CALCULATION OF GEOLOGICAL UNCERTAINTIES ASSOCIATED WITH 3-D GEOLOGICAL MODELS

THÈSE N° 2476 (2001)

PRÉSENTÉE AU DÉPARTEMENT DE GÉNIE CIVIL

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Ian POMIAN-SRZEDNICKI

géologue diplômé de l'Université de Lausanne
de nationalité suisse et originaire de Genève (GE)

acceptée sur proposition du jury:

Prof. Dr A. Parriaux, directeur de thèse
Prof. Dr H. Einstein, rapporteur
A. Isler, rapporteur
P. Kohler, rapporteur
Dr V. Labiouse , rapporteur
Dr L. Tacher, rapporteur

Lausanne, EPFL
2002

Remerciements

En premier lieu, je tiens à remercier mon directeur de thèse, le professeur Aurèle Parriaux, directeur du Laboratoire de Géologie de l'Ingénieur et de l'Environnement (GEOLEP) de l'Ecole Polytechnique Fédérale de Lausanne. Je lui suis très reconnaissant de m'avoir reçu dans son laboratoire et d'avoir mis à ma disposition tous les moyens nécessaires pour mener à bien ce travail de recherche.

Je remercie le Dr Laurent Tacher, qui m'a guidé dans la réalisation de ce travail. Je remercie aussi tous les membres du jury de cette thèse, qui ont révisé mon manuscrit et y ont apporté des critiques constructives. Ces remerciements s'adressent particulièrement au professeur Herbert Einstein du Massachusetts Institute of Technology, pour l'intérêt qu'il a porté à ce travail, au Dr Alfred Isler du bureau Kellerhals+Haefeli AG, pour avoir mis à ma disposition les données nécessaires à la construction du modèle géologique 3D du portail sud du tunnel du Lötschberg, ainsi qu'au Dr Michel Mercier du bureau Bonnard&Gardel S.A., pour avoir spontanément accepté de faire partie du jury de cette thèse.

Je tiens à remercier toutes les personnes qui m'ont apporté leur aide en répondant à mes nombreuses questions: le professeur Michel Maignan, le professeur Jean-Luc Epard et le professeur Albrecht Steck, tous trois de l'Université le Lausanne, Reihardt Fuhrer et le Dr Sylvain Sardy du Département de Mathématiques de l'Ecole Polytechnique Fédérale de Lausanne, ainsi que le Dr Jiannong Fang du GEOLEP.

Je suis reconnaissant à Mme Ruth Borloz pour avoir réalisé l'important travail que représentait la relecture et la correction de mon manuscrit, ainsi qu'à Esteban Rosales qui a été l'artisan de la modélisation géologique 3D de cette thèse.

Je souhaite également remercier les collègues et amis pour les bons moments passés ensemble à l'Ecole Polytechnique Fédérale de Lausanne: le Dr Stéphane Dupasquier, le Dr Julien Vaudan, Nicolas Rist, Javier Cordoba et, bien sûr, tous les collaborateurs du GEOLEP que je n'ai pas cité ici.

Finalement, je ne saurais oublier mes parents, que je remercie pour le soutient qu'ils m'ont toujours apporté, ainsi que mon amie Esther A. Rueda Armada, pour sa patience et ses encouragements.

Ian Pomian-Srzednicki

Version abrégée

La construction des modèles informatiques 3D du sous-sol se fait à partir des observations récoltées sur le terrain par levé cartographique, par mesures géophysiques ou encore par forages. Cependant, ces données sont toujours en nombre limité de sorte qu'un modèle géologique est bâti sur une large part d'interprétation du géologue et d'interpolation par des méthodes mathématiques. Il s'ensuit que les modèles sont porteurs d'une incertitude, variable au sein du domaine d'étude, qui est rarement précisée dans le pronostic géologique. Ce travail a donc pour but de proposer une méthode originale pour l'évaluation des incertitudes sur les structures dans les modèles 3D du sous-sol et de tester les potentialités de son application.

La méthode proposée se base sur le modèle géologique 3D fourni par le géologue et le considère comme étant le plus probable parmi toutes les alternatives possibles (notion de «best guess»). Un terme aléatoire est ajouté à ce modèle afin d'apporter à cette description purement déterministe une part de fluctuations erratiques. Les différentes surfaces qui composent le modèle en question sont alors considérées de manière individuelle comme étant chacune un champ aléatoire gaussien. En chaque point de la surface, la fonction aléatoire $Z(\mathbf{u})$ décrivant la position de l'interface est composée de la somme d'un terme déterministe $m(\mathbf{u})$ représentant la position prédite par le modèle et d'un terme aléatoire $\sigma(\mathbf{u})\varepsilon(\mathbf{u})$ décrivant les fluctuations autour de la position la plus probable. Par la suite, un modèle de variabilité spatiale (un variogramme $\gamma(\mathbf{h})$) est proposé afin de conditionner le champ aléatoire aux observations disponibles. Des critères géologiques tels que la forme des plis et l'épaisseur des couches sont alors également pris en compte dans ce modèle. A ce stade, il est possible d'évaluer la variance locale sur la position des différentes interfaces par la technique du krigeage.

Finalement, la variabilité qui concerne les interfaces du modèle est convertie en information volumique par le calcul des probabilités d'occurrence des différents types de roche présents dans le domaine étudié. Ces probabilités sont calculées en accord avec les règles d'intersection qui régissent la séquence stratigraphique du modèle géologique. Elles permettent d'apporter un modèle probabiliste des structures du sous-sol qui peut être représenté sous la forme d'un champ de probabilités tridimensionnel.

Un programme de calcul qui applique la méthodologie proposée a également été développé au cours de cette recherche.

Abstract

3-D geological models are built with data collected in the field such as boreholes, geophysical measurements, pilot shafts or geological mapