \text{and} \quad d = \frac{1}{u} = e^{-\sigma\sqrt{\Delta t}}$$

and p is

$$p = \frac{e^{\alpha\Delta t} - d}{u - d}. \tag{5.9}$$

The symmetry of (5.8) is apparent in that after two time steps the value S_0 repeats (Figure 5.13). It can also be noted that the binomial approach defined by (A1) reflects exponential growth or decay of S. In addition when using this model it is implicitly assumed that u and d are constant during the period T.

One objection to this discrete binomial valuation approach is that in reality prices may take on more than just two possible values at the end of a given period, since actual trading in the market takes place almost continuously. However the length of a period can be chosen to be arbitrarily small by successive subdivisions. And as the time interval, Δt, approaches zero, the approximation approaches the geometric Brownian motion.

[15] A consequence of assumptions (A1) and (A2) is

$$E[S_{t+\Delta t}] = p\,S_t u + (1-p)\,S_t d$$

Equating with (5.6) gives

$$e^{\alpha\Delta t} = p\,u + (1-p)\,d$$

and

$$p = \frac{e^{\alpha\Delta t} - d}{u - d}.$$

To be a valid model of probability, $0 \le p \le 1$ must hold. Thus $d \le e^{\alpha\Delta t} \le u$.

From the continuous model we have

$$E[S^2_{t+\Delta t}] = S_t^2\, e^{(2\alpha+\sigma^2)\Delta t}. \tag{5.7}$$

Equations (5.6) and (5.7) combine to

$$\text{Var}[S_{t+\Delta t}] = S_t^2\, e^{2\alpha\Delta t}\,(e^{\sigma^2\Delta t} - 1).$$

The discrete model satisfies

$$\begin{aligned}\text{Var}[S_{t+\Delta t}] &= E[S^2_{t+\Delta t}] - (E[S_{t+\Delta t}])^2 \\ &= p\,(S_t u)^2 + (1-p)\,(S_t d)^2 - S_t^2\,(pu + (1-p)d)^2.\end{aligned}$$

Equating variances leads to

$$e^{2\alpha\Delta t}\,(e^{\sigma^2\Delta t} - 1) = pu^2 + (1-p)d^2 - (e^{\alpha\Delta t})^2,$$

that is

$$e^{2\alpha\Delta t + \sigma^2\Delta t} = pu^2 + (1-p)d^2$$

The size of the binomial model is reduced by using a recombining tree, that is an up movement followed by a down movement leads to the same price as a down movement followed by an up movement (for $u\,d = 1$). At expiration time $T = N\,\Delta t$, the price S_T can only take $N + 1$ discrete values

$$S_0\,u^{N-j}\,d^j, \quad j = 0, 1, \ldots, N$$

with probability

$$C(N, j)\,p^{N-j}\,(1-p)^j, \quad j = 0, 1, \ldots, N$$

where $C(N, j)$ indicates the combinations of size j from N. The random variable S_T follows a binomial distribution. This considerably reduces the number of nodes on the tree. There are only $N\,(N+1)/2$ nodes in this tree, whereas there are $2^{N+1} - 1$ nodes in the non recombining tree.

This approximation should be applied with some care since one of the probabilities could become negative for some combinations of α, σ and Δt, that is if $|\alpha\sqrt{\Delta t}| > \sigma$. For practical purposes the possibility of negative probabilities is not regarded as a major problem, since the ratio of the drift rate to the volatility rate is normally small.

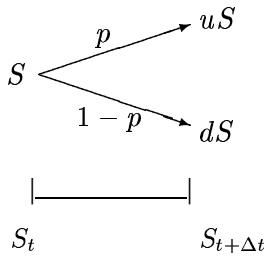


Figure 5.1: *Binomial price process: either the price will go to uS with probability p or it will go to dS with probability $(1-p)$*

Costs. Costs consists of fixed operating costs which are independent of the production level and of variable costs which are associated with extracting and processing ore and with extracting waste. These latter costs are given per produced ounce and per extracted ton of waste. The cutoff grade z_c is chosen so that the quantity of metal recovered from a block of grade z_c pays for its marginal mining and processing costs. We add a variable cost to extract waste while letting z_c include extraction, marginal mining and processing costs for blocks above cutoff. In addition, at the start of the project an initial investment must be made. We include this investment in the fixed operating costs. Costs are supposed known.

Decision epochs and decision variables

The development project is seen as a sequential problem with a finite number of phases. The project is divided into stages to reflect the sequential pattern of decisions and the uncertainty characteristics at different levels. Decisions are made annually[16].

[16]This is different from the time step of 1 month that will be used to approximate the GBM in SDP

The flexibility types considered are: initiation, termination and information. The available initial decisions are then

- abandon,
- develop only the small pit,
- develop the large pit immediately,
- wait one year before taking a decision,
- wait one year and carry out additional drilling before taking a decision,
- start to develop the large pit but with the possibility to go back to the small pit.

Once the project is started, the options available to management are

- abandon,
- continue the development.

We are particularly interested in the value of additional drilling. It will be given by the optimal value of the entire tree with the possibility of drilling the extra holes minus the optimal value of the entire tree without the possibility to drill the extra holes. We do not consider capacity flexibility. We define a constant production rate. We do not take into consideration the start/stop option either. It is questionable whether a mine can be stopped and started to follow commodity price fluctuations. Exceptions are, as Armstrong and Galli (1997) note, low cost dredging operations and low cost quarries and mines extracting industrial minerals. But capital intensive open pits and most underground mines are not flexible in stopping and starting operations.

The decision variables describe the operator's possible reactions to a changing environment. The action space at stage n consists of the following decision variables

$$I_I(n) = \begin{cases} 1 & \text{if additional information is obtained in period } n \\ 0 & \text{else} \end{cases} \quad \forall n$$

$$I_P(n) = \begin{cases} 1 & \text{if ore is produced in period } n \\ 0 & \text{else} \end{cases} \quad \forall n$$

$$I_A(n) = \begin{cases} 1 & \text{if the project is abandoned in period } n \\ 0 & \text{else} \end{cases} \quad \forall n$$

and the action vector, $a(n)$, is then

$$a(n) = \{I_I(n), I_P(n), I_A(n)\}.$$

One additional variable must be defined

$$\delta_n = \begin{cases} 0 & \text{if the project is abandoned} \\ 1 & \text{else} \end{cases} \quad \forall n$$

This defines the status of the development project at stage n.

Transition probabilities

The transition probabilities of going to one state to another are

$p_S(p,q)$: element (p,q) in the transition probability matrix for the gold price between two stages, that is the probability that price at stage $n+1$ is equal to q given that price at stage n is p;

$p_{z_I}(l,k)$: element (l,k) in the transition probability matrix for the additional information between two stages, that is the probability that additional information $\mathbf{z}_I$ at stage $n+1$ is k given that information $\mathbf{z}$ at stage n is l. In particular, we suppose that the additional information is obtained at the start of the project ($t=0$). We then need to define

$$p_{z_I} = \begin{pmatrix} p(\mathbf{z}_I = \text{rich}|\mathbf{z}) \\ p(\mathbf{z}_I = \text{average}|\mathbf{z}) \\ p(\mathbf{z}_I = \text{poor}|\mathbf{z}) \end{pmatrix}$$

that is the probability of obtaining the additional information $\mathbf{z}_I$ given the initial data $\mathbf{z}$. The distribution of the reserves must be updated given the new information. If no additional information is obtained $p_{z_I} = 1$.

Note that the transition probability for the additional information depends on the action taken by the operator. Finally the total transition probability, $p_{ij}(a_n)$, between two stages is simply given by

- for $t=0$: $p_{ij}(a_0) = p_S(p,q) \cdot p_{z_I}$;
- for $t>0$: $p_{ij}(a_n) = p_S(p,q)$.

We will assume that

$$p_{z_I} = p(\mathbf{z}_I|\mathbf{z}) = \begin{cases} 0.25 & \text{if } \mathbf{z}_I = \text{rich} \\ 0.50 & \text{if } \mathbf{z}_I = \text{average} \\ 0.25 & \text{if } \mathbf{z}_I = \text{poor} \end{cases}$$

which is an approximation as the rich, average and poor information were the P90, P50 and P10 scenarios, respectively. As we could not work with the 100 simulated data sets, we chose 3 to represent a rich, an average and a poor scenario and we supposed they represented respectively 25 %, 50 % and 25 % of the possible extra data sets.

The fundamental equation of optimality

Let the parameters used in the model be

T time horizon (in years) for the development project,

S_0 initial gold price,

r risk-free rate,

z_c cutoff grade,

VC variable operating costs per unit of waste extracted,

FC fixed operating costs,

AC abandonment costs (early abandonment),

IC cost of additional information.

The fundamental equation of optimality is then

$$\begin{aligned} V_n(i) &= \max_{a_n}\left\{R_n(i,a_n)+\frac{1}{(1+r)}E[V_{n+1}|a_n,i]\right\} \\ &= \max_{a_n}\left\{R_n(i,a_n)+\frac{1}{(1+r)}\sum_j p_{ij}(a_n)\,V_{n+1}(j)\right\} \quad \forall i, n=1,\ldots,N-1. \end{aligned}$$

where

$R_n(i,a_n)$ is the immediate return from taking action a_n in state i at stage n,

$V_n(i)$ is the maximum expected value in state i at stage n.

For the final stage ($n=T$) we have

$$V_T(i)=\max_{a_T}\{R_T(i,a_T)\}.$$

The immediate return function, R, is the net income at the present stage for a given state and action. The return function for an arbitrary stage n and state is

$$R_n(a_n)=\delta_n\left[I_P(n)\,B_n\,S_n-I_P(n)\,W_n\cdot VC-I_I(n)\,IC-I_A(n)\,AC\right]$$

where B_n and W_n denote the annual reserves and waste, respectively. Fixed costs are assigned at the start of the project. If the project is delayed half of the fixed costs will be assigned at $t=0$ and half at $t=1$. The cost of additional drilling is assigned at $t=0$. That is $I_I(n)$ can be equal to 1 just at $t=0$.

Now, to make the model consistent, decisions taken at a stage must be compatible with previous decisions and with the realised stochastic values. For this the following constraint is imposed

- $\delta_n=1-\sum_{j=1}^{n-1}I_A(j),\ \ n=1,\ldots,T$ (if the project was abandoned in a previous period, the project status is zero in the present and all subsequent periods).

This constraint means that once abandoned the project can not be started again.

5.7.2 Financial parameters

In this section we define the parameters for the gold price process, the riskfree rate and the constants used in the model. Then we present the results for the development of just the lower part of the large pit as its reserves are highly uncertain. The initiation and termination options are evaluated. The information flexibility is considered and its value is highlighted. Finally, in Section 7.4, we compare the results for the small pit and the large one, and determine the optimal strategy.

Gold price parameters. The present price of gold, S_0, is assumed to be $ 320 per ounce. It was the current price of gold when we started this project as it can be seen in Figure 5.8. As the price is modelled using a

geometric Brownian motion we require estimates of the drift rate and volatility. These are estimated from historical data, via the logs. Over a finite time interval Δt, the change in the logarithm of S is normally distributed[17] with mean $(\alpha - \frac{1}{2}\sigma^2)\Delta t$ and variance $\sigma^2 \Delta t$. Let S_i denote the gold price at time i and suppose we have $n+1$ observations. Given the data series

$$U_i = F_{i+1} - F_i = \log(S_{i+1}) - \log(S_i), \quad i = 1, \ldots, n$$

its sample mean, $\overline{U}$, and variance, V^2, are computed. These statistics are estimates of $(\alpha - \frac{1}{2}\sigma^2)\Delta t$ and $\sigma^2 \Delta t$, respectively. Thus

$$\hat{\alpha} = \frac{\overline{U} + V^2/2}{\Delta t} \quad \text{and} \quad \hat{\sigma} = \frac{V}{\sqrt{\Delta t}}.$$

It is not easy to choose an appropriate value for n. More data lead to more accuracy; but σ changes over time and data that are too old may not be relevant for predicting the future. This is confirmed by Figure 5.6 which shows the monthly average price of gold in US dollars per ounce from 1968 to 2002. The question is to choose a period that is long enough to be statistically meaningful but short enough to be relevant.

Hull (1993) suggests to use closing prices from data over the most recent 90 to 180 trading days[18]. Figure 5.8 shows daily gold prices in US dollars per ounce for the period from 2001 to 2002. The last 180 data were used to estimate these parameters. This gave $\alpha = 0.11$ and $\sigma = 12.52\%$. Figure 5.9 shows the daily series of gold price in bold type together with the estimated trend. In addition 10 simulated series are shown (dotted lines). Figure 5.10 presents the estimated trend for the same model for 10 years. As can be seen the drift rate is unrealistically high.

As the minimum life time of the deposit is 5 years, we estimated the model parameters using the data of the last 5 years of the monthly series. The parameters are $\alpha = 0.02$ and $\sigma = 11.41\%$. Figure 5.11 presents the estimated trend and 10 realisations of this process. Figure 5.12 presents the estimated trend for the same model for 10 years and 10 realisations of this process. Is it reasonable to take a constant volatility of 11.41%? The implied volatility[19] of gold was around 13 % at the end of 2001 (source: http://www.pmpublishing.com/volatility/gc.html).

For the binomial model to give a good approximation the period Δt must be chosen sufficiently small. Comparing Figures 5.13 and 5.14 the importance of the choice of Δt in the approximation emerges. By definition, the mean of S_t obtained with a discrete approximation, for any Δt, is the same as the mean of the GBM. On the contrary, as it can be seen in Table 5.2, the standard deviations differ. Increasingly smaller

[17] Let $F(S) = \log(S)$. Since $\partial F/\partial t = 0$, $\partial F/\partial S = 1/S$ and $\partial^2 F/\partial S^2 = -1/S^2$, we have using Ito's Lemma (see Equation D.4)

$$\begin{aligned} dF &= \frac{1}{S}\, dS - \frac{1}{2S^2}\,(dS)^2 \\ &= \alpha\, dt + \sigma\, dz - \frac{1}{2}\sigma^2\, dt = (\alpha - \frac{1}{2}\sigma^2)\, dt + \sigma\, dz \end{aligned}$$

[18] There are 250 trading days per year

[19] Implied volatility is the current volatility of a stock, as estimated by its option price. If the price of an option is known, then the implied volatility can be deduced from it. Because there are many options on a stock, each option can have a different implied volatility. It is then necessary to calculate a representative implied volatility for a stock. This is an average of the implied volatilities of the different options on that stock (usually at the money options for the next few expirations are used). Note that implied volatility is the only not directly observable parameter in option pricing

values of Δt were considered and the respective parameters and project values were compared. We finally set Δt equal to $1/12$ year. This gives $u = 1.03$ and $p = 0.51$.

	$1/4$ year	$1/12$ year	$1/144$ year	**GBM**
$T = 1$ year	36.85	36.93	36.96	**37.00**
$T = 2$ years	52.82	52.92	52.97	**53.02**
$T = 3$ years	65.56	65.68	65.74	**65.81**

Table 5.2: *Standard deviations of the price for the binomial approximation for different Δt with $S_0 = \$\,320$, $\alpha = 0.02$, $\sigma = 11.41\%$. In bold type the theoretical values of the GBM are given. It is obvious that as Δt goes to 0 the approximation gets better*

The risk- free rate. The risk free rate is measured by the 6-month LIBOR (London Interbank Offer Rate) which has fluctuated over the last 10 years between 1 % and 7 % (Figure 5.15). In the model an annual rate of return of 2 % is used. It was the value of the 6-month LIBOR when we started this project.

Note that using the risk- neutral valuation we do not need to specify the drift rate α for the gold price process. This is replaced by r[20]. This gives the risk- neutral probability $p = 0.51$ (see Appendix D.2).

Parameter values. The parameters used in the model are

$T = 8$ years,

$S_0 = \$\,320$ per ounce of gold,

$r = 0.02$,

$z_c = 1$ g/t,

$VC_1 = \$\,5.5$ and $VC_2 = \$\,4.5$ per ton,

$FC = \$\,100$ million,

$AC = \$\,5$ million,

$IC = \$\,10$ million.

5.7.3 Results for the lower pit

Two years is assumed to be the life of the lower pit, that is the lower part of the large pit. We assume that the pit must be developed in at maximum 3 years. For such a short period of time the uncertainty on the price is small (Figure 5.12). This allows to highlight the importance of accounting for the uncertainty on the reserves.

First the uncertainty on gold price is considered while the reserves are assumed known, that is they are fixed at their expected value. Then both the price and reserves are supposed unknown.

[20] In particular it is replaced by $r - \delta$, where δ is the convenience yield for gold. The convenience yield for gold is approximately zero (Casassus and Dufresne, 2001)

Uncertainty on gold prices

The only source of operating flexibility included in the model is the option to abandon the project (termination flexibility). The initial decisions we defined are

- abandon the project,
- develop the pit immediately,
- wait for one year (to resolve price uncertainty) before to develop the pit.

Different versions of the model have been considered to illustrate the importance of including flexibility in the evaluation of the project. The first one is a deterministic version of the model: the gold price is replaced by its initial value and is considered constant. In the second version of the model the stochastic representation of the price is considered, but no flexibility is included. The deterministic model is expected to give a lower project value than the stochastic model as the drift rate is positive. The third version of the model allows for flexibility, that is taking actions, during the project. It is expected to give an equal or higher value than the stochastic model. The results obtained are presented in Table 5.3. The project is evaluated for the different initial decisions. Note that if the project is abandoned at $t = 0$, its value is 0.

The optimal initial decision is *develop the pit immediately* for the deterministic model. Whereas the optimal initial decision is *wait* for the stochastic models. Thus the deterministic model leads to a non optimal initial decision. If the initial decision is based on the deterministic model, we incurr in an expected loss of $ 7 (182 - 175) million. The option to wait one year has value: $ 7 million. This corresponds to an increase in project value of 4.00 %. Comparing the stochastic model with no flexibility and the model with flexibility, it can be noted that operating (termination) flexibility has in this case no value.

Project evaluation model	**Expected value (NPV)**
Initial decision: start to develop the pit immediately	
deterministic model (S_0)	169
stochastic model: no flexibility	175
stochastic model: flexibility	175
Initial decision: wait	
deterministic model (S_0)	164
stochastic model: no flexibility	182
stochastic model: flexibility	**182**

Table 5.3: *Expected value of the project in $ million ($r = 0.02$, $\sigma = 11.41\%$)*

These results can be compared to the results presented in Davis (1996) who studied the flexibility values found in the literature for mineral and oil projects: operating flexibility has little or no value, while initiation flexibility can amount to a significant increase in the NPV.

Uncertainty on gold prices and reserves

In addition to the previous initial decisions there is the possibility to

- wait for one year to carry out additional drilling before to develop the pit.

The first decision node of the tree is presented in Figure 5.2. As it was said previously, once the project is started every year we can decide whether to continue the development or abandon the project. If we wait 1 year at $t = 2$ we can decide whether to start the project or abandon it. This second decision node is also presented in Figure 5.2.

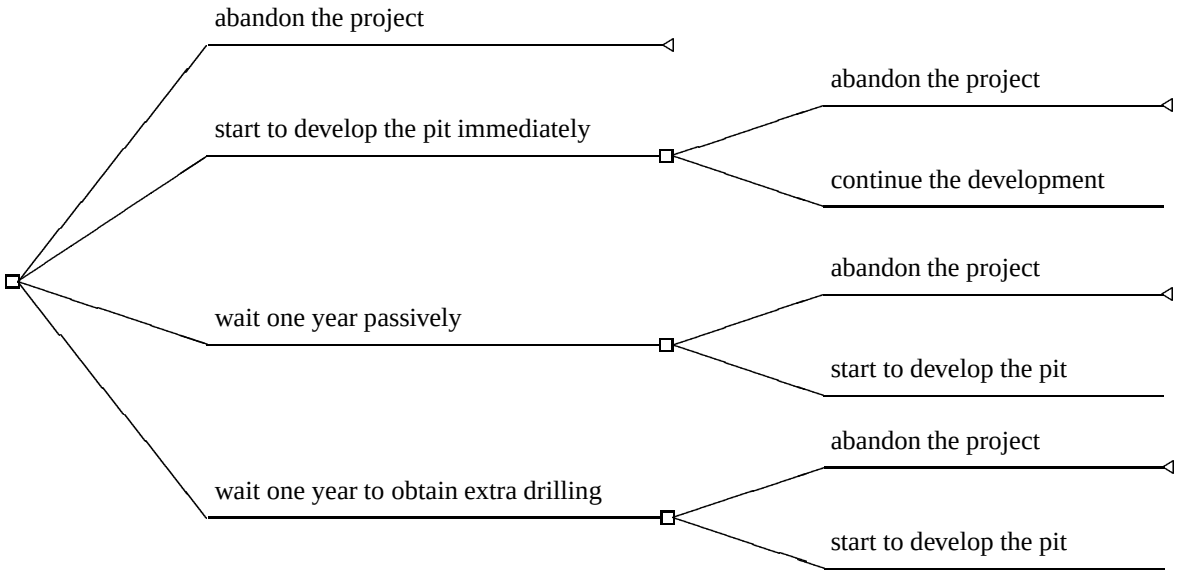


Figure 5.2: *First and second decision nodes*

The value of information will be given by the optimal value of the entire tree with the possibility of drilling the extra holes minus the optimal value of the entire tree without the possibility to drill the extra holes.

As previously, different versions of the model have been considered to illustrate the importance of including flexibility in the evaluation of the project. The first one is the deterministic version of the model and the uncertain variables are replaced by their expected value. In the second version of the model the stochastic representation of the two uncertain variables is considered, but no flexibility is included. The third version of the model allows for flexibility, that is taking actions, during the project but it does not consider the possibility of getting new information. And finally version four is defined so that management has the flexibility to make adjustments as the project proceeds including the possibility to obtain more information. The results obtained are presented in Tables 5.4 and 5.5.

As it can be seen in Table 5.4 the deterministic model gives a lower project value, as we expected, than the stochastic model. Comparing the initial decisions *abandon the project*, *start to develop the pit immediately*

Project evaluation model	**Expected value (NPV)**
Initial decision: start to develop the pit immediately	
deterministic model (S_0)	169
deterministic model ($E[S_t]$)	175
stochastic model: no flexibility	180
stochastic model: flexibility	185
Initial decision: wait passively	
deterministic model (S_0)	164
deterministic model ($E[S_t]$)	181
stochastic model: no flexibility	187
stochastic model: flexibility	**192**

Table 5.4: *Expected value of the project in* $ *million for the initial decisions* start to develop the pit immediately *and* wait passively ($r = 0.02$, $\sigma = 11.41\%$)

and *wait* for the 3 models, the optimal initial decision is *start to develop the pit immediately* for the deterministic model with S_0, that is the current gold price. Whereas it is *wait* for the deterministic model with $E[S_t]$ and the two stochastic models. Initiation flexibility has value: $ 7 (192-185) million. This corresponds to an increase in project value of 3.78 %. By including flexibility- without the possibility to get additional information- the expected present value is higher and thus the value of flexibility is in this case greater than zero. The value of operating flexibility (termination flexibility) is $ 5 (192-187) million. This corresponds to an increase in project value of 2.67%.

Introducing the uncertainty on the reserves increases the value of flexibility. Note that as the gold price has a positive drift it is the uncertainty on the reserves that mostly gives value to flexibility.

Project evaluation model	**Expected value (NPV)**
Initial decision: wait and get additional information	
Additional information: rich	
stochastic model: no flexibility	328
stochastic model: flexibility	328
Additional information: average	
stochastic model: no flexibility	222
stochastic model: flexibility	224
Additional information: poor	
stochastic model: no flexibility	126
stochastic model: flexibility	133

Table 5.5: *Expected value of the project in* $ *million for the initial decision* wait 1 year and get additional information ($r = 0.02$, $\sigma = 11.41\%$)

Additional information

Table 5.5 presents the value of the project given additional information. This value depends on whether the new information turns out to be richer, average or poorer than expected. The value of operational flexibility is the highest for the poor scenario as the probability of poor reserves is in this case the highest (Figure 4.19). It is $ 7 (133-126) million, that is an increase in project value of 5.56 %. It should be noted that although the additional information reveals that the project is poorer than anticipated, it is still worthwhile to develop the pit because the expected value of this project is greater than 0. Thus it remains optimal to *develop the deposit*. But is it better to develop it immediately or is it better to wait to get additional information on the reserves and then develop it? It would be worthwhile to obtain this information if it helps better defining the project and if it changes, for example, the optimal initial decision leading it to *abandon the project*. In our case it is always optimal to develop the deposit.

Now, suppose the poor additional information is obtained. The expected value of the project would then be $ 133 million, whereas the expected value of the project without this information is $ 192 million. Knowing before developing the pit that the reserves are poor is perhaps better than finding it out when the development of the deposit has started. Figure 5.16 compares the histograms of the project values with flexibility given the initial data and given the initial data + poor information. In this last case, the uncertainty on the project value is smaller. The probability of a high value (> $ 400 million) is zero, whereas the probability of a low project value is higher. This information would allow management to better define the project.

Suppose that the rich, average and poor scenarios have a probability of occurring of 0.25, 0.50 and 0.25, respectively. The expected NPV value with respect to the additional information is then: $\$\ 0.25 \times 328 + 0.5 \times 224 + 0.25 \times 133 = 227.25$ million, which is higher than the expected NPV value without additional information. The expected value of additional information is $ 35.25 million, that is additional information increases the project value of 18.36 %. Thus it is worthwhile to obtain it. The optimal development strategy is then *wait, obtain additional information and then develop the pit*.

An alternative approach to value information could be to consider the changes in flexibility due to extra information. Observe that the value of flexibility can be, at one extreme, zero if the additional information is rich and, at the other, it can be quite high if the additional information is poor. In these cases it seems worthwhile to obtain the additional data. To compute the value of additional information in helping decide whether a project must be carried on or not, we consider the expected change in flexibility values. This implies that an information is worthwhile to be obtained if it greatly increases or diminishes the flexibility value. Note that the approach presented in Section 6.6 cannot be considered when deciding whether a project must be carried on or not. The risk value is zero as operating flexibility eliminates the possibility of a negative NPV.

The value of flexibility is $ 5 million given the initial data and it is $ 0, 2 and 7 million given the rich, average and poor scenarios, respectively. The expected value of the change in flexibility is $\$\ 5 - 0.25 \times 0 - 0.5 \times 2 - 0.25 \times 7 = 2.25$ million. It is reduced by 45 %. Thus it is worthwhile to wait and get additional information.

Considering the changes in flexibility is not a valid approach to measure the value of information, because it does not take into account the value of the project and the fact that additional information helps better defining the project. In this case it overestimates the value of information.

Note that only the expected values are considered and not the dispersion of the project values.

Another potential gain from additional information that we did not take in consideration is that it helps in exploiting optimally the reserves. As Dias (2002) observes, this is particularly important if the additional information turns out to be rich. For this a penalty factor could have been introduced in the model.

Sensitivity to the initial gold price

A sensitivity study was carried out to measure the impact of the initial gold price on the project value and the value of flexibility. The initial price is altered over the range from $ 100 to 320. Figure 5.17 shows the value of initiation flexibility. The optimal initial decision is *wait* and thus the value of initiation flexibility is greater than zero. It increases as the price increases.

Figure 5.18 shows the value of operating flexibility. The dashed line shows that for gold prices below $ 220, it is not worthwhile to start the project without flexibility. The solid line shows that for the project with flexibility the threshold gold price drops to $ 125. As expected as the gold price rises, the difference between the lines decreases, showing that the value of flexibility decreases. However, the option to run the mine would then be strongly-in-the money. So for low prices the termination flexibility is of high value; on the contrary for high gold prices it is of little value.

Figure 5.19 presents the expected value of additional information for initial gold prices varying in the interval from $ 100 to 400. For low initial prices (up to $ 175), information flexibility has no or little value. For higher prices the value of information flexibility increases. As the optimal decision among *start to develop the pit immediately* and *wait passively* is always *wait passively*, it is optimal to wait and get additional information.

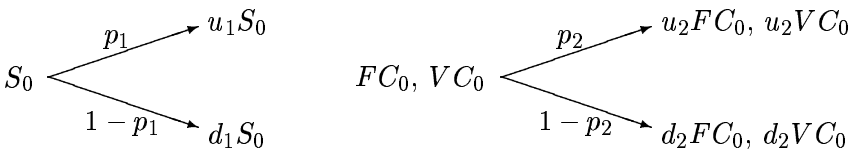
Uncertainty on gold prices, on reserves and on costs

So far costs were supposed known. However, sometimes costs are more uncertain than prices and payoffs (Slade, 2000). One care point in mining is the unpredictable cost of obtaining environmental approval. Examples in other industries include nuclear power plants for which construction costs are hard to predict due to engineering and regulatory uncertainties, the development of a newline of aircraft, urban construction projects, and many R&D projects, such as the development of a new drug by a pharmaceutical company (Dixit and Pindyck, 1994).

For projects that take time to complete and involve a sequential investment problem, Dixit and Pindyck describe two different kinds of cost uncertainty: technical uncertainty and input cost uncertainty. The first relates to the physical difficulty of completing the project: how much time, effort and materials will be required? Technical uncertainty can be resolved only by undertaking the project. The second arises when the prices of labor, land and materials needed to build a project fluctuate unpredictably, or when unpredictable changes in government regulations change the cost of construction. Prices and regulations change regardless of whether or not the firm is investing. These two types of uncertainty affect the investment decision differently and it could be important to incorporate both in the analysis (for more details see Dixit and Pindyck, 1994, p. 346). Here we do not differenciate these two kinds of costs. The objective is to introduce some uncertainty on the costs into the model and check the value of flexibility.

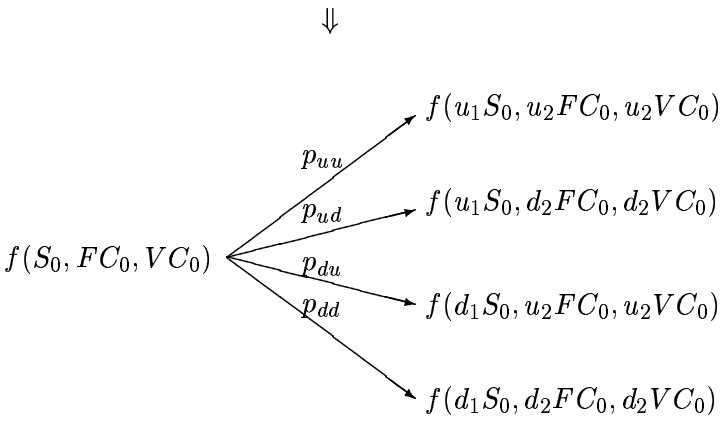
For this we assume that gold prices and costs are uncorrelated. Both fixed and variable costs are assumed

to follow geometric Brownian motions with zero drift and 10 % volatility. For simplicity, we assume that these two variables are perfectly correlated. Thus two perpendicular binomial trees are used to approximate the joint bivariate lognormal distribution. The time step, Δt, is still equal to $1/12$ year. The combination of these two processes is represented in Figure 5.3.



gold price increases by u_1 with probability p_1 or decreases by d_1 with $1 - p_1$

fi xed and variable costs increase by u_2 with probability p_2 or decrease by d_2 with $1-p_2$



function of gold price, fi xed and variable costs

Figure 5.3: *One period value tree*

Nodal probabilities are calculated as in Cox, Ross and Rubinstein (1979) for each independent binomial tree and are then multiplied together according to probability rules, that is

$$\begin{aligned} p_{uu} &= p_1 p_2, \\ p_{ud} &= p_1(1-p_2), \\ p_{du} &= (1-p_1)p_2, \\ p_{dd} &= (1-p_1)(1-p_2) \end{aligned}$$

where $p_1/\,(1-p_1)$ is the risk- neutral probability of an up/ (down) movement for the price and $p_2/\,(1-p_2)$ is the risk- neutral probability of an up/ (down) movement for costs. Note that as a zero drift is taken for costs there is no risk- neutral adjustment for costs.

The results are presented in Table 5.6. The uncertainty on the fixed costs does not have a marked impact as the fixed costs are assigned at the start of the project (i.e. either at $t = 0$ or at $t = 1$ year). The optimal initial decision among: *start to develop the pit immediately*, *wait* and *abandon* is *wait*. Initiation flexibility has value: \$ 5 (190-185) million. This corresponds to an increase in project value of 2.70 %. The value of operating flexibility is \$ 6 (190-184) million that is 3.26 % of project value.

Project evaluation model	**Expected value (NPV)**
Initial decision: start to develop the pit immediately	
stochastic model: no flexibility	180
stochastic model: flexibility	185
Initial decision: wait passively	
stochastic model: no flexibility	184
stochastic model: flexibility	**190**
Initial decision: wait and get additional information	
stochastic model: flexibility (*rich*)	324
stochastic model: flexibility (*average*)	221
stochastic model: flexibility (*poor*)	130

Table 5.6: *Expected value of the project in \$ million with unknown costs modelled as GBM's with zero drift and volatility of 10 % ($r = 0.02$, $\sigma = 11.41\%$)*

The expected value of additional information is now \$ 34 million, that is an increase in project value of 17.89%. We can compare these values with the ones in Tables 5.4 and 5.5. Note that the project values are equal or smaller than the values obtained with known costs. This is due to the fact that as time increases the GBM gives a higher probability to high costs than to low costs. The uncertainty on costs reduces the value of initiation flexibility. The introduction of the uncertainty on costs increases the value of operating flexibility. Because the project is more uncertain, flexibility has a higher value. The value of information is slightly smaller than the expected value of information with known costs.

Taking into account the uncertainty on costs increases, although just slightly in this case, the value of flexibility and modifies the expected NPV. As we do not have any data to define the random process for costs we will suppose them known. Moreover, the uncertainty on costs makes the hypothesis of a replicating portfolio dubious. Costs, as well as development strategy, depend on the geological uniqueness of the reserve. In this case a SDP approach should be used and an adjusted discount rate defined.

5.7.4 Value of the four development options

In this section we compare the small pit and the large one, and determine the optimal project and the optimal strategy. As we have seen previously the four development options available to management are

1. develop the small pit;
2. develop the large pit;
3. carry out additional drilling and then choose between the large or small pit;

4. start to develop the large pit with the possibility to revert to the small pit.

At first we evaluate the project defined by the development options 1, 2 and 3. Then we will evaluate the entire project defined by the four options and finally value additional drilling.

The first node of the decision tree is presented in Figure 5.4.

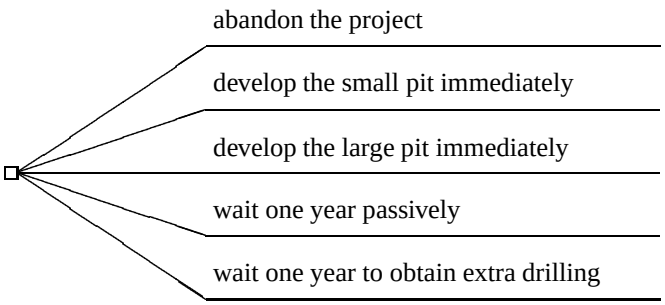


Figure 5.4: *First decision node*

We present the results for the small and large pit separately to highlight the value of flexibility which is expected to be higher for the large pit as its reserves and the the gold price are more uncertain.

The value of project U_1 (develop the large pit) and of project U_2 (develop just the small pit) are presented in Table 5.7. Recall that the large pit is not given by the small pit plus the lower part of large pit. Comparing the initial decisions *abandon the project*, *start to develop the pit immediately* and *wait*, the optimal initial decision is *wait* and the optimal project is U_1. Initiation flexibility has then value: \$ 36 (595 - 559) million. This corresponds to an increase in project value of 6.44 %. Note that the fact of waiting 1 year before starting the development increases the difference between the 2 project values.

Project evaluation model	**Expected value (NPV)** Large pit (U_1)	Small pit (U_2)
Initial decision: start to develop the pit immediately		
deterministic model ($E[S_t]$)	544	472
stochastic model: no flexibility	548	475
stochastic model: flexibility	559	476
Initial decision: wait 1 year before to develop the pit		
deterministic model	578	495
stochastic model: no flexibility	582	497
stochastic model: flexibility	**595**	499

Table 5.7: *Expected value of the project in \$ million for the initial decisions* start to develop the pit immediately *and* wait 1 year before to develop the pit *($r = 0.02$, $\sigma = 11.41\%$)*

In the model with flexibility the options available to management are: continue or abandon the project. That is every each year management can decide to continue or abandon the project in response to low gold prices and/or low reserves. The value of operating flexibility is for U_1: \$ 13 (595-582) million, that is an increase in project value of 2.23 %. The value of operating flexibility is for U_2: \$ 2 (499-497) million, that is an increase in project value of 0.40 %. The value of flexibility is higher, as it was expected, for U_1, as the uncertainty on its reserves and on the price is higher. Figure 5.20 presents and compares the histograms of the expected values for the two projects without and with flexibility. Note that flexibility enhances the project value. Note that project U_1 can be strictly preferred to U_2: the probability of low profit values is just slightly higher for U_1 than for U_2. The risk of taking the wrong decision is small. This can also be seen in Figure 5.21 which compares the expected values, computed without flexibility, for the two projects for each simulation. Observe that when project U_1 is preferred to U_2 it is strongly preferred.

Project evaluation model	**Expected value (NPV)** Large pit (U_1)	Small pit (U_2)
Initial decision: wait 1 year and get additional information		
Additional information: rich		
stochastic model: no flexibility	747	513
stochastic model: flexibility	**754**	515
Additional information: average		
stochastic model: no flexibility	629	522
stochastic model: flexibility	**640**	524
Additional information: poor		
stochastic model: no flexibility	499	464
stochastic model: flexibility	**516**	466

Table 5.8: *Expected value of the project in* \$ *million for the initial decision* wait 1 year and get additional information *($r = 0.02, \sigma = 11.41\%$)*

Additional information

Table 5.8 presents the value of the project if additional information is obtained. For the 3 different scenarios the project is still worthwhile to be undertaken. Note that the expected values of project U_2 given the initial data or given the initial data + additional information are very similar. The differences are due to the statistical fluctuations of the simulations and to the few extra data that test the upper zone. The optimal project is U_1 for the rich, average and poor data.

The value of operating flexibility is higher for the case of poor additional information: it is of \$ 17 (516-499) million, that is an increase in project U_1 of 3.41 %.

Now we must choose the optimal strategy: is it worthwhile or not to get the additional data to better define the 2 projects and choose between them? We need to quantify the value of this additional information. The expected value of information, computed as the difference between the project values with and without information and assuming that the high scenario has a probability of 0.25, the average of 0.50 and the poor of 0.25, is \$ 42.50 million. It increases the project value of 7.14 %. Thus it is worthwhile to obtain it.

Moreover, the reduction of risk due to additional information can be taken into account. Looking at Figure 5.21 note that the rich extra data discriminates best between the 2 projects. The average and poor data sets do not help choosing the optimal project. Observe that given the poor extra data the 2 projects are now closer. Consider, for example, the poor scenario. First we present the risk values with the additional information available in $t = 0$. This is to compute the value of additional information at $t = 0$.

Data	Optimal initial decision	Risk	Value of additional information
initial data	develop the large pit	27	-
additional information: rich	develop the large pit	1	26
additional information: average	develop the large pit	21	6
additional information: poor	develop the small pit	35	0

Table 5.9: *Risk and value of additional information (\$ million)*

The risk values are presented in Table 5.9. It can be noted that although the posterior distribution of the recovered grade given the rich additional information had the highest variance (Table 4.3), the rich additional information discriminates best between the two projects. It has the lowest risk value and provides the decision maker with the most sure scenario. On the contrary the average and poor additional information have a high risk value. In particular the poor extra data gives the less sure scenario as the expected values of the two projects are the closest. Note that as we have defined a constant production rate the gain procured by additional information is also the optimal definition of this rate.

If we assume that the high scenario has a probability of 0.25, the average of 0.50 and the poor of 0.25, the expected value of additional information is \$ $0.25 \times 26 + 0.5 \times 6 + 0.25 \times 0 = 9.5$ million. The experience is thus worthwhile to be undertaken. The efficiency of getting new information is given by $9.5/27 = 0.35$. Note that the more efficient the experience the more this value is close to 1. Thus if at the start of the project we had to choose between develop immediately the large pit given the initial data or given the initial + extra data at a cost of \$ 10 million, the optimal decision would be to develop the large pit after having obtained additional information. The extra data help discriminating the two projects.

Now as the additional information is not available at $t = 0$ we need to take in consideration the time factor. As we have seen the option to wait has value: the optimal initial decision is *wait and then develop the large pit*. The risk values are presented in Table 5.10. Note that the risk values are smaller than in the previous table. This is due to the discounting and to the fact that it is optimal to wait. Here the highest risk corresponds to the poor additional information as the two project values are the closest.

If we assume that the high scenario has a probability of 0.25, the average of 0.50 and the poor of 0.25, the value of additional information is \$ $0.25 \times 21 + 0.5 \times 5 + 0.25 \times 0 = 7.75$ million. Additional information reduces the risk by 37 %. The efficiency of getting new information is $7.75/21 = 0.36$. It is then worthwhile to obtain it.

Sensitivity to the initial gold price

A sensitivity study is performed to measure the impact of the initial gold price on the project value and on the value of flexibility. The initial price is altered over the range from \$ 180 to 320.

Data	Optimal initial decision	Risk	Value of additional information
initial data	develop the large pit	21	-
additional information: rich	develop the large pit	0	21
additional information: average	develop the large pit	16	5
additional information: poor	develop the small pit	28	0

Table 5.10: *Risk and value of additional information (*$ *million)*

Develop immediately or wait? Figure 5.22 compares the value of the 2 projects for different initial prices. The initial price is altered over the range from $ 180 to 320. For the initial decision *develop immediately* project U_1 becomes more interesting than U_2 for prices around $ 260. For the initial decision *wait* project U_1 becomes more interesting than U_2 for prices around $ 240. Figure 5.23 compares the value of the two projects given the poor additional information. For prices around $ 280 or larger project U_1 becomes more interesting than U_2.

Initiation flexibility. A sensitivity study is performed to measure the impact of the initial gold price on the initial decision (develop immediately/ wait) for project U_1. Figure 5.24 presents the value of initiation flexibility: the option to wait has value for all the prices in the interval considered. For lower prices it has little value.

Operating flexibility. A sensitivity study is performed to measure the impact of the initial gold price on the value of operating flexibility. Figure 5.25 shows the operating flexibility value. For prices smaller than $ 180 the project valued with flexibility is not worthwhile to be undertaken, while for prices smaller than $ 230 the project valued without flexibility is not worthwhile. It can be noted that near the break- even gold price of $ 230 the option value is at his peak. As the gold price increases the option value decreases.

Information flexibility. The development of the large pit is considered. Figure 5.26 presents the expected value of additional information for initial gold prices varying in the interval from $ 190 to $ 420. Additional information has value.

Figure 5.27 also shows the expected value of additional information for different initial gold prices. The initial gold prices vary in the interval from $ 150 to 450. For prices around $ 270 the value of information is at its peak. Thus, in the presence of costly learning, there exists an upper and a lower critical boundary within which it is optimal to exercise the learning action. Outside this range it is not optimal to pay a cost to learn. These results are consistent with the results of Martzoukos and Trigeorgis (2001).

Sensitivity to the variable cost of waste

Figure 5.28 highlights the value of flexibility as a function of the variable costs. They are altered over the range from $ 5 to 10 per ton of waste. It can be noted that near the break- even cost of $ 7.5 per ton of waste the option value is at its peak. The minimum and maximum costs for flexibility to have value are around $ 5 and $ 10 per ton of waste.

Choice of the discount rate

In Section 5 we saw that several authors including Lund (1999), consider that private risk (such as uncertainty on the reserves) can not be hedged and that consequently the discount rate has to be increased to account for this. This is why we decided to vary the rate.

Figure 5.29 compares the 2 projects for varying interest rates. The small pit becomes more interesting for rates greater than 17 %. Figure 5.29 also presents the value of initiation flexibility for different interest rates. The development of the large pit is considered. The option to wait does not have value for interest rates greater than 7 %. This highlights the importance of correctly defining the discount rate. It greatly influences the project values and the optimal strategy. Thus if a SDP approach was considered with a risk adjusted discount rate the results could be quite different.

Suppose that the discount rate, ρ, is 7 %. It is higher than the risk free rate as it must account for the project's risk. We have seen that in the case of a risk- neutral valuation the drift rate for the price process is replaced by the risk- free rate. Here, on the contrary, we need to define the drift rate for the price process. It can be estimated using historical data. This was done in Section 7.2. Table 5.11 presents the project values for the initial decisions *start to develop the pit immediately* and *wait 1 year before to develop the pit*. The optimal decision is now develop the large pit immediately. Initiation flexibility has no value. Operating flexibility increases the project of 2.97 %.

Project evaluation model	**Expected value (NPV)** Large pit (U_1)	Small pit (U_2)
Initial decision: start to develop the pit immediately		
stochastic model: no flexibility	404	393
stochastic model: flexibility	**416**	395
Initial decision: wait 1 year before to develop the pit		
stochastic model: no flexibility	389	375
stochastic model: flexibility	404	377
Initial decision: wait and get additional information		
stochastic model: flexibility (*rich*)	**527**	389
stochastic model: flexibility (*average*)	**438**	396
stochastic model: flexibility (*poor*)	339	**348**

Table 5.11: *Expected value of the project in* \$ *million* ($\rho = 0.07$, $\alpha = 0.02$, $\sigma = 11.41\%$)

Note that the two projects are now closer. This is due to the higher discount rate. The table also presents the results for additional information. In particular, note that if the deposit turns out to be poor it is now optimal to develop the small pit. The expected value of additional information is \$ 21.75 million. It increases project value by 5.23 %, which is worthwhile. However, its value is smaller than that obtained with the risk- neutral approach as here the optimal initial decision is to develop immediately the large pit.

The two approaches, risk- neutral valuation or risk- adjusted SDP, lead to different results. The major drawback of the risk- adjusted SDP is that it gives no indication on how to define the rate. Moreover, here we would need to define a rate for the development of the large pit and one for the small pit as their risks are different. Observe that, the risk- neutral approach, if not appropriate, overvalues the project.

Option to shift from large pit to small pit?

We now add the possibility to shift from project U_1 to U_2, that is we suppose that if the large pit is developed immediately management can decide after one year (at $t = 1$) to revert to the small pit if the project is not as interesting as expected. The first node of the decision tree is shown in Figure 5.5.

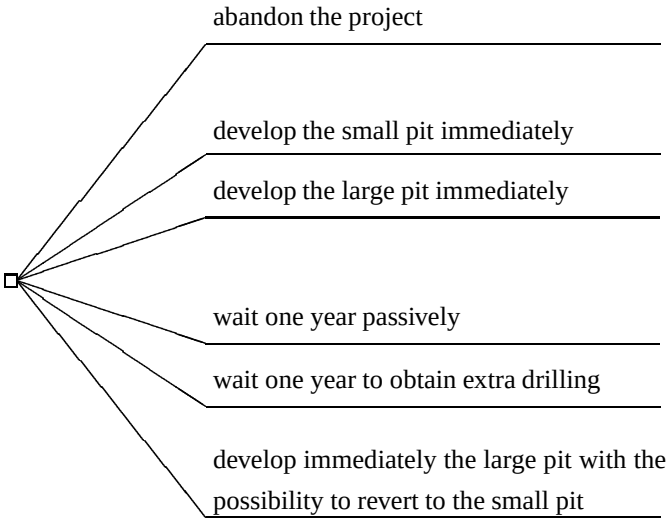


Figure 5.5: *First decision node*

The results are shown in Table 5.12 and are compared to the results obtained for the development of the large pit but with no option to shift to the small pit. The cost of reverting to the small pit is $ 2 million, that is smaller than the cost of abandoning the project. The value of the project with this additional flexibility is $ 621 million, that is an increase of 11.24 % in project value, for the immediate development. It is $ 654 million, that is an increase of 9.83 % in project value, for the delayed development. Thus, if additional information cannot be obtained, the optimal initial decision is *wait 1 year and then start to develop the large pit with the possibility to revert to the small pit.*

Now, if additional information is obtained and the large pit is developed, at year $t = 2$ management can decide whether to continue developing the large pit or shift to the small pit. The value of the project is $ 763 million if the information is rich, that is an increase of 1.22 % in project value. The value of the project is $ 693 million if the information is average, that is an increase of 8.26 % in project value. And finally, if the information is poor the value of the project is $ 584 million, that is an increase of 13.17 % in project value. The expected value of additional information is $ 29 million, that is information increases the project value of 4.51 %. It is obtained comparing the optimal value of the entire tree with the possibility of drilling extra holes and the optimal value of the tree without the possibility of drilling extra holes.

The risk values of taking the optimal initial decision is clearly reduced. The opportunity to modify the

Project evaluation model	**Expected value (NPV)** Option to shift	No option to shift
Initial decision: start to develop the pit immediately		
stochastic model: flexibility	621	559
Initial decision: wait 1 year before to develop the pit		
stochastic model: flexibility	**654**	**595**
Initial decision: wait and get additional information		
stochastic model: flexibility (*rich*)	763	754
stochastic model: flexibility (*average*)	693	640
stochastic model: flexibility (*poor*)	584	516

Table 5.12: *Expected value of the project in* \$ *million* ($r = 0.02$, $\sigma = 11.41\%$)

project after one year, that is to revert to the small pit, greatly reduces the risk of the project. Here it is 0 while it was of \$ 21 million without this additional flexibility. The initial risk is small and extra information is not useful to help choosing between the 2 projects. Additional information helps better defining the project as it augments the project value by 4.51 % and is thus worthwhile to be obtained. The optimal development strategy is then *wait 1 year and get additional information, then develop the large pit with the possibility to revert to the small pit*.

What if the plug-in reserves were used?

Up to this part, the values of the reserves were those computed using the Bayesian approach. As an aside, we consider what would have happened had we used the plug- in reserves (Chapter 2). As we saw the reserves distributions for the two approaches, Bayesian and plug- in, mostly differ in the tails. Thus, although the optimal project is expected to be the same whether the Bayesian or the plug-in results are considered, the 2 projects, and in particular the development of the large pit, are expected to differ. Table 5.13 presents the values of project U_1 for the initial decisions *start to develop the pit immediately*, *wait passively* and *wait 1 year and get additional drilling* for both approaches.

Table 5.13 shows that for the development of the large pit the value of flexibility is slightly higher for the *plug-in* reserves: \$ 18 million (that is an increase in project value of 3.20 %) while it was of \$ 13 million (2.23 %) for the Bayesian reserves. This is due to the fact that in this case the *plug-in* gives slightly larger weight to the low values as it was seen in Figure 4.8.

The value of initiation flexibility is essentially the same whether the plug-in approach or the Bayesian approach is used, even though the absolute values of the project are different.

Having said that, there is a significant difference between the expected value of information. For the plug- in approach, the expected value of information is \$ 81.25 million ($0.25 \times 798 + 0.5 \times 661 + 0.25 \times 529 - 581$), that is an increase in project value of 13.98 %. It almost doubles the value obtained with the Bayesian approach. This shows that the model for the reserves is important.

Project evaluation model	**Expected value (NPV) of large pit** Plug- in	Bayesian
Initial decision: start to develop the pit immediately		
stochastic model: no flexibility	525	548
stochastic model: flexibility	545	559
Initial decision: wait		
stochastic model: no flexibility	563	582
stochastic model: flexibility	**581**	**595**
Initial decision: wait 1 year and get additional information		
Additional information: rich		
stochastic model: no flexibility	792	747
stochastic model: flexibility	**798**	**754**
Additional information: average		
stochastic model: no flexibility	651	629
stochastic model: flexibility	**661**	**640**
Additional information: poor		
stochastic model: no flexibility	512	499
stochastic model: flexibility	**529**	**516**

Table 5.13: *Expected value of the development of the large pit in* \$ *million (*$r = 0.02$, $\sigma = 11.41\%$*)*

5.8 Conclusions

In this chapter we have presented the real option methodology for evaluating natural resource projects subject to both market and technical risk. Its strong point compared to NPV is that it can be used to value flexibility inherent in projects. Stochastic dynamic programming has been used to carry out the computations because it allows us to value early exercise options. The initial decisions that have been considered are

1. start immediately,
2. wait one year passively,
3. wait one year to acquire new information by extra drilling

while the development options considered are

1. develop the small pit,
2. develop the large pit,
3. start developing the large pit but revert to the small one if results are not promising.

The results highlighted the value of managerial flexibility, which can be substantial. In particular, the option to shift to the small pit greatly enhances the project value. Initiation flexibility is found to have value, but operating flexibility has little value. The importance of accounting for reserves uncertainty was also highlighted. For this we focussed on the lower part of the deposit as it is the most uncertain zone. Moreover, the comparison of the results obtained with the Bayesian and conditional simulation approaches permitted us to

highlight the importance of the reserves distribution in the project evaluation.

One of the main aims of this work was to build a model that evaluates the option *wait and obtain additional information*. We measured the value of information as the difference between the project values with and without additional information and as the expected reduction of risk. The first quantity highlights the fact that additional information permits us to better define the project. The second quantity highlights the fact that additional information allows to better discriminate the 2 projects. We showed that it is worthwhile to obtain extra data although it is expensive. In particular, the rich additional information has the highest value as discriminates best the 2 projects. Whereas the poor extra holes have the lowest value. However, note that, even without getting extra information on the reserves, the optimal initial decision is *wait*. As we have seen for the decision approach with $\rho = 7\%$ it is worthwhile to obtain additional information although it delays the start of the project. Immediate investment is justified if additional information will not change the investment decision or information will not be available before the investment must be taken or abandoned. Thus purchase information that will impact the upcoming decisions, if the value increase justifies the cost of the information. Or invest now or abandon the project if there is no information to be gained (or its expense is too great) that will significantly change the project's outcome or impact the investment's decision process. Information has value not only if it leads us to modify the initial decisions: it allows to better define the project.

One generalisation of the model used would be to take into account the possibility for technical risk to evolve over time. This could be particularly important in the production phase. For this Connell (2002) used a GBM. Cortazar *et al.* (2001) suggested that this could be also important in the exploration phase and in our case for the lower zone of the deposit which is sparsely sampled. For this Cortazar *et al.* used a GBM. However, note that during the exploration phase the reserves change at discrete time intervals.

In a mining project other sources of flexibility may be valuable. The rate of production was assumed constant per year and all the amount produced was assumed sold. The possibility to vary the production or the cutoff grade in response to varying prices or costs could be considered. Of course the complexity of the decision problem would then increase.

The rate of return is important to the analysis. It is true that the existence of a replicating portfolio and thus the use of the riskfree rate is dubious and must be checked. In decision or SDP models little interest is given to its determination. The rate of return could be obtained using relevant market information for the gold price. If a replicating portfolio is defined, the risk neutral valuation can be applied and then in the SDP equation the drift rate α will be replaced by $r - \delta$ where δ is the convenience yield. Thus the results are quite different whether market information is used or not to define market uncertainties. In addition, note that we used a constant interest rate. This is an acceptable assumption when the life of the project is short (less than one year). Models that include a stochastic interest rate could be used. To develop practical valuation models and risk management strategies around physical assets, treated as options, one must first and foremost set limits on what is practical and achievable. We believe that for projects that are dominated by market risk CCA should be used, whereas for projects that are dominated by private risk SDP should be used.

The results depend strongly on the model that is chosen for the price and on its parameters. It could have been interesting to compare the results obtained with the GBM with a mean reverting process. However, as it was seen in Figure 5.6, for a period of time of 10 years or shorter, as it is our case, it is difficult to choose between a GBM or a mean reverting. Note that from an economic point of view there are reasons why the GBM cannot be expected to hold for natural resources such as copper and oil (Stensland and Tjøstheim,

1989). This model essentially asserts that past price information can be neglected given today's price. Although this is a reasonable assumption for most stock markets, for natural resources one would believe that it is of interest for the decision making to know the price history prior to the present price. The supply of a raw material is affected by a change in the price after a time lag. The memoryless property of this model neglects the nature of the production process. High prices will bring more mines into production, and this will again lower the price. This tends to produce a negative autocorrelation in the price process. Furthermore the future price in a raw material market could be both higher and lower than today's spot price. For example, the price of storage may be so high that it pays to buy on future contracts, thus putting pressure on this price so that it exceeds the spot price. In contrast, buyers will often need the resource in consumption and production on an immediate basis. They cannot wait for lower prices, and thus the firms producing in periods with high prices will make profit that they cannot expect to make if they store the raw material for future sale. In this way it is possible that the future price is lower than the present spot price. Such a situation is not likely to occur in stock markets. Who will buy a stock now when it is possible to buy it cheaper on future contracts (Karma, 1982).

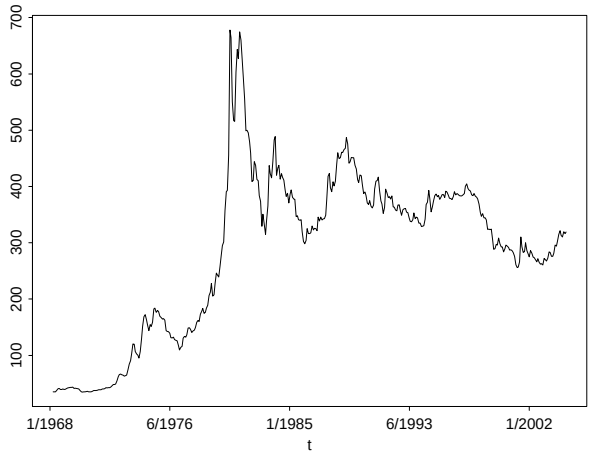


Figure 5.6: *Monthly price of gold in US dollars per ounce (the data are the monthly average prices) from 1/1968 to 11/2002. In the mid 80's and 90's the price of gold in US dollars has fallen drastically*

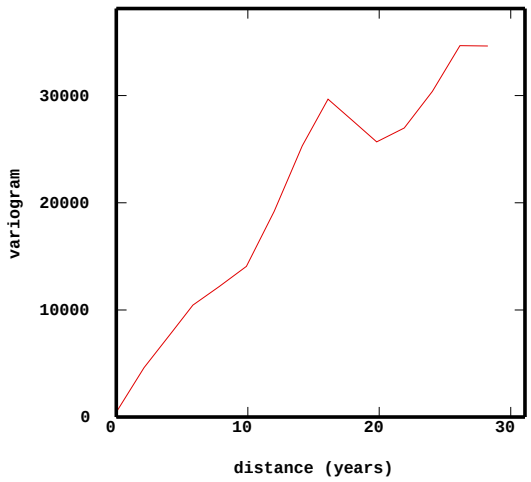


Figure 5.7: *Experimental variogram of the monthly series of the gold price*

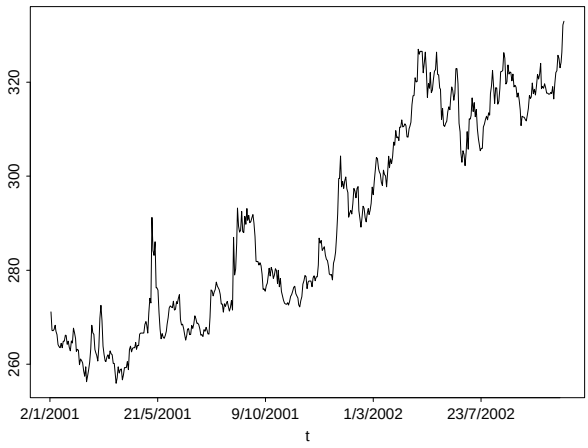


Figure 5.8: *Daily price of gold in US dollars per ounce (PM closing values) for the years 2001 and 2002 (from 2/1/2001 to 16/12/2002)*

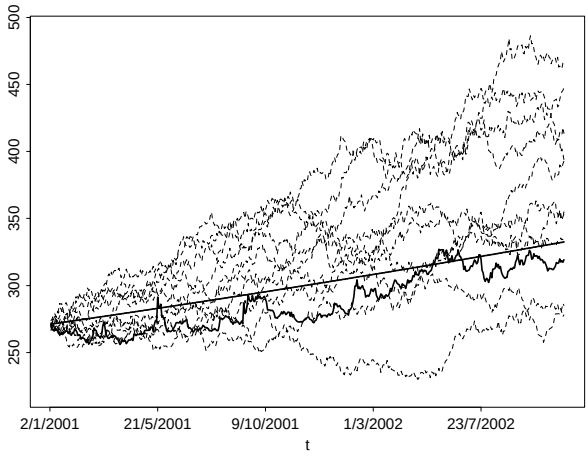


Figure 5.9: *Daily series of gold price (bold) and GBM's estimated trend with* $\alpha = 0.11$ *and* $\sigma = 12.52\%$*. Also* 10 *realisations of this process are presented (*$S_0 = \$\,271.1$*)*

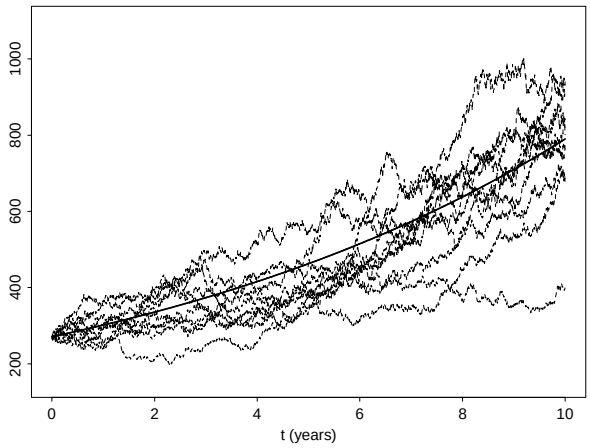


Figure 5.10: *GBM's estimated trend with* $\alpha = 0.11$ *and* $\sigma = 12.52\%$ *and* 10 *realisations of this process for 10 years (*$S_0 = \$\,271.1$, $t_0 = 2/1/2001$*)*

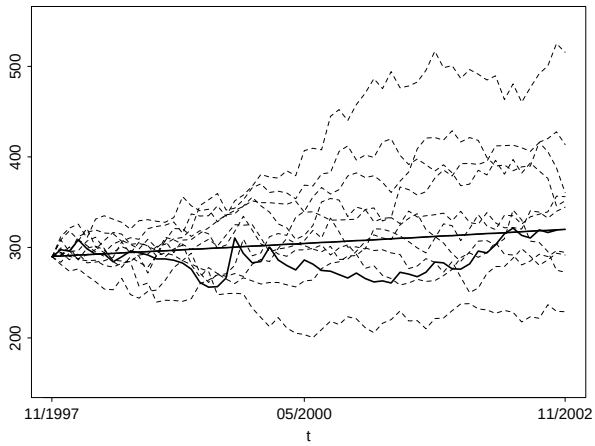


Figure 5.11: *Monthly series of gold price (bold) and GBM's estimated trend with* $\alpha = 0.02$ *and* $\sigma = 11.41\%$*. Also* 10 *realisations of this process are presented (*$S_0 = \$\,289.45$*)*

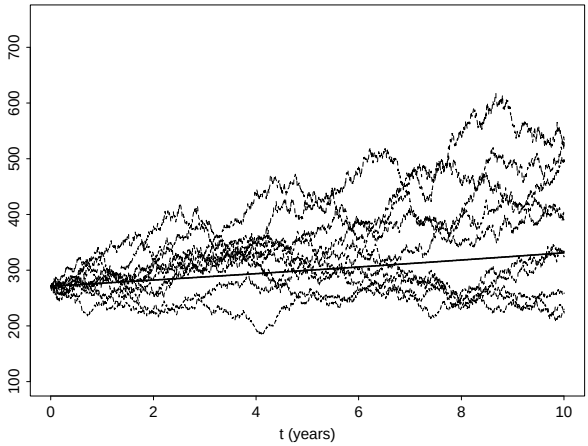


Figure 5.12: *GBM's estimated trend with* $\alpha = 0.02$ *and* $\sigma = 11.41\%$ *and* 10 *realisations of this process (*$S_0 = \$\,271.1, t_0 = 2/1/2001$*)*

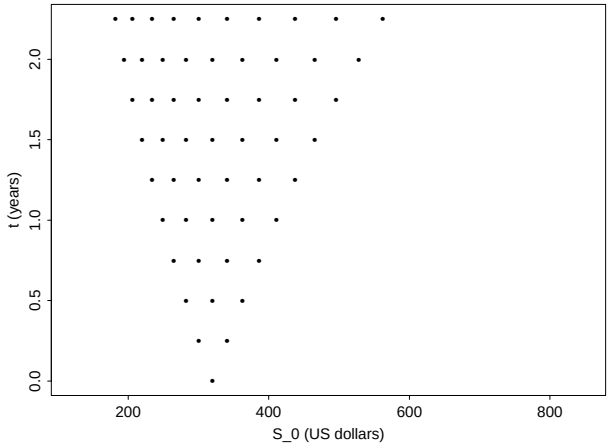


Figure 5.13: *Binomial tree with* $S_0 = \$\,320$, $\Delta t = 1/4$ *year,* $\alpha = 0.11$ *and* $\sigma = 12.52\%$

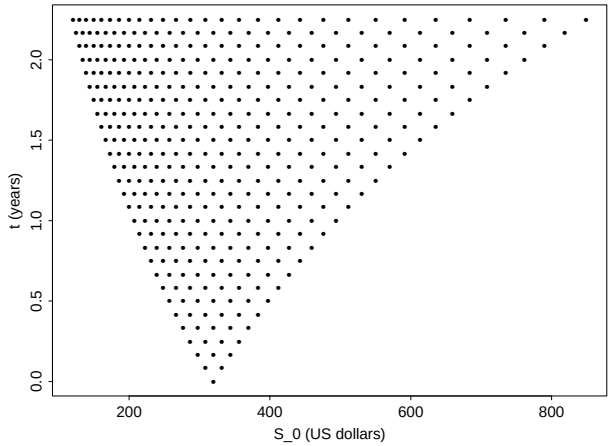


Figure 5.14: *Binomial tree with* $S_0 = \$\,320$, $\Delta t = 1/12$ *year,* $\alpha = 0.11$ *and* $\sigma = 12.52\%$

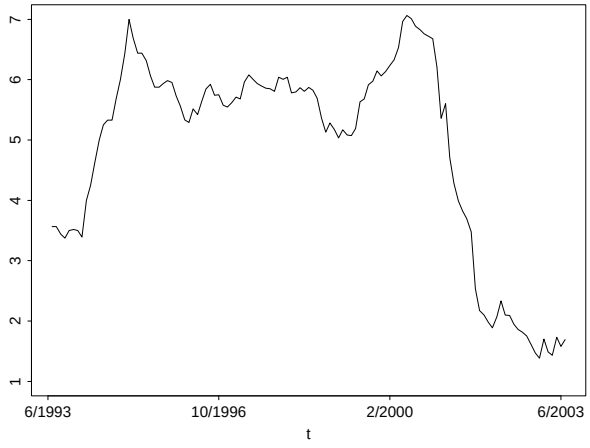


Figure 5.15: *6-month LIBOR (last trading day of each month) series from 6/1993 to 6/2003*

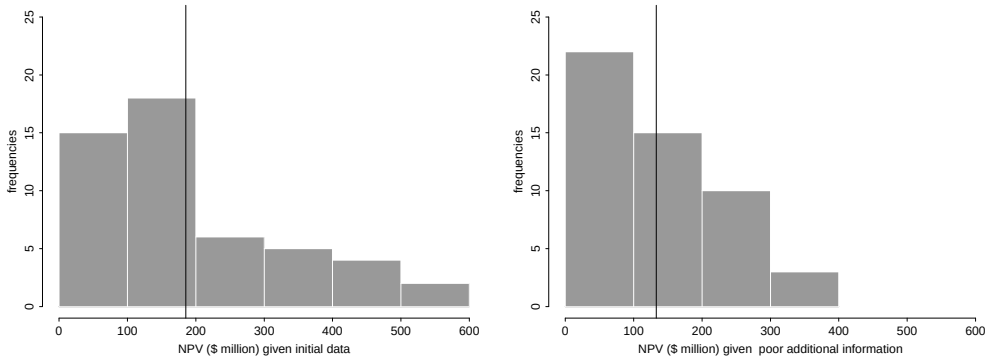


Figure 5.16: *Development of the lower part of large pit. Histograms of the project values given the initial data and given the initial data + poor additional information. The projects are valued with flexibility*

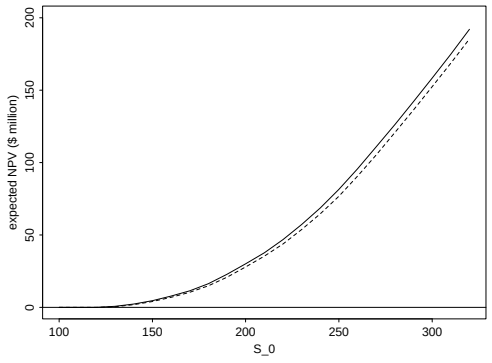


Figure 5.17: *Initiation flexibility for the development of lower part of large pit as a function of the initial price* S_0. *The continuous line represents the project value for the initial decision* wait passively. *The dashed line represents the project value for the initial decision* develop the pit immediately

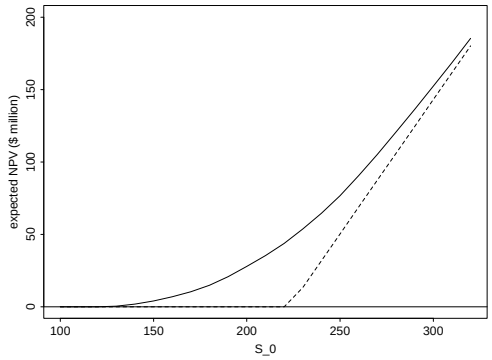


Figure 5.18: *Operating flexibility for the development of lower part of large pit for different initial gold prices. The continuous line represents the project value for the stochastic model with flexibility. The dashed line represents the stochastic model without flexibility*

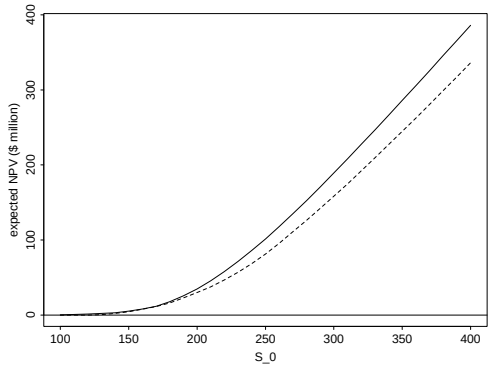


Figure 5.19: *Value of additional information for the development of lower part of large pit as a function of the initial gold price. The continuous line represents the project value for the initial decision* wait and get additional information. *The dashed line represents the project value for the initial decision* wait passively. *The projects are valued with flexibility*

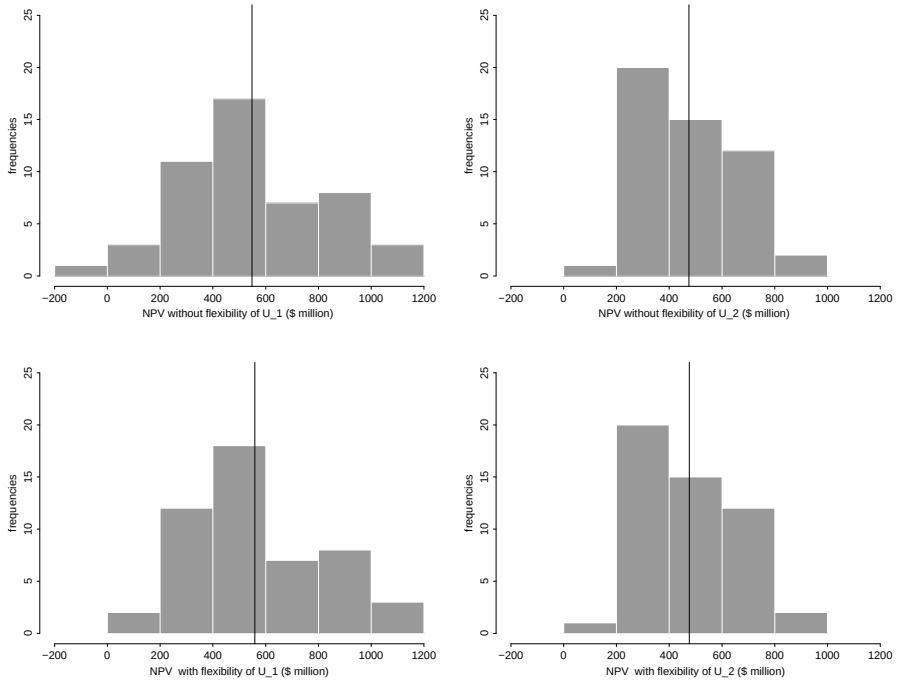


Figure 5.20: *Histograms of the 2 projects values computed without flexibility (above) and with flexibility (below). The initial decision is* develop immediately. *Project U_1, on the left, is, as expected, more uncertain than project U_2, on the right. The variation coefficients are 0.50 and 0.38 for U_1 and U_2, respectively. For U_1 the probability of low values ($< \$\,200$ million) is slightly higher than for U_2. Whereas the probability of high values ($> \$\,800$ million) is much higher. Taking into account flexibility the frequence of small NPV decreases for U_1. It is unchanged for U_2. The variation coefficients are now 0.45 and 0.37 for U_1 and U_2, respectively. Taking into account flexibility enhances, although slightly in this case, the project value*

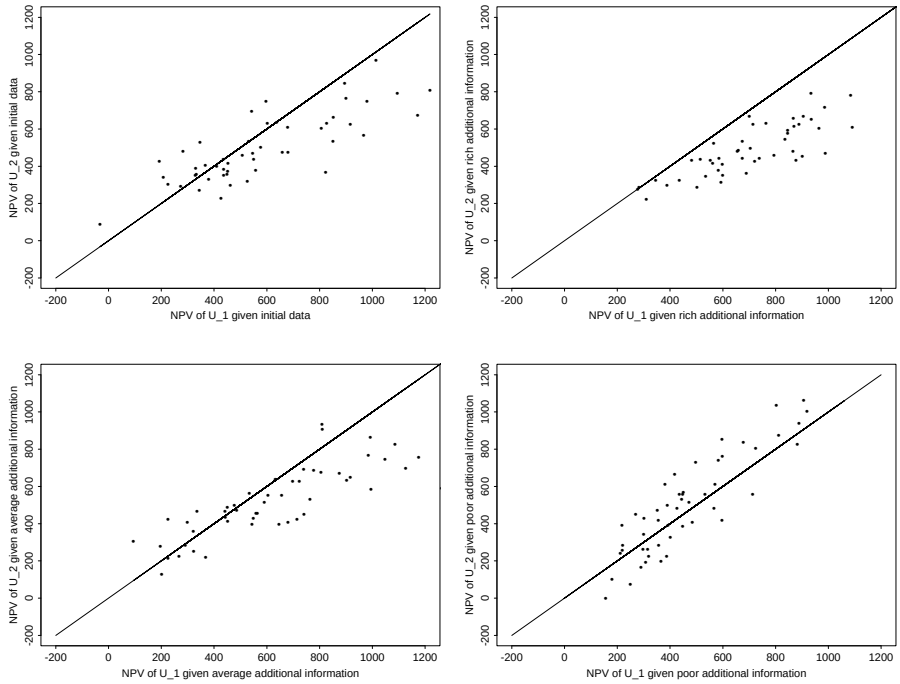


Figure 5.21: *Comparison of the NPV values ($ million) of the 2 projects given the initial data and given the initial + the rich, average and poor extra data. Given the rich additional information, project U_1 can be strictly preferred to U_2*

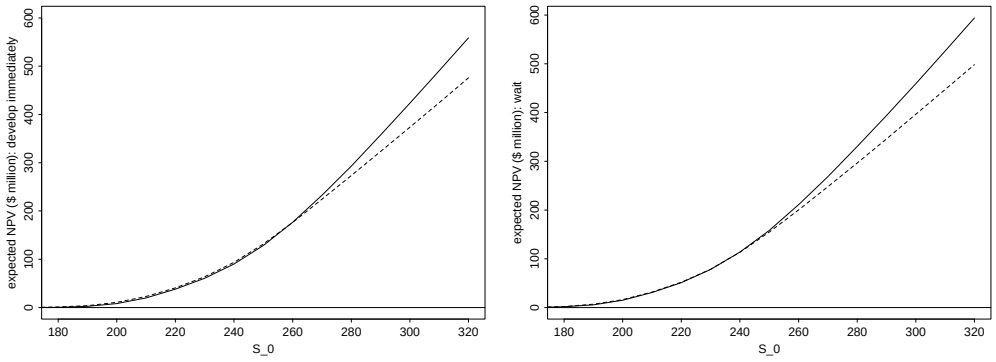


Figure 5.22: *Comparison of the 2 projects for different initial gold prices. The projects are valued with flexibility for the initial decisions* develop immediately the pit *and* wait. *The continuous line represents project U_1 and the dashed line represents project U_2. In the figure on the left project U_1 becomes more interesting than U_2 for prices around $ 260. In the figure on the right project U_1 becomes more interesting than U_2 for prices around $ 240*

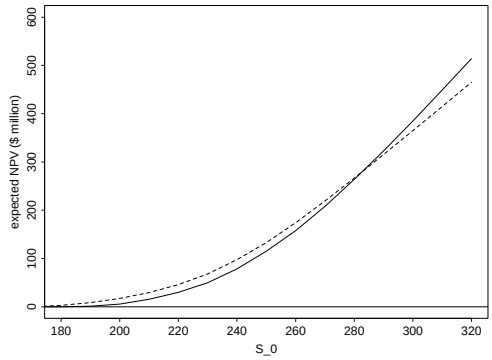


Figure 5.23: *Comparison of the 2 projects for different initial gold prices given the poor additional information. The continuous line represents project U_1 and the dashed line represents project U_2. The projects are valued with flexibility. For prices around $ 280 project U_1 becomes more interesting than U_2*

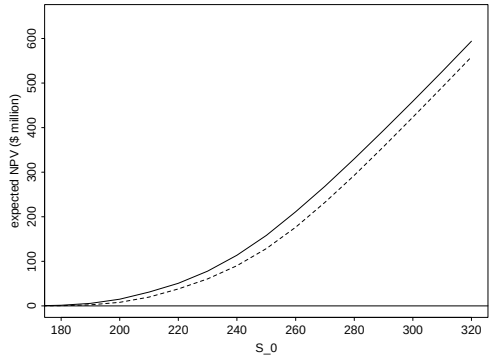


Figure 5.24: *Initiation flexibility for the development of large pit for different initial gold prices. The projects are valued with flexibility. The continuous line represents the project for the initial decision* wait *and the dashed line represents the project for the initial decision* start the development immediately. *The option to wait has value for all the prices in the interval considered. Its value is small for low prices*

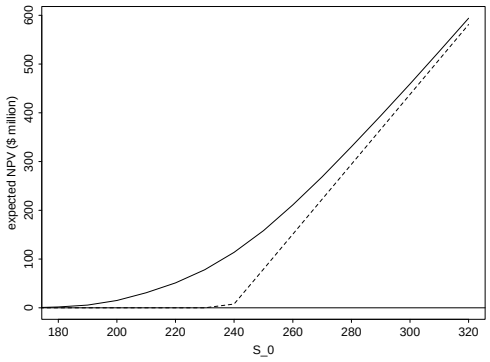


Figure 5.25: *Operating flexibility for the development of large pit for different initial gold prices. The initial decision is wait. The continuous line represents the project valued with flexibility and the dashed line represents the project valued without flexibility. It looks like a standard option payoff diagram*

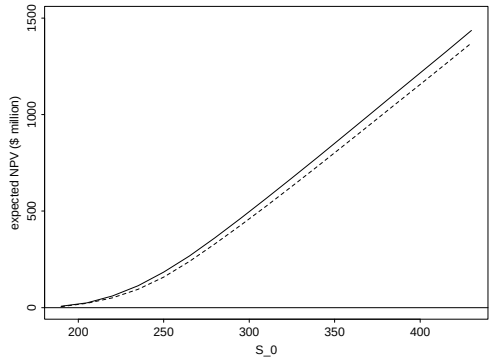


Figure 5.26: *Value of additional information for the development of large pit as a function of the initial gold price. The continuous line represents the project valued with additional information and the dashed line represents the project valued without additional information. As the optimal initial decision is* wait *it is worthwhile to* to wait and get extra data

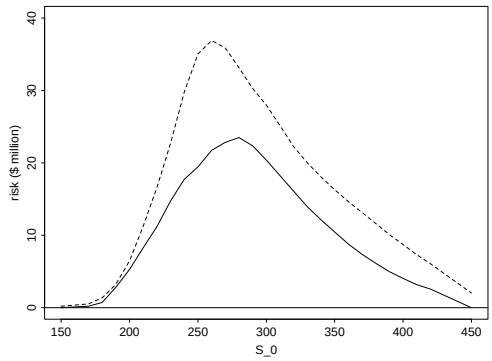


Figure 5.27: *Information flexibility as a function of S_0. The continuous line represents the expected risk for the initial data + additional information. The dashed line represents the risk given the initial data. Note that for low or high prices the value of information is small, that is the project is either too bad or too good to invest in order to learn more or to worry about possibly lower realized cash flows*

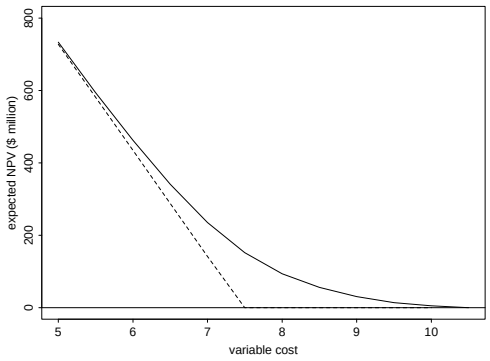


Figure 5.28: *Operating flexibility as a function of the variable cost. The large pit is developed at* $t = 1$. *The continuous line represents the stochastic model with flexibility and the dashed line the stochastic model without flexibility. The value of flexibility is higher for higher costs and reaches its maximum for around* $ *7.5 per ton of waste. Then flexibility starts to decrease again to a maximum cost of around* $ *10 per ton of waste*

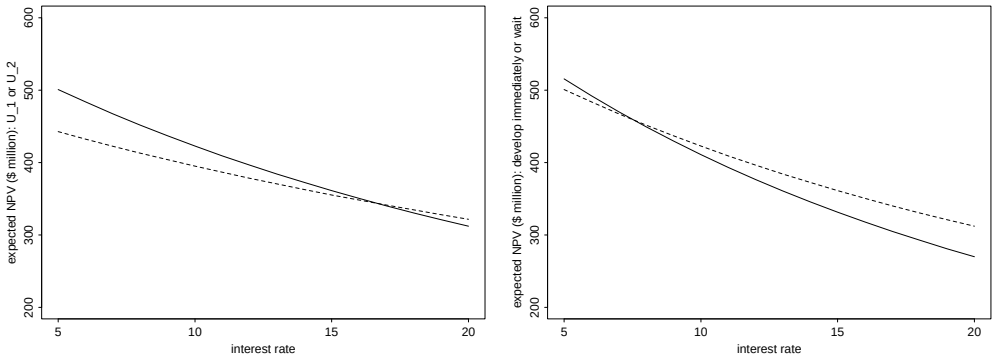


Figure 5.29: *Choice of the project and initiation flexibility as a function of the interest rate. The projects are valued with flexibility. In the graphic on the left the continuous line represents* U_1 *and the dashed line* U_2. *In the graphic on the right the continuous line represents* U_1 *for the initial decision* wait *and the dashed line for* develop immediately

Chapter 6

Conclusions and perspectives

The problem that motivated this work was the evaluation of an open pit gold mine project, and in particular the selection of the best development option. At the time we started this work it was common practice to run conditional simulations of the deposit and combine them with the NPV approach or to consider a fixed scenario for the reserves and combine it with the real options framework. This thesis focuses on modelling the gold reserves using conditional simulations or a Bayesian approach and on the financial evaluation of the project using real options.

Conditional simulations were used because they are well accepted by the mining industry (Dowd, 1994; Sanguinetti *et al.*, 1997; Thwaites, 1998). However, strong hypotheses on the covariance function are needed. For this the Bayesian framework was introduced. A Bayesian approach to modelling was considered for the reserves of the orebody for the available information: the upper part of the deposit has been densely sampled and the orebody seems to continue downwards. Although the continuity between the upper and lower parts was assumed, the Bayesian approach permitted to leave some uncertainty on the covariance structure.

A central concern in spatial statistics by gaussian random fields is the identification of the covariance structure. The kriging procedure is widely used for prediction in mining because it produces optimal predictions when the covariance structure of the random field is known. In reality the covariance is not known and needs to be estimated. The optimality of kriging is then questionable. The Bayesian paradigm provides a framework in which to analyse the performance of the kriging predictor. As important as the quality of the predictor itself is the quality of the measure of uncertainty attached to that predictor. Bayesian prediction is based on the complete predictive distribution that accounts for the variability of model parameters.

The gaussian model was described. The importance of accounting for the uncertainty on the model parameters was highlighted through several examples and the application to the test case. For the gaussian random fields accounting for the uncertainty on the correlation parameter greatly influences the predictive distribution, while accounting for the uncertainty on the variance parameter does not have a large influence. The most critical point seems to be the behavior at small distances for which there is no available information. Although we would expect a larger variance for the case where all parameters are unknown it is not always the case.

The size of data sets in earth sciences is usually large- very large. This makes calculations time con-

suming. For example, the conditional simulations of the deposit had to be run overnight in batches of 20. The Bayesian approach is even heavier as it needs in addition to simulate the unknown parameters and the covariance matrix has to be inverted for each value of the correlation parameters.

As the data are not gaussian, the transformed gaussian model proposed by De Oliveira *et al.* (1997) was considered. The Box- Cox family of power transformations is combined with the Gaussian random field model. The Bayesian framework allows to easily take into account the uncertainty on the transformation. However, this model is difficult to use as the range of the transformed data depends on the transformation parameter. Another major difficulty of this model is the definition of the joint prior distribution for the model parameters. Moreover, the physical interpretation of the parameters is problematic.

In any case, we recommend the transformed gaussian model as a first tool to analyse non- gaussianity in geostatistical data. Then it would be possible either to use a simple transformation (as the log -transformation, which can be used for positive skewed data, although this hypothesis must be checked carefully) or, as Christensen *et al.* (2001) state, when the sampling mechanism suggests a specific non Gaussian distributional assumption, to incorporate this within the framework of the generalised linear mixed model. We underline that the Bayesian framework provides the optimal predictor for the log- normal model while it is not the case for kriging when the mean in unknown. The transformed gaussian model can help deciding whether the log- transformation, for example, is appropriate or not for the data under study. And more generally, it can avoid the selection of an incorrect transformation due, for example, to the influential effect of a few observations.

The most critical point of the Bayesian approach is the definition of the prior distribution. It has an important role, as expected, for small samples. Several prior distributions, either noninformative or informative, were defined and the sensitivity to the prior of the parameters posterior distribution and of the predictive distribution was checked. Note that noninformative distributions are often improper distributions: it is then necessary to verify whether the posterior distribution is a proper distribution or not. The transformed gaussian model seems sensitive to the prior that is chosen.

A more general Bayesian approach was defined to evaluate the reserves of the case study. The results were then compared to ones obtained with the plug- in approach. As it is usual practice in geostatistics, the gaussian anamorphosis function was used instead of the Box and Cox trasformation to transform the data in normally distributed values. The gaussian model was then defined for the transformed data. We recall that the hypotheses for the Bayesian framework are stronger than for kriging. The spatial distribution of the variable of interest must be specified. However, the approach we followed required, as for conditional simulations, only the gaussianity of a transformation of the random function under study. Both a constant mean and a quadratic vertical drift were introduced in the model. A noninformative prior was considered for the mean parameters. Several covariances were fixed a priori and were assumed equally likely. This permitted us to invert the covariance matrix only once in the sampling algorithm. The geometric anisotropy was easily allowed for in the model. For each Monte Carlo sample of the mean and covariance, classic geostatistical conditional simulations were carried out. The importance of taking into account the uncertainty on the covariance parameters was highlighted. The Bayesian and *plug- in* distributions of the reserves mostly differ in the tails.

In conclusion, we recommend the Bayesian approach, but it should be applied with some care as it is situation dependent, for example, for the definition of the prior distributions. We feel that this is a strength of the approach as it forces one to look critically at the problem under study.

The real options framework was presented for evaluating natural resource projects subject to both market and technical risk. Its strong point is that it can be used to value flexibility inherent in projects. Neglecting flexibility can lead to undervalue projects and consequently to refuse projects that could be profitable.

The risk- neutral valuation was considered. Gold price was assumed to follow a geometric Brownian motion. It is a non stationary process. It is true that for short periods of time, relative to the mean reversion time, it is difficult to differenciate a GBM from a mean- reverting process. We used a stochastic dynamic programming approach to carry out the computations as it allows us to value early exercise options. The binomial model was used to discretise the GBM.

The model allowed for initiation flexibility, operating flexibility and information flexibility. Initiation flexibility is found to have value. Operating flexibility has little value: as expected the more uncertain the reserves and the costs, the larger the value of flexibility. The importance of accounting for the uncertainty on the reserves was also highlighted comparing the Bayesian and *plug- in* approaches. They provide more or less the same results, that is the same optimal project. However, the value of information flexibility differ greatly. Our principal interest was information flexibility. Given the initial information the optimal decision was to delay the start of the project of 1 year and then develop the large pit. The value of the rich additional information was high as it allowed to well differenciate the two projects that is the development of the large pit is found to be much more interesting than the development of the small pit. The reduction of risk due to the average and poor additional information was smaller as the two projects were still close. The additional data reduced the risk by 40 % and so it is worthwhile to obtain it. Information has value not only if it brings to modify the initial decisions, but also if it allows to better define the project.

Sensitivity of the flexibility value to the initial gold price was checked. Initiation flexibility, that is the option to wait, has value for high gold prices. Information flexibility has little value for either very low or very high prices. That is it is not worthwhile to double the information on the lower part of the deposit if the gold price is either very low or very high.

The option to shift to the small pit if the lower part of the large pit is not interesting greatly enhances the project value.

The risk- neutral valuation should be used if a replicating portfolio can be defined or equivalently if cash-flows are assumed to be traded in the market. This is a dubious hypothesis when valuing real options as a mine (or oil) project has risks that are project specific and are not traded in the market. We considered a decision approach with a fixed discount rate representing the risk of the project. As expected the results are quite different from the results obtained with the risk- neutral approach: the development strategy is different. It is thus really important to correctly define the discount rate.

What is interesting about this thesis is that it combines geostatistical methods and financial tools to value a mine project. In particular, conditional simulations and Bayesian analysis were combined to obtain the reserves. We think that it is important to use a model that takes both technical and market uncertainty into consideration when evaluating natural resource projects. Moreover, the Bayesian approach we finally used is more general than the model proposed by De Oliveira *et al.* (1997). The difficulties for defining the prior distribution do not arise. And the range of the gaussian variable is well defined.

6.1 Perspectives for future work

We resume the principal limitations of this work and suggest alternatives that could have been explored.

The additional information was simulated. This permitted to work with different scenarios, but on the other side it is time consuming. For this we chose only 3 of the 100 simulated data sets and supposed they represented 25 %, 50 % and 25 % of the possible extra data sets. Moreover, in the place of fixing the location and number of extra holes, a strategy for the optimal location and number of extra holes could be defined. Alternatively, as Dias (2002), we could consider different scenarios for the reduction in variance of the reserves due to extra data. In this case we would have to carry out only the simulations conditional on the initial data set. However, note that additional data does not necessarily imply a reduction of the variance.

An important part of this work is dedicated to the choice of the model for prediction. The model would have been more general if a family of covariance functions was taken instead of choosing a specific model. The Matérn class of covariance functions which include the exponential model could have been considered. It would have permitted to account for different behaviours at the origin.

The Bayesian approach we have considered strongly relies on the gaussian assumption. For data that do not seem to follow a gaussian distribution but that still retain some broad similarity with it, the skew normal distribution (Azzalini and Dalla Valle, 1996) could be defined instead. This is a quite manageable distribution allowing ample flexibility in skewness and kurtosis. Thus when the histogram of the data presents a continuous and not too skewed distribution a gaussian transformation is appropriate. But, although the anamorphosis transformation can be applied to any continuous variable, if the histogram is skewed and has long tails a gaussian transformation will limit the amplitude of large values and magnify unimportant differences between low values. This could be interesting, if, for example, the proportion of zero values is important.

The model allowed for the uncertainty on the reserves and on the gold price. Unknown costs were considered to value the lower part of the large pit. Other sources of uncertainty could have been included in the model, as for example the rate of production and the cutoff. This would increase the complexity of the decision problem. In addition it could be interesting to introduce in the model the correlation between the production and the gold price or analogously between the cutoff and the gold price or between the cutoff and the costs. Higher prices lead to lower cutoff grades. And at higher prices the company could sell more and thus shorten the mine's life (Lane, 1991).

We assumed that the reserves were updated only if additional drilling was obtained. It could be interesting to consider continuous updating of the reserves that would be due to the production information. For this Connell (2002) introduces a GBM to model reserves uncertainty evolving over time. However, this implies that the reserves variance grows with time.

One drawback of the approach considered is that it does not address the question of the correct risk- adjusted discount rate. The discount rate should then account for both market and technical risks. A risk premium that accounts for private risk should be added to the risk- free rate. The approaches by Smith and Nau (1995) and Slade (2000), although different, define risk- adjusted rates.

In addition note that the hypothesis of a constant discount rate is acceptable when the life of the project is short (less than 1 year). Models that include stochastic discount rate could be used.

The approach considered could take into account market opportunities to hedge the project risks by trading securities. It could be interesting to value the test case assuming that, for example, 50 % of the annual production is hedged. The gold derivatives market has grown rapidly over the last decade (Neuberger, 2001). Mining companies are often hedging a certain proportion of their production forward so that financial institutions are sure debts will be repaid. That is mining companies use the derivative market to hedge their production and the principal reason is risk reduction or risk control. In practice, few companies sell more than a small fraction of their reserves forward. Following Neuberger forward sales amount to less than two years future production. Management knows that a strategy of forward selling (forward contract) which looks like prudent risk management over a period in which the gold price has fallen would look very foolish in a period in which the gold price has risen sharply. The full hedged producer derives no benefit from the improvement in the price of its main product. Another possibility for the mine company is to acquire a put option. Derivatives increase the flexibility of managing the reserves.

More information on the company exploiting the deposit could help in better defining the decision rule for the optimal project. We assumed that the decision maker was risk- neutral and thus the project with the highest expected profit was chosen even if more uncertain. A small company which is risk- adverse or conservative would prefer a more certain project rather than a project with higher value but with a greater uncertainty. On the contrary, a big company that has interests in more projects would not hesitate in choosing the project with the higher payoff although more risky. Utility theory should be used for decision analysis for a conservative risk attitude. Utility is a measure of value reflecting the preferences of the decision maker based upon beliefs. The decision rule is then to choose the alternative having the highest expected utility.

Appendix

Appendix A

Conditional simulations of a gaussian random function

The method used to simulate from a gaussian random function Y given observations of Y in certain points is called *conditioning by kriging*: a non conditional simulation is first obtained and it is then conditioned using kriging. The turning bands algorithm is used to carry out non conditional simulations. For a description of this algorithm see Matheron (1973) and Chilès and Delfiner (1999).

Let $\{Y(x), x \in \mathcal{D}\}$, $\mathcal{D} \subseteq \mathbb{R}^d$, be the gaussian random field of interest and $Y_1 = (Y(x_1), \ldots, Y(x_n))^T$ a set of n observations from a single realisation of this random field, where $x_i, i = 1, \ldots, n$ are known distinct locations in $\mathcal{D}$. We look for a conditional simulation of Y at points $x_0 = (x_{01}, \ldots, x_{0m})^T$ based on the data x_i, that is Y_0. Let m be the mean of Y and C its covariance matrix. Denote by m_0 and m_1 the mean of Y_0 and Y_1, respectively. Denote by $C_{00}^{m \times m}$ the covariance of Y_0, $C_{11}^{n \times n}$ the covariance of Y_1 and $C_{01}^{m \times n} = C_{10}^T$ the covariance between Y_0 and Y_1. It is known that the conditional distribution of Y_0 given Y_1 is gaussian (Mardia *et al.*, 1992). The mean of this distribution is

$$Y_0^* = E[Y_0|Y_1] = C_{01}C_{11}^{-1}Y_1 + (m_0 - C_{01}C_{11}^{-1}m_1),$$

that is the simple kriging predictor of Y_0 given Y_1 with known mean m and covariance C, and the kriging error $Y_0 - Y_0^*$ has zero mean and covariance

$$C_{00} - C_{01}C_{11}^{-1}C_{01}^T.$$

The fact that Y_0^* and $Y_0 - Y_0^*$ are independent suggests the following decomposition

$$Y_0 = Y_0^* + [Y_0 - Y_0^*].$$

The kriging error is unknown as Y_0 is unknown. *Conditioning by kriging* consists of simulating the error $Y_0 - Y_0^*$ and adding it to the kriging predictor Y_0^*. The algorithm is finally

- simulate the error W with zero mean and covariance C;
- compute W_0^* given W_1;
- add the simulated error $W_0 - W_0^*$ to Y_0^*.

The conditional simulation is then

$$T_0 = Y_0^* + [W_0 - W_0^*].$$

Since kriging is an exact interpolator, at a sample point x we have $Y^*(x) = Y(x)$ and $W^*(x) = W(x)$, so that $T(x) = Y(x)$. In the case of an unknown but constant mean the algorithm is exactly the same with ordinary kriging replacing simple kriging.

Appendix B

Grid parameters for simulations

	Direction X_1	Direction X_2	Direction X_3
Origin (m)	16995	23050	-200
Mesh size (m)	5	5	5
N nodes (m)	41	121	131
Min (m)	16995	23050	-200
Max (m)	17195	23650	450

Table B.1: *Grid parameters for the punctual simulations (where Min and Max indicate the area estimated)*

	Direction X_1	Direction X_2	Direction X_3
Origin (m)	16905	22912.5	-195
Mesh size (m)	10	25	10
N nodes (m)	65	32	65
Min (m)	16900	22900	-200
Max (m)	17550	23700	450

Table B.2: *Grid parameters for block simulations (where Min and Max indicate the area estimated)*

Appendix C

Bayesian analysis

The basic notions and tools of Bayesian analysis that are used in the thesis are introduced. Section 1 presents the notation. Section 2 describes the basic notions of Bayesian inference and the two widely used prior distributions: the noninformative and conjugate. Finally Section 3 presents the idea of Markov Chain Monte Carlo algorithms and in particolar Gibbs sampling.

C.1 Notation

Let u, v and w be three random variables or vectors. Let $f(\cdot)$ denote any density under study where the arguments identify the random variables in question. Then, $f(u, v)$, $f(u|v)$ and $f(u)$ denote respectively the joint density of $(u, v)^T$, the conditional density of u given v and the marginal density of u. The following properties will be repeatedly used

- the Bayes' Theorem
$$f(v|u, w) = \frac{f(u|v, w)\, f(v|w)}{f(u|w)};$$
- the decomposition of joint densities as products of conditional densities
$$f(u, v|w) = f(u|v, w)\, f(v|w);$$
- the computations of marginal densities from joint densities
$$f(u|w) = \int f(u, v|w)\, dv = \int f(u|v, w)\, f(v|w)\, dv.$$

We will often write
$$f(u|v, w) \propto h(u, v, w)$$
for some function $h(\cdot)$, where $h(u, v, w)$ is considered a function only of u, with $(v, w)^T$ fixed and the proportionality constant does not depend on u,
$$f(u|v, w) = c(v, w)\, h(u, v, w)$$
for some function $c(\cdot)$.

Let now Z denote the data and $\theta = \{\theta_1, \ldots, \theta_p\}$ denote the vector containing the unknown quantities: for example, model parameters and future observations. From a Bayesian view point there is no distinction between observables, Z, and unobservables, θ: all are considered random variables, and uncertainty at any stage is quantified in terms of probability density functions. Inference about the uncertain quantities θ requires setting up a probability model for the joint distribution $f(Z, \theta)$, which is usually determined by $f(Z|\theta)$, called the *likelihood function* when viewed as a function of θ, and $f(\theta)$, called the *prior distribution*. This distribution can be specified with the help of parameters which are called hyperparameters. If a hierarchical Bayesian analysis is considered then the uncertainty on the prior is modeled by a distribution on these parameters. After having observed the data, z, inference about θ is based on the conditional density of θ given $Z = z$, called the *posterior distribution*, which from Bayes' Theorem is given by

$$\begin{aligned} f(\theta|z) &= \frac{f(\theta)\, f(z|\theta)}{\int f(\theta)\, f(z|\theta)\, d\theta} = \frac{f(\theta)\, f(z|\theta)}{f(z)} \\ &\propto f(\theta)\, f(z|\theta). \end{aligned}$$

The posterior distribution combines the prior beliefs about θ with the information about θ contained in the sample z.

Note that the marginal posterior distribution of θ_j, $j = 1, \ldots, p$, is

$$f(\theta_j|z) = \int f(\theta_j, \theta^j|z)\, d\theta^j = \int f(\theta_j|\theta^j, z)\, f(\theta^j|z)\, d\theta^j$$

where $\theta^j = \{\theta_i, i \neq j\}$. The full conditional distribution of θ_j is

$$f(\theta_j|z, \theta^j) = \frac{f(\theta|z)}{\int f(\theta|z)\, d\theta_j} = \frac{f(\theta|z)}{f(\theta^j|z)}. \tag{C.1}$$

As we will see later in this chapter, this distribution plays an important role in the application of Markov chain Monte Carlo methods.

We will denote by $m(\cdot)$ the marginal density of Z, and by $\pi(\cdot)$ the density function of the model parameters.

The notions presented here come from Zellner (1971), Berger (1985) and Robert (1992).

C.2 Bayesian inference

Inference problems concerning θ can easily be dealt with using Bayesian analysis. The idea is that, since the posterior distribution contains all the available information about θ, both sample and prior information, any inferences concerning θ should consist solely of features of this distribution. Bayesian inference consists of reporting the entire posterior distribution $\pi(\theta|Z)$. And a visual inspection of the graph of this posterior provides the best insight concerning θ.

However, the simplest inferential use of the posterior distribution is to report a point estimate for θ with an associated measure of accuracy.

To estimate θ, a number of classical techniques can be applied to the posterior distribution. The most common is maximum likelihood estimation, which takes, as the estimate of θ the value $\hat{\theta}$ which maximizes the likelihood function $L(\theta|Z) \propto f(Z|\theta)$. The analogous Bayesian estimate is defined as follows.

Definition 1 *The generalized maximum likelihood estimate of θ is the mode of $\pi(\theta|Z)$, that is the value $\hat{\theta}$ which maximizes $\pi(\theta|Z)$.*

It is the most likely value of θ given the prior and the sample z. It is called the *maximum a posteriori*.

Other common Bayesian estimates of θ include the mean (which is optimal under squared error loss) and the median (which is optimal under absolute error loss) of $\pi(\theta|Z)$. In some cases these are better estimates of θ than the mode. For example, if the posterior is strongly skewed the median might be a more sensible measure of location.

The customary Bayesian measure of the accuracy of an estimate is the posterior variance of the estimate which is defined as follows

Definition 2 *If θ is a real valued parameter with posterior distribution $\pi(\theta|Z)$, and $\hat{\theta}$ is the estimate of θ, then the posterior variance of $\hat{\theta}$ is*

$$\text{Var}[\hat{\theta}|Z] = E[(\theta - \hat{\theta})^2|Z]. \tag{C.2}$$

When $\hat{\theta}$ is the posterior mean, $\hat{\theta} = E[\theta|Z]$, then $\text{Var}[\hat{\theta}|Z] = \text{Var}[\theta|Z]$ which is called the posterior variance. The posterior mean minimizes (C.2) and is the estimate with smallest standard error. For this reason, generally the posterior mean is used as the estimate of θ and the posterior standard deviation as the standard error.

C.2.1 Prior distributions

The most critical and criticized point of Bayesian analysis is the definition of prior distributions. The prior distribution is the key to Bayesian inference. But it seldom occurs that the available prior information is precise enough to lead to an exact determination of the prior distribution. The prior is a tool summarizing available information as well as uncertainty related with this information and ungrounded prior distributions produce unjustified posterior inference.

Noninformative priors

Bayesian analysis is used even when no prior information is available. In these situations a noninformative prior is needed. A noninformative prior is a prior which "favors" no possible values of θ over others. A noninformative prior is frequently an improper prior, that is, it has infinite mass.

Various suggestions have been advanced for determining a noninformative prior. The most widely used method is that of Jeffreys which is based on the expected Fisher information,

$$I(\theta) = E\left[\frac{\partial \log f(Z|\theta)}{\partial \theta}\right]^2.$$

This, under commonly satisfied assumptions, is given by

$$-E\left[\frac{\partial^2 \log f(Z|\theta)}{\partial \theta^2}\right].$$

The corresponding law is

$$\pi(\theta) = [I(\theta)]^{1/2}$$

in the 1 dimensional case. If θ is a vector Jeffreys suggests the use of

$$\pi(\theta) = |I(\theta)|^{1/2}$$

where $I(\theta)$ is the Fisher information matrix and $|I(\theta)|$ is its determinant. For example, if Z is a $N(\beta, \sigma^2)$ and $\theta = (\beta, \sigma)$ is unknown, then

$$\pi(\theta) \propto \frac{1}{\sigma^2}$$

which is improper as it does not integrate to one. Alternatively, β and σ can be assumed independent *a priori* and the one- dimensional Jeffreys prior for each of the parameters can be used. This gives

$$\pi(\theta) \propto \frac{1}{\sigma} \tag{C.3}$$

which is the noninformative prior recommended by Jeffreys. It is standard practice to use (C.3) as the non-informative prior for a location-scale problem (Berger, 1985).

A proper posterior distribution can be obtained even if $\pi(\theta)$ is an improper prior.

Jeffreys' prior is invariant under reparametrisation. If we have no information about θ, we also have no information about, for example, $1/\theta$. In addition, the choice of this law is justified by the fact that $I(\theta)$ is accepted as a measure of the quantity of information contained in the model on θ. The prior must give a larger weight to the values of θ with a larger $I(\theta)$. This is equivalent to minimize the influence of the prior and is thus as noninformative as possible.

Often the results, estimate and standard error, from Bayesian analysis with a noninformative prior coincide with the results from classical approaches (maximum likelihood, for example). Of course the interpretations differ, but the values are the same. It is argued that when the two approaches give the same results, it consolidates or gives strength to the classical approaches, but when they differ, the classical approach suffers in the comparison.

Conjugate families

In general, $m(Z)$ and $\pi(\theta|Z)$ are not easily calculable. As Berger states, a large part of the Bayesian literature is devoted to finding prior distributions for which $\pi(\theta|Z)$ is easily calculated. These are the conjugate priors.

Definition 3 *Let $\mathcal{F}$ denote the class of density functions $f(Z|\theta)$. A class $\mathcal{P}$ of prior distributions is said to be a conjugate family for $\mathcal{F}$ if $\pi(\theta|Z)$ is in the class $\mathcal{P}$ for all $f \in \mathcal{F}$ and $\pi \in \mathcal{P}$.*

For a given class of densities $\mathcal{F}$, a conjugate family can be determined by examining the likelihood functions $L(\theta|Z)$, and choosing as a conjugate family, the class of distributions with the same functional form as these likelihood functions. The resulting priors are called natural conjugate priors. Conjugate priors allow to begin with a certain functional form for the prior and end up with a posterior of the same functional form, but with parameters updated by the sample information. This provides an easy way of seeing the effect of prior and sample information.

In general, when dealing with conjugate priors there is no need to explicitly calculate $m(Z)$. The reason is

that since $\pi(\theta|Z) = \pi(\theta)\, f(Z|\theta)/m(Z)$, the factors involving θ in $\pi(\theta|Z)$ must be the same as the factors involving θ in $\pi(\theta)\, f(Z|\theta)$. Hence it is only necessary to look at the factors involving θ in $\pi(\theta)\, f(Z|\theta)$, and see if these can be recognised as belonging to a particular distribution. If so, $\pi(\theta|Z)$ is that distribution.

The use of conjugate priors should not be justified only by their attractive properties but if a conjugate prior can be chosen which gives a reasonable approximation to the true prior. However, many Bayesians argue that in dealing, for example, with a normal mean, the class of normal priors is rich enough to include approximations to most reasonable priors. Berger doubts this belief, observing that using a normal prior can sometimes result in unappealing conclusions. Moreover, in order to overcome problems with improper priors, proper priors, and in particular conjugate priors, are often used as approximations to improper priors: a normal distribution with a large variance, for example. This is not always a good solution as the prior can dominate the data. In conclusion, the prior should be chosen according to the problem at hand. In Berger (1985; chapter 4.7) the robustness (or sensitivity) of Bayesian analysis to possible misspecification of the prior distribution is studied and examples are presented to show the importance of the prior distribution. Several robust Bayesian techniques are introduced. To investigate robustness one simple technique is to try different reasonable priors and see what happens. This is called *sensitivity analysis*.

Generally if, conjugate priors are not, or cannot, be used, the computation of the posterior law, or of its mean and variance, cannot be done analytically.

C.3 Markov chain Monte Carlo

The notions presented here come from Gilks *et al.* (1996) and Robert (chapter 6 in Droesbeke *et al.*, 2002). In the following sections we present the tools we used and are widely used in Bayesian analysis.

C.3.1 Monte Carlo integration

Monte Carlo methods are used to approximate an integral of the form

$$\int_{\Theta} g(\theta)\, \pi(\theta|z)\, d\theta = E_{\pi}[g(\theta)]. \tag{C.4}$$

As $\pi(\theta|z)$ is a density function, the idea is to generate $\theta_1, \ldots, \theta_l$ independently from π. The large numbers law ensures that

$$\frac{1}{l} \sum_{i=1}^{l} g(\theta_i) \;\rightarrow\; \int g(\theta)\, \pi(\theta|z)\, d\theta$$

as $l \rightarrow \infty$. It is not necessary to simulate from $\pi(\theta|z)$ to approximate (C.4). Let h be an arbitrary density function on Θ, the *importance sampling algorithm* (Geweke, 1989) consists of

- generate $\theta_1, \ldots, \theta_l$ from h and
- approximate (C.4) by

$$\frac{\sum_{i=1}^{L} g(\theta_i)\, w(\theta_i)}{\sum_{i=1}^{L} w(\theta_i)}$$

  where $w(\theta_i) = \pi(\theta_i)\, f(z|\theta_i)/h(\theta_i)$.

As

$$\int_\Theta g(\theta)\, f(z|\theta)\, \pi(\theta)\, d\theta = \int_\Theta \frac{g(\theta)\, f(z|\theta)\, \pi(\theta)}{h(\theta)}\, h(\theta)\, d\theta$$

the large numbers law ensures that this approximation converges to (C.4). And thus

$$E_\pi[g(\theta)|z] = \frac{\int g(\theta)\, f(z|\theta)\, \pi(\theta)\, d\theta}{\int f(z|\theta)\, \pi(\theta)\, d\theta} \approx \frac{\sum_i g(\theta_i)\, w(\theta_i)}{\sum_i w(\theta_i)}. \tag{C.5}$$

The choice of the importance function h is crucial: it must be easy to simulate and it must be relatively close to $\pi(\theta|z)$ to reduce the variability of (C.5) and for the weights $w(\theta_i)$ to be not too low. Note that the tails of h should be at least as heavy as the tails of $\pi(\theta|z)$ for the convergence to be not too slow.

In general drawing samples $\{\theta_i\}$ independently from $\pi(\cdot)$ is not feasible and determining or choosing $h(\cdot)$ is not easy. However the $\{\theta_i\}$ need not necessarily be independent. The $\{\theta_i\}$ can be generated by any process which draws samples throughout the support of $\pi(\cdot)$ in the correct proportions. This can be done through a Markov chain having $\pi(\cdot)$ as its stationary distribution. This is then *Markov chain* Monte Carlo. We start by introducing the 2 most used MCMC algorithms: the Gibbs sampler and the Metropolis- Hastings algorithm. The main theoretical concepts of the markovian simulation method are then described. The problem is to simulate a random variable $\mathbf{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(l)})$ with distribution $\pi(\cdot)$ within the state space E. As there is no known efficient direct method to solve this problem an iterative method has been built.

C.3.2 The Gibbs sampler

It was given its name by Geman and Geman (1984) and it originated in the context of image restoration. The Gibbs sampler permits to generate random variables that follow approximately the distribution of interest $\pi(\cdot)$ so that the importance function h is obtained directly from $\pi(\cdot)$.

The Gibbs sampler succesively and repeatedly simulates from the conditional distributions of each component given the other components, that is

Algorithm 1

1. propose a starting value $\mathbf{x_0} = (x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(l)})$;
2.
   - simulate $x_1^{(1)}$ from the conditional $(x^{(1)}|x_0^{(2)}, \ldots, x_0^{(l)})$;
   - simulate $x_1^{(2)}$ from the conditional $(x^{(2)}|x_1^{(1)}, x_0^{(3)}, \ldots, x_0^{(l)})$;
   - ...
   - simulate $x_1^{(l)}$ from the conditional $(x^{(l)}|x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(l-1)})$;
3. iterate this procedure.

This means that the full conditionals are completely known and can be sampled from. The realisations $\mathbf{x_1}$, $\mathbf{x_2}$, ... obtained from iterating step 2 are simulated from a Markov chain. Under weak regularity conditions[1], convergence of this chain to the stationary distribution $\pi(x^{(1)}, \ldots, x^{(l)})$ is guaranteed, so that for

[1] If $\pi(X^{(i)}|X^{(j)}) > 0$ over E, $j \neq i$, $\{X_{(n)}\}$ is an ergodic Markov chain with stationary distribution π (Robert, p. 168 in Droesbeke *et al.*, 2002)

sufficiently large k, the values $(x_k^{(1)}, \ldots, x_k^{(l)}), \ldots, (x_n^{(1)}, \ldots, x_n^{(l)})$ can be regarded as realisations from this distribution. The initial value $\mathbf{x_0}$ has no importance.

In our case the vector $\mathbf{X}$ represents the parameter vector or the vector of predictions. The application of the Gibbs sampler is natural in Bayesian problems which exhibit conditional conjugacy, that is the full conditional distributions are available. Often it is not possible to sample directly from these conditional distributions, in this case the following Metropolis- Hastings algorithm is used in the Gibbs sampler (*Metropolis within Gibbs*).

It must be observed that if the components $X^{(1)}, X^{(2)}, \ldots, X^{(l)}$ exhibit heavy dependence, the algorithm will converge slowly and give highly serially correlated output.

C.3.3 The Metropolis- Hastings algorithm

Let $q(x, y)$ be any arbitrary transition probability, that is $q(x, y)$ is the probability density of moving to y from x, and from which simulation is direct. The Metropolis- Hastings algorithm is

Algorithm 2

1. propose a transition $x \to y$ following the $q(x, \cdot)$ law;
2. calculate $a(x, y) = \min\left\{1, \frac{\pi(y)\, q(y,x)}{\pi(x)\, q(x,y)}\right\}$, then accept the proposal y with probability $a(x, y)$, otherwise keep x;
3. iterate this procedure.

The advantage of this algorithm over the Gibbs sampler is that it is not necessary to know all the conditional distributions. However, the transition mechanism now depends on a proposed transition q and a subsequent step of evaluation of this proposal. If q is poorly chosen, then either the proportion of rejections can be high, or the constructed Markov chain might move around the space too slowly giving a very low efficiency of the Monte Carlo estimation. A useful monitoring device of the method is given by the average percentage of iterations for which moves are accepted. The success of the method depends on having a reasonable acceptance rate. On the contrary, making the chain move very slowly so that the acceptance probability is close to 1, makes the chain take many iterations to converge as the whole parameter space must be visited. Note that a particular choice of q might work well on one target density, but be extremely poor on another.

The Metropolis- Hastings framework is very general since it does not impose any restriction on the form of q except some weak regularity conditions that are required to ensure irreducibility and aperiodicity. All currently used MCMC algorithms can be expressed in this form.

Examples

Canonical forms of proposal distributions are

- *the independence sampler*: the simplest choice for the proposal distribution is $q(x, y) = q(y)$, that does not depend on x. This gives

$$a(x, y) = \min\left\{1, \frac{w(y)}{w(x)}\right\},$$

where $w(x) = \pi(x)/q(x)$ (it is the importance weight function that would be used in importance sampling to weight a sample from q toward the distribution π).

- *The Metropolis dynamic*: $q(y,x) = q(x,y) \ \forall x, y$ is a symmetric proposal and gives

$$a(x,y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}.$$

For example, $q(\cdot, y)$ might be a multivariate normal distribution with mean x and a constant covariance matrix.

- *Single- component Metropolis- Hastings*: it is often more convenient and computationally efficient to divide X into components, $X = \{X^{(1)}, X^{(2)}, \ldots, X^{(d)}\} = \{X^{(i)}; i = 1, \ldots, d\}$, and update these components one by one. Let X^j be the variable composed with all the $X^{(i)}$ components of X except for the jth one, $X^j = \{X^{(i)}; i = 1, \ldots, d; i \neq j\}$. The candidate $y^{(j)}$ is generated from a proposal distribution $q_j(y^{(j)}|x^{(j)}, x^j)$, where in x^j the components $1, \ldots, j-1$ have been already updated. The jth proposal distribution q_j generates a candidate only for the jth component of X, ad may depend on the current values of the components of X. This gives

$$a(x^j, x^{(j)}, y^{(j)}) = \min\left\{1, \frac{\pi(y^{(j)}|x^j)\, q(x^{(j)}|y^{(j)}, x^j)}{\pi(x^{(j)}|x^j)\, q(y^{(j)}|x^{(j)}, x^j)}\right\}$$

where $\pi(x^{(j)}|x^j)$ is the full conditional distribution of $X^{(j)}$ (it was given in (C.1)).

- *Gibbs sampling*: it is a special case of the single- component Metropolis- Hastings. The proposal distribution for updating the jth component of X is $q_j(y^{(j)}|x^{(j)}, x^j) = \pi(y^{(j)}|x^j)$. It gives an acceptance probability of 1.

C.3.4 Markov chains simulations

We follow the introduction to general state space Markov chain theory by Tierney (chapter 4 in Gilks *et al.*, 1996). The distribution π is defined on a set E. Often $E = \mathcal{R}^d$. The only technical requirement is that the collection $\mathcal{E}$ of subsets of E on which π is defined must be a countably generated sigma algebra. All subsets of E and all functions defined on E are assumed to be $\mathcal{E}$- measurable.

Definition 4 *A Markov chain $\{X_k; k \geq 0\}$ on the state space E is specified by its initial state X_0 and its transition kernel Q,*

$$Q(x, A) = P(X_k \in A | X_{k-1} = x)$$

for all $x \in E$ and $A \subset E$. That is, $Q(x, \cdot)$ is the distribution of the Markov chain after one step given that it starts at x.

For a probability distribution ν on E, define the distribution νQ by

$$\nu Q(A) = \int Q(x, A)\nu(dx)$$

which is the distribution of the Markov chain at time k if the transition kernel is Q and the distribution at time $k-1$ is ν. A Markov chain has *invariant (or stationary) distribution* π if $\pi Q = \pi$. The product PQ of two transition kernels P and Q is the transition kernel defined by

$$PQ(x, A) = \int P(x, dy)\, Q(y, A)$$

for all $x \in E$ and $A \subset E$. Using this notation we can write

$$P(X_k \in A | X_0 = x) = Q^k(x, A)$$

for any $k \geq 0$. To study the convergence of the chain the total variation distance for two probability measures ν_1 and ν_2,

$$||\nu_1 - \nu_2|| = 2 \sup_A (\nu_1(A) - \nu_2(A)),$$

is introduced. If $Q^k(x, \cdot) \to \pi(\cdot)$ for π- almost all x, as $k \to \infty$, under the total variation distance, then the chain is called *ergodic* and it admits π as its stationary law. Ergodicity implies invariance of π

$$\int_E Q(x, A)\pi(dx) = \pi(A),$$

that is starting from any value generated from π, the probability of getting to A after a transition with K is $\pi(A)$. A sufficient, but not necessary, condition for a transition kernel to produce Markov chains that are invariant is *reversibility*. Reversibility implies that[2]

$$Q(x, dy)\, \pi(dx) = Q(y, dx)\, \pi(dy)$$

which balances the probability of going from x to y with that of going from y to x. Reversibility implies that π is a stationary distribution since

$$\pi Q(dy) = \int Q(x, dy)\, \pi(dx) = \int Q(y, dx)\, \pi(dy) = \pi(dy).$$

Definition 5 *A Markov chain is φ-irreducible for a probability distribution φ on E if $\varphi(A) > 0$ implies that $P_x(\tau_A < \infty) > 0$ for all $x \in E$ where τ_A is the time of the first return to A. A chain is irreducible if it is φ-irreducible for some probability distribution φ.*

Irreducibility means that all interesting sets can be reached. This is equivalent to imposing the existence of an integer n such that $Q^n(x, A) > 0$. Irreducibility implies the unicity of the invariant distribution π

Theorem 1 *If the Markov chain is irreducible and has invariant distribution π, then the chain is π-irreducible and π is the unique invariant distribution of the chain.*

It can be shown that irreducibility implies convergence of the proportion of time spent in a set A to $\pi(A)$. Stronger distributional results are possible but the additional condition of aperiodicity is necessary. An m-cycle for an irreducible chain with transition kernel Q is a collection $\{E_0, \ldots, E_{m-1}\}$ of disjoint sets such that $Q(x, E_j) = 1$ for $j = i + 1 \bmod m$ and all $x \in E_i$. The period d of the chain is th largest m for which an m- cycle exists. The chain is aperiodic if $d = 1$. Irreducibility and aperiodicity imply ergodicity.

[2] $Q(x, dy)$ is the probability of moving to a small measurable subset $dy \in E$ given that the move starts at x

Theorem 2 (Ergodic theorem) *If X is irreducible, aperiodic with transition kernel Q and invariant distribution π, then X is ergodic.*

The Ergodic theorem does not give any information on the convergence rate: how close the marginal distribution π^n of the current iteration n is to the target distribution $\pi()$? We will not discuss this issue, although extremely important.

The simulation of a law π using Markov chains consist of constructing a Markov chain with stationary distribution π. Now the samples we obtain from π by the means of a Markov chain are not necessarily independent. However, under ergodicity the convergence of the Monte Carlo integral is preserved. If X is an irreducible Markov chain with invariant law π, then using the Ergodic theorem we have

$$\frac{1}{l}\sum_{i=1}^{l} g(x_i) \to E_\pi[g]$$

as $l \to \infty$.

Example: Metropolis- Hastings dynamic

Let q be an irreducible and aperiodic transition law on E. Let a be the acceptance function, $E \times E \to (0,1]$: $(x,y) \to a(x,y)$, where $a(x,y)$ describes for a couple of states (x,y) the acceptance probability of the transition $x \to y$.

The acceptance- rejection algorithm works in two steps

Algorithm 3

1. propose a transition $x \to y$ following the $q(x,\cdot)$ law;
2. accept this transition with the probability $a(x,y)$, otherwise stay at x.

The transition kernel Q of this algorithm is thus

$$Q(x,y) = \begin{cases} q(x,y)\,a(x,y) & \text{if } x \neq y \\ q(x,x) + \int q(x,y)\,(1-a(x,y))dy = 1 - \int q(x,y)\,a(x,y)dy & \text{otherwise} \end{cases}$$

and there are two independent random variables to draw: one following the transition law q on E and the second following an uniform law U on $[0,1]$. If $U \leq a(x,y)$ the transition $x \to y$ is accepted. Otherwise there is no transition.

For the transition kernel Q to be reversible, the acceptance function $a(x,y)$ must verify

$$\forall x, y \in E, x \neq y : \pi(x)\,q(x,y)\,a(x,y) = \pi(y)\,q(y,x)\,a(y,x),$$

or equivalently

$$\frac{a(x,y)}{a(y,x)} = \frac{\pi(y)\,q(y,x)}{\pi(x)\,q(x,y)}.$$

Some classic acceptance laws are

- *the Barker dynamic*: $a(x,y) = \frac{\pi(y)\, q(y,x)}{\pi(x)\, q(x,y) + \pi(y)\, q(y,x)}$;

- *the Metropolis- Hastings dynamic*: $a(x,y) = \min\left\{1, \frac{\pi(y)\, q(y,x)}{\pi(x)\, q(x,y)}\right\}$.

As q is irreducible and aperiodic, so is Q[3].

[3]If the support of q includes the support of π, the chain is irreducible, ergodique and has stationary distribution π (Robert, p. 164 in Droesbeke *et al.*, 2002)

Appendix D

Notions of finance

Section 1 presents the Wiener process and a broad class of continuous- time stochastic processes, called Ito processes. Section 2 introduces the main concepts about Options and describes the Black and Scholes' formula to value a European call option. These notions are fundamental to follow Chapter 5.

D.1 Wiener process

A Wiener process (or a **Brownian motion**) z_t is a continuous time stochastic process with the following properties

1. $z_0 = 0$;
2. $z_t \sim N(0, t)$, $\forall t \geq 0$. That is, for each t the random variable z_t is normally distributed with mean $E[z_t] = 0$ and variance $\text{Var}[z_t] = E[z_t^2] = t$.
3. The increments Δz_t on non-overlapping time intervals are independent. They are normally distributed with a mean of 0 and a variance that increases linearly with the time interval, $\text{Var}[\Delta z_t] = \Delta t$. That is the increments Δz_t can be obtained from standard normally distributed random variables ϵ_t

$$\Delta z_t = \epsilon_t \sqrt{\Delta t}.$$

   By letting Δt become infinitesimally small, we can represent the increment of a Wiener process, dz, in a continuous time as $dz = \epsilon \sqrt{dt}$. It must be noted that a Wiener process is non stationary.

A Wiener process, z_t, is then a Markov process: the probability distribution for all future values of the process depends only on its current value and is unaffected by past values of the process or by any other information, that is the current value of the process is all one needs to make the best forecast of its future value.

Brownian motion with drift

It is represented by

$$dx = \alpha\, dt + \sigma\, dz$$

where dz is the increment of a Wiener process, α is called the drift parameter and σ the variance parameter or volatility. Over any finite interval, Δt, the change in x, Δx, is gaussian with mean $\alpha\,\Delta t$ and variance $\sigma^2\,\Delta t$. It can be observed that the trend is the dominant determinant of Brownian motion, whereas in the short run, the volatility of the process dominates.

Generalized Brownian motion- Ito processes

An Ito process, x_t, is a continuous time stochastic process represented by

$$dx = \alpha(x,t)\,dt + \sigma(x,t)\,dz \tag{D.1}$$

where dz is the increment of a Wiener process, $a(x,t)$ and $b(x,t)$ are known (non random) functions. The drift and variance coefficients are functions of the current state and time.

Since $E[dz] = 0$, $E[dx] = \alpha(x,t)\,dt$. The variance of dx is equal to $E[dx^2] - (E[dx]^2)$, which contains terms in dt, in $(dt)^2$ and in $(dt)(dz)$, which is of order $(dt)^{3/2}$. For dt infinitesimally small, terms in $(dt)^2$ and $(dt)^{3/2}$ can be ignored, and to order dt the variance is $\text{Var}[dx] = \sigma^2(x,t)\,dt$.

Geometric Brownian motion

A special case of equation (D.1) is the *geometric Brownian motion with drift*

$$dx = \alpha\,x\,dt + \sigma\,x\,dz \tag{D.2}$$

Here, $\alpha(x,t) = \alpha\,x$ and $\sigma(x,t) = \sigma\,x$, where α and σ are constants.

The percentage changes in x, dx/x, are normally distributed. Since these are changes in the natural logarithm of x, absolute changes in x, dx, are lognormally distributed. The expected value of x at time t, if currently $x(0) = x_0$, is given by

$$E[x_t] = x_0\,e^{\alpha\,t}$$

and the variance is given by

$$\text{Var}[x_t] = x_0^2\,e^{2\alpha t}\,(e^{\sigma^2 t} - 1).$$

Figure D.1 shows three sample paths of equation (D.2), with a drift rate of $\alpha = 0.08$, that is 8% per year, and $\sigma = 0.18$, that is 18% per year. The time interval is 1 month. Then x_t is calculated using the equation

$$x_t = \left(1 + \frac{\alpha}{12}\right)\,x_{t-1} + \frac{\sigma}{\sqrt{12}}\,x_{t-1}\,\epsilon_t$$

where ϵ_t is drawn from a standard normal distribution. The starting value is $x_{50} = 100$. The trend line is also presented.

It can be observed that a gold price that follows a geometric Brownian motion tends to move away from its initial value (if the drift rate is not zero). The variance increases with time, giving a higher probability for high values. Figure D.2 presents 50 realisations of this model.

Mean-reverting process

Although it can be argued that the prices of commodities- such as copper or oil- should be related to long-run marginal production costs, they are often modelled as geometric Brownian motions, which tend to wander far from their starting points. In fact, while in the short run the price of oil, for example, might fluctuate randomly up and down, in the longer run it ought to be drawn back towards the marginal cost of producing oil. Thus the price of oil should be modelled instead as a *mean reverting process*. The simplest mean-reverting process, the *Ornstein-Uhlenbeck process*, is

$$dx = \eta\,(\overline{x} - x)\,dt + \sigma\,dz \tag{D.3}$$

where η is the speed of reversion (and controls the correlation distance for the prices) and $\overline{x}$ is the level to which x tends to revert (if x is a commodity price, then $\overline{x}$ might be the long-run marginal cost of production of this commodity). This process, although satisfying the Markov property, does not have independent increments[1]. If the value of x is currently x_0, then its expected value at any future time t is

$$E[x_t] = \overline{x} + (x_0 - \overline{x})\,e^{-\eta t}$$

and the variance of $(x_t - \overline{x})$ is

$$\text{Var}[x_t - \overline{x}] = \frac{\sigma^2}{2\eta}\,(1 - e^{-2\eta t}).$$

It can be observed that as t becomes large the expected value of x_t converges to $\overline{x}$ and the variance converges to $\frac{\sigma^2}{2\eta}$. In addition

- as $\eta \to \infty$: $\text{Var}[x_t] \to 0$, which means that x can never deviate from $\overline{x}$; and

- as $\eta \to 0$: x becomes a simple Brownian Motion, and $\text{Var}[x_t] \to \sigma^2 t$.

Figure D.3 shows three realisations of an Ornstein-Uhlenbeck model with initial value $x_{50} = 100$ and mean of 130. This model is calmer than the geometric Brownian motion. The volatility gives the amplitude of the fluctuations. Figure D.4 presents 50 realisations of this model with scale parameter equal to 5 years.

It can be observed that equation (D.3) is the continuous-time version of the first-order autoregressive process in discrete time. Specifically equation (D.3) is the limiting case as $\Delta t \to 0$ of the following autoregressive process of order 1 (AR(1))

$$x_t - x_{t-1} = \overline{x}\,(1 - e^{-\eta}) + (e^{-\eta} - 1)\,x_{t-1} + \epsilon_t$$

where ϵ_t is normally distributed with mean zero and standard deviation σ_ϵ and

$$\sigma_\epsilon = \frac{\sigma^2}{2\eta}(1 - e^{-2\eta}).$$

[1]The expected change in x depends on the difference between x and $\overline{x}$. If x is greater (less) than $\overline{x}$, it is more likely to fall (rise) over the next short interval of time

Ito's Lemma

The Ito process of equation (D.1) is continuous in time, but is not differentiable in the usual sense. To differentiate or integrate functions of Ito processes we need to make use of Ito's Lemma.

Suppose that $x(t)$ follows the process of equation (D.1), and consider a function $F(x,t)$ that is at least twice differentiable in x and once in t. The total differential of this function is

$$dF = \frac{\partial F}{\partial t}\,dt + \frac{\partial F}{\partial x}\,dx + \frac{1}{2}\,\frac{\partial^2 F}{\partial x^2}\,(dx)^2 \tag{D.4}$$

where $(dx)^2 = \sigma^2(x,t)\,dt$.

Example: Geometric Brownian motion

Let $F(x) = \log(x)$. Since $\partial F/\partial t = 0$, $\partial F/\partial x = 1/x$ and $\partial^2 F/\partial x^2 = -1/x^2$, we have from equation (D.4)

$$\begin{aligned} dF &= \frac{1}{x}\,dx - \frac{1}{2x^2}\,(dx)^2 \\ &= \alpha\,dt + \sigma\,dz - \frac{1}{2}\sigma^2\,dt = (\alpha - \frac{1}{2}\sigma^2)\,dt + \sigma\,dz, \end{aligned}$$

so over a finite time interval t, the change in the logarithm of x is normally distributed with mean $(\alpha - \frac{1}{2}\sigma^2)t$ and variance $\sigma^2 t$. The exact formula for x_t is then

$$x_t = x_0\, e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma z_t}.$$

The discrete time analog is

$$x_t = x_{t-1}\, e^{(\alpha - \frac{1}{2}\sigma^2) + \sigma\epsilon}$$

where ϵ is a white noise process consisting of independent identically distributed standard gaussian random variables. Thus

$$F_t - F_{t-1} = (\alpha - \frac{1}{2}\sigma^2) + \sigma\epsilon \tag{D.5}$$

is a white noise process with mean $\delta = \alpha - \frac{1}{2}\sigma^2$, whereas F_t is a random walk process[2] with drift δ. Using standard statistical procedures the parameters δ and σ may be estimated from discrete data points.

An important difference between the two stochastic processes is the dependence of the state space on the number of periods. While the maximum state space for the approximated Ornstein-Uhlenbeck process is independent of the number of stages, the random walk representation of the geometric Brownian motion implies that the state space grows linearly with an increase in stages.

D.2 Options

We now describe some basic concepts in options theory. The notions presented in this section are reproduced according to Hull (1993).

[2] A random walk is the discrete analogue of a Wiener process

A *derivative security* (or contingent claim) is a security whose value depends on the values of other more basic underlying variables. An option is a derivative security. The underlying assets include stocks, stock indices, foreign currencies, debt instruments, commodities, and future contracts. Options are contracts.

There are two basic types of options. A *call option* gives the holder the right to buy the underlying asset by a certain date for a certain price. A *put option* gives the holder the right to sell the underlying asset by a certain date for a certain price. The price in the contract is known as the *exercise* price (or *strike price*); the date in the contract is known as *expiration* date (or *maturity*). If the option can be exercised before maturity, it is called an *American* option; if only at maturity, a *European* option. The returns from an option are asymmetric: the holder has the right to do something, but he does not have to exercise this right. This distinguishes options from forwards and futures contracts, where the holder is obligated to buy or sell the underlying asset. Note that to enter an option contract an investor must pay, while it costs nothing to enter into a forward or futures contract. The final outcome of a simple option can be a net loss of the limited cost of acquiring the option. The key feature of an option is that the cost of exercising it is defined in advance and does not depend on later conditions. It is in this respect that an option has value. Options are attractive because they offer the prospect of high gains with limited losses.

Option analysis consists of a set of procedures for calculating the value of options. Essentially, they estimate the expected value of the asymmetrical distribution of possible outcomes associated with options.

The decision whether to exercise an option or not is reflected by the formula

$$C = \max(S - E, 0),$$

where C is the value of the option, S is the current price of the underlying asset and E is the exercise price. If the exercise price of a call option is higher than the stock price at the moment of decision ($E > S$), the option will be worthless and will not be exercised. Unlike call options, which increase in value with favorable movements in the underlying asset price, put options -just like an insurance- pay off when the asset drops in value (Trigeorgis, 1996). It is assumed that the decision maker can cash the value immediately.

Financial options are usually priced using some variant of the CCA that was developed by Black and Scholes (1973) and Merton (1973). The idea is to value a financial option as part of a riskless portfolio. For example, suppose that the underlying security is a stock, it is possible to take a *long position* in the derivative security (the stock option) and a *short position* in the underlying asset (the stock). That is with a short (long) position in the futures market, the investor promises to sell (buy) an asset at a fixed price (the futures price) on a fixed date. Since both positions are affected by the same source of uncertainty (the stock price), the capital gains associated with one investment are exactly offset by the losses associated by the other. The rate of return on the portfolio is thus riskless and should therefore equal the risk-free rate. The building block of the Black and Scholes method is a partial differential equation that relates the expected future value of the option to the price of the underlying asset and the riskless rate (Hull, 1993, p.219). It is assumed that the price of the underlying asset follows a geometric Brownian motion.

The Black and Scholes' formula for a European call option

Black and Scholes showed that the value of a European call option of maturity T on a non-dividend paying stock, $C(S, t; E)$, satisfies the following partial differential equation

$$\frac{\partial C}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0 \tag{D.6}$$

where $C(T) = \max(S - E, 0)$, obtained constructing a riskless hedge portfolio

$$SN - C = B \tag{D.7}$$

by selling short one call option, purchasing $N = \partial C/\partial S$ shares of the underlying stock at the price S and borrowing B at the riskless rate. The value of the call option is

$$C(S, t; E) = S\mathcal{N}(d_1) - Ee^{-r(T-t)}\mathcal{N}(d_2)$$

where

$$d_1 = \frac{\ln(S/E) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{(T - t)}},$$

$d_2 = (d_1 - \sigma\sqrt{(T - t)})$, r is the risk-free rate of return (continuously compounded), $(T - t)$ the time to option expiration, and $\mathcal{N}()$ is the cumulative standard normal distribution function. The call option is equivalent to a levered position in the stock where the number of shares of the stock held in the replicating portfolio, N, is here given by $\mathcal{N}(d_1)$, and the amount borrowed is $B = Ee^{-r(T-t)}\mathcal{N}(d_2)$. N and B fluctuate continuously with the underlying stock price and with time and require frequent adjustment to maintain the above equivalence. If the future return of the stock is certain, that is σ^2 equals zero, the Black and Scholes formula becomes

$$C = S - Ee^{-r(T-t)}.$$

Now the value of the call equals the current stock price minus the present value of the exercise price. When the future price is uncertain, the Black and Scholes formula will use the weights $\mathcal{N}(d_1)$ and $\mathcal{N}(d_2)$ for S and $Ee^{-r(T-t)}$, respectively.

One key property of the Black and Scholes differential equation is the **risk-neutral valuation**: the equation does not involve variables that are affected by the risk preferences of investors. In particular, it does not involve the expected return on the stock, α, which does depend on risk preferences. Thus any set of risk preferences can be used when valuing C.

The Black and Scholes formula can be applied to options on stocks paying a continuous dividend yield at a constant rate δ[3]. This was derived by Merton (1973). The payment of a dividend causes a stock price to drop by an amount equal to the dividend. It follows that the payment of a continuous dividend yield rate δ causes the growth rate in the stock price to be less than it would otherwise be by an amount δ. Thus a European option on a stock with price S paying a continuous dividend yield of δ has the same value as the corresponding European option on a stock with price $Se^{-\delta(T-t)}$ that pays no dividend. To value the option it is simply necessary to reduce the current stock price from S to $Se^{-\delta(T-t)}$ and use the Black and Scholes formula. The expected proportional growth rate in the stock price is $r - \delta$.

It is this Black and Scholes formula that is used to value real options.

Merton (1973) considered the valuation of options when the interest rate is stochastic. Its model is the same as the Black and Scholes model with

- the instantaneous interest rate r replaced by the interest on a riskless bond maturing at the same time as the option;

[3]The dividend yield is the yearly dividends divided by the stock price

- the stock price volatility, σ, is replaced by $\hat{\sigma}$ which is a function of σ, of the bond volatility assumed to be a function of time and of the instantaneous correlation between the stock and bond prices. For most traded options it can be shown that $\hat{\sigma}$ is close to σ (Hull, 1993, p.436).

The binomial tree approach

The risk-neutral valuation principle states that any security dependent on a stock price can be valued on the assumption that the world is risk neutral. Thus to value an option we can assume that

- the expected return from all traded securities is the risk-free interest rate;
- future cash flows can be valued by discounting their expected values at the risk-free interest rate.

When using the binomial model the risk neutral valuation principle can be used and the expected value of the price at the end of a time interval Δt will be $S_t e^{r\Delta t}$. Thus (5.9) becomes[4]

$$p = \frac{e^{r\Delta t} - d}{u - d}.$$

This gives the risk- neutral probability of S. Now we use the binomial model to value a call with an expiration date one year away. Let C be the current call price, C_u be its value next period if the stock price goes to uS, and C_d be its value next period if the stock price goes to dS. If the price goes to uS, the value of the call will be $C_u = \max(uS - E, 0)$. Likewise, if the price goes to dS, the value will be $C_d = \max(dS - E, 0)$. The portfolio in (D.7) replicates exactly the option's value

$$C = SN - B \begin{cases} \xrightarrow{p} C_u = uSN - e^r B \\ \xrightarrow{1-p} C_d = dSN - e^r B \end{cases}$$

We have two equations with two unknowns, N and B. Solving for these unknowns we would define our duplicating portfolio. Having defined it, we could then determine its current value. As it has exactly the same value as our option, we would also have determined the option's value. The number of shares is found to be

$$N = \frac{C_u - C_d}{(u - d)\, S}$$

and the quantity that must be borrowed is

$$B = \frac{dC_u - uC_d}{u - d}\, e^{-r}.$$

The current value of the call is obtained substituting for N and B in $C = SN - B$. Alternatively, the risk neutral valuation can be used to value the call. The probability p can be determined from the information available about S. The value of the call is then simply given by

$$C = (pC_u + (1 - p)C_d)\, e^{-r}.$$

[4] If dividends are considered the expected value of the price at the end of a time interval Δt will be $S_t e^{(r-\delta)\Delta t}$

Options are evaluated by starting at the end of the tree (time T) and working backward. The value of the option is known at time T (for a call: $\max(S_T - E, 0)$, for a put: $\max(E - S_T, 0)$). Since a risk neutral world is assumed, the value at each node at time $T - \Delta t$ can be calculated as the expected value at time T discounted at rate r for a time period Δt. Similarly, the value at each node at time $T - 2\Delta t$ can be calculated as the expected value at time $T - \Delta t$ discounted at rate r for a time period Δt, and so on (Hull, 1993). If the option is American, it is necessary to check at each node to see whether early exercise is preferable to holding the option for a further time period Δt. Finally, by working back through all the nodes, the value of the option at time zero is obtained.

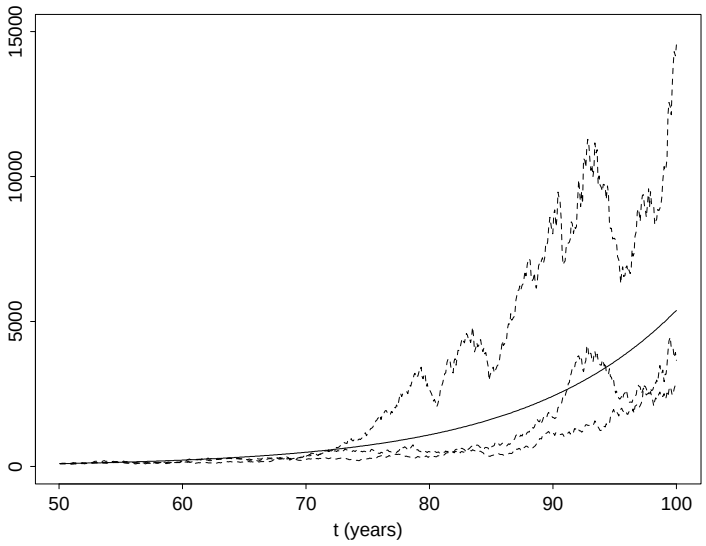


Figure D.1: *Sample paths of geometric Brownian motion with initial value* 100*, drift rate* 8%*, volatility* 18% *(*$T = 50$ *years,* $\Delta t = 1/12$ *year). The continuous line represents the trend. Because the spot price has a lognormal distribution, the upward swings are more marked than the downward ones*

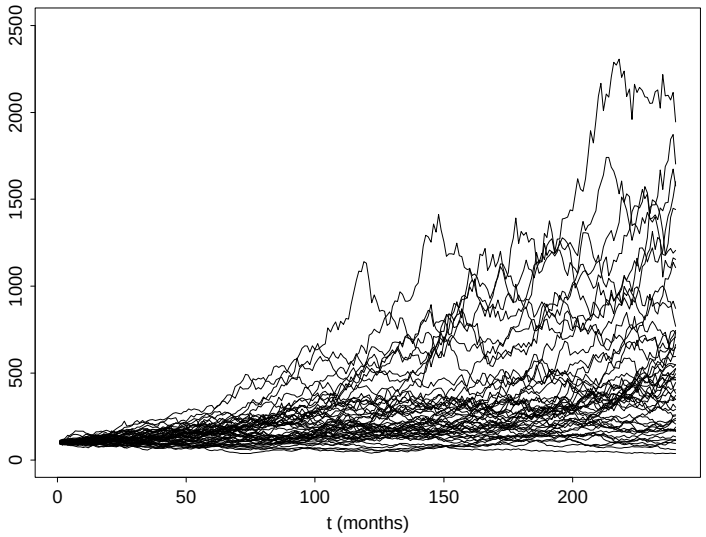


Figure D.2: *Fifty simulations of a geometric Brownian motion with initial value* 100*, drift rate* 8%*, volatility* 18% *(*$T = 20$ *years,* $\Delta t = 1/12$ *year)*

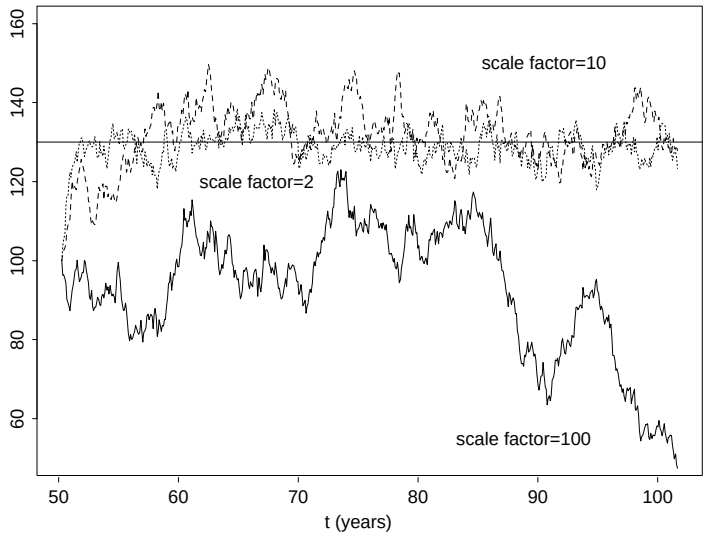


Figure D.3: *Sample paths of mean- reverting process. The process for η (1/scale factor) $\to 0$ tends to a simple brownian Motion without drift. It tends to wander far from its initial value of* 100. *While the larger $\eta(= 1/10, 1/2)$ the less it tends to drift away from $\overline{x}(= 130)$ ($T = 50$ years, $\Delta t = 1/12$ year)*

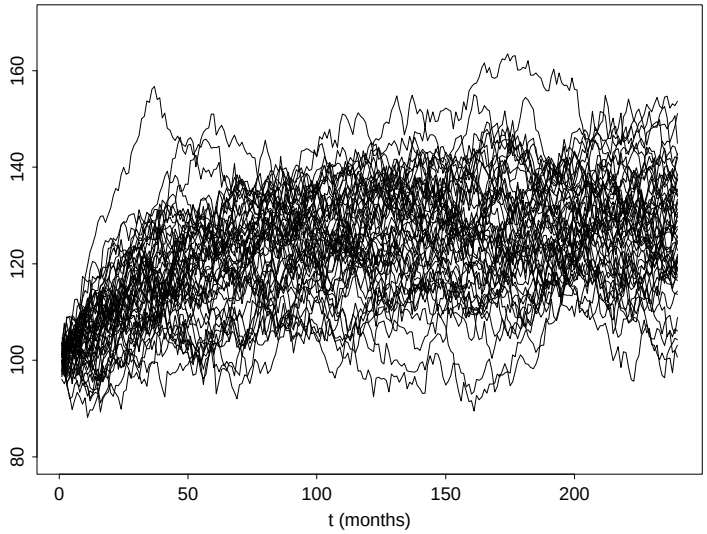


Figure D.4: *Fifty simulations of a mean- reverting process with initial value* 100, $\overline{x}$ 130, *volatility* 2, *scale factor* 5 *years ($T = 20$ years, $\Delta t = 1/12$ year)*

Appendix E

Dynamic programming

Stochastic dynamic programming is a solution technique with numerous applications within the field of operations research. Due to the recursive solution technique, SDP provides an efficient method to value flexibility.

E.1 Sequential decision process: finite horizon

At any stage a decision has to be made which will have an influence, an immediate return or cost, on the system's state at the subsequent stage. Given the initial state of the system the objective is to choose the sequence of decisions over time so as to maximise (or minimize) some given function of the immediate returns for all stages and states the system goes through. Information is not available all at once but at different moments. Dynamic programming is the set of algorithms and mathematical tools to study the sequential decision process and compute the eventual optimal strategies.

The following variables are naturally associated to the system's dynamic evolution

- the time variable, t, taking discrete values in the interval $[0, T]$;
- the state variable, x, with $x(t) \in \Re^n$, representing a point at t of the system's *trajectoire* (identifying the status of the project at each stage);
- the decision variable, a, with $a(t) \in \Re^m$, describing the operator's possible reactions to a changing environment.

It is assumed that the evolution of the system is described by the following state equation

$$x(t+1) = f(x(t), a(t), t).$$

Suppose the current date is t and the current state is x_t. Let $F_t(x_t)$ be the expected net present value when the firm makes all decisions optimally from this point onwards. Let $\pi_t(x_t, a_t)$ be the immediate profit flow when the firm chooses a_t. At the next period $(t+1)$, the state will be x_{t+1}. Optimal decisions will yield $F_{t+1}(x_{t+1})$. This is random from the perspective of period t, so we must take its expected value $E_t[F_{t+1}(x_{t+1})|x_t, a_t]$, that is called the continuation value. Discounting back to t, the sum of the immediate payoff and the continuation value is

$$\pi_t(x_t, a_t) + \frac{1}{1+\rho} E_t[F_{t+1}(x_{t+1})|x_t, a_t].$$

The firm will choose a_t to maximize this, and the result will be $F_t(x_t)$. Thus

$$F_t(x_t) = \max_{a_t} \left\{ \pi_t(x_t, a_t) + \frac{1}{1+\rho} E_t[F_{t+1}(x_{t+1})] \right\}. \tag{E.1}$$

This is formally stated in Bellman's principle.

The Bellman principle

The *optimality principle* was expressed by Bellman as follows: *An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.*

Equation E.1 is called the Bellman equation or the fundamental equation of optimality. It must be noted that it is not a linear equation. The optimal decision a_t depends on all the values F_t weighted by their appropriate probabilities. To solve this problem, as it is of finite horizon, we start at T and work backward. At T we have

$$F_T(x_T) = \max_{a_T} \{\pi_T(x_T, a_T)\}.$$

At $T-1$ we have

$$F_{T-1}(x_{T-1}) = \max_{a_{T-1}} \left\{ \pi_{T-1}(x_{T-1}, a_{T-1}) + \frac{1}{1+\rho} E_{T-1}[F_T(x_T)] \right\}.$$

Thus we know the value function at $T-1$, that in turn allows us to solve the maximization problem for a_{T-2} leading to $F_{T-2}(x_{T-2})$, and so on.

E.2 Sequential decision process: infinite horizon

In this setting, the problem one period hence looks exactly like the problem now, except for the new starting state. Therefore the value function is common to all periods, although it will be evaluated at different points x_t. The function $F(x_t)$ can then be written without any time label. Moreover since x_t and x_{t+1} could be any of the possible states, we get

$$F(x) = \max_{a} \left\{ \pi(x, a) + \frac{1}{1+\rho} E[F(x')|x, a] \right\}. \tag{E.2}$$

To find $F(x)$ an iterative procedure is needed. Start with any guess for the true value function, $F^1(x)$. Use it on the right-hand side of (E.2) and find the corresponding optimal choice rule a^1. Substituting it back, the right-hand side becomes a new function of x, $F^2(x)$, which is used as the next guess of the true value function, and repeat the procedure. Convergence is guaranteed no matter how bad the initial guess: the factor $1/(1+\rho)$ scales down any errors in the guess from one step to the next. For a comprehensive description of the procedure and other references see Dixit and Pindyck (1994).

References

References

Abramowitz M., Stegun I. A. (1965), *Handbook of Mathematical Functions*, Dover Publications, New York.

Armstrong M., Galli A. (1997), Option pricing: a new approach to valuing mining projects, *CIM Bulletin*, vol. 90, 37–44.

Azzalini A., Dalla Valle A. (1996), The multivariate skew normal distribution, *Biometrika*, **83**, 715–726.

Berger J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer- Verlag, New York.

Berger J. O., De Oliveira V., Sanso B. (2000), Objective Bayesian analysis of spatially correlated data, Technical Notes, Institute of Statistics and Decision Sciences, Duke University.

Besag J. (1989), A candidate's formula: a curious result in Bayesian prediction, *Biometrika*, **76**, 183.

Biver P., Mostad P. F., Guillou A. (1997), An overview of different techniques to quantify uncertainties on global statistics for geostatistical modelling, *E. Y. Baafi and N. A. Schofield (eds.), Geostatistics Wollongong '96*, Kluwer Academic Publishers, vol. 1, 573–584.

Black F., Scholes M. (1973), The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637–659.

Box G. E. P., Cox D. R. (1964), An analysis of transformations (with discussion), *Journal of the Royal Statistical Society*, B, **26**, 211–252.

Box G. E. P., Tiao G. C. (1973), *Bayesian Inference in Statistical Analysis*, Addison- Wesley Publishing Company.

Brealey R. A., Myers S. C. (1991), *Principles of Corporate Finance*, McGraw-Hill.

Brennan M. J., Schwartz E. S. (1982), Consistent regulatory policy under uncertainty, *Bell Journal of Economics*, **13**, 506–521.

Brennan M. J., Schwartz E. S. (1985), Evaluating natural resource investments, *Journal of Business*, **58** (2).

Carvalho R. M., Remacre A. Z., Suslick S. B. (2000), Geostatistical simulation and option pricing techniques: a methodology to integrate geological models in the mining evaluation projects, in *Kleingeld W. J. and Krige D. G. (eds.), Geostatistics 2000 Cape Town, Proceedings of the 6th International Geostatistics Congress held in Cape Town, South Africa*, vol. 2, 667–676.

Casassus J., Dufresne P. C. (2001), Maximal convenience yield model implied by commodity futures, Technical Notes.

Chilès J-P., Delfiner P. (1999), *Geostatistics, Modeling Spatial Uncertainty*, Wiley and Sons, New York.

Chorn L. G., Carr P. P. (1997), The value of purchasing information to reduce risk in capital investment projects, in *1997 SPE Hydrocarbon Economics and Evaluation Symposium held in Dallas, Texas, 16- 18 March*, SPE 37948.

Chorn L. G., Croft M. (1998), Resolving reservoir uncertainty to create value, in *1998 SPE Annual Technical Conference and Exhibition held in New Orleans, Louisiana, 27- 30 September*, SPE 49094.

Christensen O. F., Møller J., Waagepetersen R. (2000), Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo, Technical Report, Department of Mathematical Sciences, Aalborg University.

Christensen O. F., Diggle P. J, Ribeiro JR P. J. (2001), Analysing positive- valued spatial data: the transformed gaussian model, in *GeoENV III- Goestatistics for Environmental Applications*, Kluwer, Dordrecht, 287–298.

Connell S. (2002), Valuation of petroleum reserves with quantity and price uncertainties: the case of woodside energy, in *Real Options and Energy Management*, Ronn E. I. editor, Risk Books, chapter 17.

Coombes J., Thomas G., Glacken I., Snowden V. (2000), Conditional simulation- which method for mining, in *W. J. Kleingeld and D. G. Krige (eds.), Geostats 2000 Cape Town*.

Cortazar G., Casassus J. (1998), Optimal timing of a mine expansion: implementing a real options model, *The Quarterly Review of Economics and Finance, Special Issue*, **38**, 755–769.

Cortazar G., Schwartz E. S., Casassus J. (2001), Optimal exploration investments under price and geological- technical uncertainty: a real options model, *R & D Management Journal*, **31** (2), 181–189.

Cox J. C., Ingersoll J. E., Ross S. A. (1985), An intertemporal, general equilibrium model of asset prices, *Econometrica*, **53**, 363–384.

Cox J. C., Ross S. A. (1976), The valuation of options for alternative stochastic processes, *Journal of Financial Economics*, **3** (1 and 2), 145–166.

Cox J. C., Ross S. A., Rubinstein M. (1979), Option pricing: a simplified approach, *Journal of Financial Economics*, **7**, 229–263.

Cressie N. (1993), *Statistics for Spatial Data*, John Wiley & Sons, New York.

Cressie N. (1994), Comment: An approach to statistical spatial-temporal modeling of meteorological fields, *Journal of the American Statistical Association*, **89** (426), 379–382.

Davis G. A. (1996), Option premiums in mineral asset pricing: are they important? *Land Economics*, **72** (2), 167–186.

Davis G. A. (1998), Estimating volatility and dividend yield when valuing real options to invest or abandon, *Quarterly Review of Economics and Finance*, **38**, Special Issue, 725–754.

de Neufville R. (2001), Real options: dealing with uncertainty in systems planning and design, communication in *5th International Conference on Technology and Policy Innovation, Technical University of Delft, Netherlands*.

De Oliveira V., Kedem B., Short D. A. (1997), Bayesian prediction of transformed gaussian random fields, *Journal of the American Statistical Association*, **92** (440), 1422–1433.

Deraisme J. (1977), Utilisation de simulations de gisement pour la planification de l'exploitation, in *Symposium on Science and Technique, Mathematical Methods in Geology, Pribram, Tcheckoslovakia*, 203–216.

Dias M. A. G. (2002), Investment in information in petroleum, real options and revelation, communication in *6th Annual Conference on Real Options, Paphos, Cyprus*.

Dickens R. N., Lohrenz J. (1996), Evaluating oil and gas assets: option pricing methods prove no panacea, *Journal of Financial and Strategic Decisions*, **9** (2), 11–19.

Diggle P. J., Tawn J. A., Moyeed R. A. (1998), Model-based geostatistics (with discussion), *Applied Statistics*, **47** (3), 299–350.

Dixit A. K., Pindyck R. S. (1994), *Investment under Uncertainty*, Princeton University Press.

Dowd P. A. (1994), The use of conditional simulation in grade control, in *Proceedings of the Conference in Mining Geostatistics, Kruger National Park, South Africa*, Geostatistical Association of South Africa, 11–25.

Droesbeke J-J., Fine J., Saporta G. Editeurs (2002), *Méthodes Bayesiennes en Statistique*. Editions TECHNIP, Paris.

Ecker M. D., Gelfand A. E. (1999), Bayesian modeling and inference for geometrically anisotropic spatial data, *Mathematical Geology*, **31** (1), 67–83.

Galli A., Armstrong M. (1997), Option pricing: estimation versus simulation for Brennan & Schwartz natural resource model, in *Baafi E. Y. and Schofield N. A. (eds.), Geostatistics Wollongong '96*, Kluwer Academic Publishers, vol. 2, 719–730.

Galli A., Armstrong M., Jehl B. (1999), Comparison of three methods for evaluating oil projects, *Journal of Petroleum Technology*, 44–49.

Galli A., Jung T., Armstrong M., Lhote O. (2001), Real option evaluation of a satellite field in the North Sea, in *2001 SPE Annual Technical Conference and Exhibition held in New Orleans, Louisiana, 30 September- 3 October*, SPE 71410.

Geman S., Geman D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Geweke J. (1989), Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57** (6), 1317–1339.

Gibson R., Schwartz E. S. (1990), Stochastic convenience yield and the pricing of oil contingent claims, *Journal of Finance*, **45**, 959–976.

Gibson R., Schwartz E. S. (1993), The pricing of crude oil futures options contracts, *Advances in Futures and Options Research*, **6**, 291–311.

Gilks, Richardson, Spiegelhalter (1996), *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Grenadier S. R. (1999), Information revelation through option exercise, *Journal of Financial Studies*, **12** (1), 95–129.

Halvorsen K. B., Omre H., Berteig V. (1989), A Bayesian approach to kriging, in *M. Armstrong (ed.), Geostatistics*, Kluwer Academic Publishers, vol. 1, 109–126.

Handcock M. S., Stein M. L. (1993), A Bayesian analysis of kriging, *Technometrics*, **35** (4), 403–410.

Handcock M. S., Wallis J. R. (1994), An approach to statistical spatial-temporal modeling of meteorological fields, *Journal of the American Statistical Association*, **89** (426), 368–378.

Hinkley D. V., Runger G. (1984), The analysis of transformed data, *Journal of the American Statistical Association*, **79**, 386, 302–320.

Hoeting J. A., Ibrahim J. G. (1998), Bayesian predictive simultaneous variable and transformation selection in the linear model, *Journal of Computational Statistics and Data Analysis*, **28**, 87–103.

Huchzermeier A., Loch C. H. (2001), Project management under risk: using the real options approach to evaluate flexibility in R&D, *Management Science*, **47** (1), 85–101.

Hull J. C. (1993), *Options, Futures, and other Derivative Securities*, Prentice-Hall.

Journel A. G. (1977), *Géostatistique minière*, Thèse de Doctorat de l'Etat, Université de Nancy.

Kall P., Wallace S. W. (1994), *Stochastic Programming*, John Wiley & Sons.

Karma A. (1982), Issues in future markets: A survey, *Journal of Future Markets*, **2**, 261–294.

Kemna A. G. Z. (1993), Case studies in real options, *Financial Management*, **22**, 259–270.

Kitanidis P. K. (1986), Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resources Research*, **22**, 499–507.

Kleingeld W. J., Thurston M. L., Prins C. F., Lantuéjoul C. (1997), The conditional simulation of a Cox process with application to deposits of discrete particles, in *E. Y. Baafi and N. A. Schofield (eds.), Geostatistics Wollongong '96*, Kluwer Academic Publishers, vol. 2, 683–694.

Lane K. F. (1991), *The Economic Definition of Ore*, Mining Journal Books Limited, London.

Loch C. H., Bode-Greuel K. (2001), Evaluating growth options as sources of value for pharmaceutical research projects, *Management*, **31** (2), 231–248.

Lohrenz J., Dickens R. N. (1993), Option theory for evaluation of oil and gas assets: the upsides and downsides, in *1993 SPE Hydrocarbon Economics and Evaluation Symposium held in Dallas, Texas, 29- 30 March*, SPE 25837.

Lund M. W. (1999), Real options in offshore oil field development projects, communication in *3rd Annual Conference on Real Options: Theory Meets Practice*.

Majd S., Pindyck R. (1987), Time to build, option value, and investment decisions, *Journal of Financial Economics*, 7–27.

Mardia K. V., Kent J. T., Bibby J. M. (1992), *Multivariate Analysis*, Academic Press, London.

Maréchal A., Shrivastava P. (1977), Geostatistical study of a lower protozoic iron orebody in the Pilbara region of Western Australia, in *15th APCOM Symposium, Brisbane Australia, 4- 8 July, 1977*, 221–230.

Martzoukos S. H., Trigeorgis L. (2001), Resolving a real options paradox with incomplete information: after all, why learn?, communication in *5th Annual Conference on Real Options: Theory Meets Practice*.

Matheron G. (1973), The intrinsic random functions and their applications, *Advances in Applied Probability*, **5**, 439–468.

Matheron G. (1978), L'estimation globale des réserves récupérables, *Course notes C-75*, Centre de Géostatistique, ENSMP, Paris.

McCarthy J., Monkhouse P. (2003), To open or not to open- or what to do with a closed copper mine, *Journal of Applied Corporate Finance*, **15** (2), 56–66.

McDonald R., Siegel D. (1985), Investment and the valuation of firms when there is an option to shutdown, *International Economic Review*, **26**, 331–349.

McDonald R., Siegel D. (1986), The value of waiting to invest, *Quarterly Journal of Economics*, **101**, 707–727.

Merton R. C. (1973), The theory of rational option pricing, *Bell Journal of Economics and Management Science*, **4**, 141–183.

Moel A., Tufano P. (2000), Bidding for the Antamina mine: valuation and incentives in a real options context, in *Brennan M. J. and Trigeorgis L. (eds.), Project Flexibility, Agency, and Competition*, Oxford University Press, chapter 8, 128–150.

Moel A., Tufano P. (2002), When are real options exercised? An empirical study of mine closings, *Review of Financial studies*, **15**, 35–64.

Moyen N., Slade M. E., Uppal R. (1996), Valuing risk and flexibility: a comparison of methods, *Resources Policy*, **22**, 63–74.

Murtha J. A. (2001), Using pseudocases to interpret P10 for reserves, NPV and production forecasts, in *2001 SPE Hydrocarbon Economics and Evaluation Symposium held in Dallas, Texas, 2- 3 April*, SPE 71789.

Neuberger A. (2001), Gold derivatives: the market impact, Report prepared for the World Gold Council May 2001, London Business School.

Omre H. (1987), Bayesian kriging- merging observations and qualified guesses in kriging, *Mathematical Geology*, **19** (1), 25–39.

Omre H., Halvorsen K. B. (1989), The Bayesian bridge between simple and universal kriging, *Mathematical Geology*, **21** (7), 767–786.

Paddock J. L., Siegel D. R., Smith J. L. (1988), Option valuation of claims on real assets: the case of offshore petroleum leases, *Quarterly Journal of Economics*, **103** (3), 479–508.

Palm S. K., Pearson N. D., Read J. A. (1986), Option pricing: a new approach to mine valuation, *CIM Bulletin*, **79** (889), 61–66.

Papritz A., Moyeed R. A. (2001), Parameter uncertainty in spatial prediction: checking its importance by cross-validating the Wolfcamp and Rongelap data sets, in *GeoENV III- Goestatistics for Environmental Applications*, Kluwer, Dordrecht, 369–380.

Pérez J. M., Berger J. O. (2002), Expected- posterior prior distributions for model selection, *Biometrika*, **89** (3), 491–511.

Pericchi L. R. (1981), A Bayesian approach to transformations to normality, *Biometrika*, **68** (1), 35–43.

Pilz J., Spoeck G., Schimek M. G. (1997), Taking account of uncertainty in spatial covariance estimation, in *E. Y. Baafi and N. A. Schofield (eds.), Geostatistics Wollongong '96*, Kluwer Academic Publishers, vol. 1, 302–313.

Pindyck R. S. (1980), Uncertainty and exhaustible resource markets, *Journal of Political Economy*, **88** (6), 1203–1225.

Putter H., Young G. A. (2001), On the effect of covariance function estimation on the accuracy of kriging predictors, *Bernoulli*, **7** (3), 421–438.

Ravenscroft P. J. (1994), Conditional simulation for mining: practical implementation in an industrial environment, in *M. Armstrong and P. A. Dowd (eds.), Geostatistical Simulations*, Kluwer Academic Publishers, 79–87.

Ribeiro JR P. J., Diggle P. J. (1999), Bayesian inference in Gaussian model-based geostatistics, Technical report ST-99-08, Department of Mathematics and Statistics, Lancaster University.

Rivoirard J. (1994), *Introduction to Disjunctive Kriging and Non-Linear Geostatistics*, Oxford, Clarendon Press.

Robert C. (1992), *L'Analyse Statistique Bayesienne*, Economica, Paris.

Ross S. (1986), *Introduction to Stochastic Dynamic Programming*, Academic Press.

Sahin A., Fuseni A. (1998), Conditional simulation of grain-size distributions in a beach sand deposit, in *APCOM, London UK, Trans. Inst. Min. Met. 1998*, 215–224.

Sanguinetti H., Shrivastava P., Deraisme J., Guibal D., Humphreys M. (1997), Control of product variability in a large open pit iron mine, in *4th Large Open Pit Mining Conference, Perth Australia*, 133–138.

Schwartz E. S. (1997), The stochastic behavior of commodity prices: implications for valuation and hedging, *Journal of Finance*, **52** (3), 923–973.

Slade M. E. (2000), Valuing managerial flexibility: an application of real-option theory to mining investments, Technical Report, Department of Economics, The University of British Columbia.

Smith J. E., Nau R. F. (1995), Valuing risky projects: option pricing theory and decision analysis, *Management Science*, **41** (5), 795–816.

Smith J. E., McCardle K. F. (1998), Valuing oil properties: integrating option pricing and decision decision analysis approaches, *Operations Research*, **46** (2), 198–217.

Smith J. E., McCardle K. F. (1999), Options in the real world: lessons learned in evaluating oil and gas investments, *Operations Research*, **47** (1), 1–15.

Stein M. L. (1999), *Interpolation of Spatial Data, Some Theory for Kriging*, Springer-Verlag, New York.

Stensland G., Tjøstheim D. (1989), Optimal investments using empirical dynamic programming with application to natural resources, *Journal of Business*, **62** (1), 99–120.

Sweeting T. J. (1984), On the choice of prior distribution for the Box-Cox transformed linear model, *Biometrika*, **71** (1), 127–134.

Thwaites A. M. (1998), Assessment of geological uncertainty for a mining project, in *APCOM, London UK, Trans. Inst. Min. Met. 1998*, 391–404.

Trigeorgis L. (1990), A real- options application in natural resource investments, *Advances in Futures and Options Research*, **4**, 153–164.

Trigeorgis L. (1996), *Real Options: Managerial Flexibility and Strategy in Resource Allocation*, MIT Press, Cambridge, MA.

Zellner A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.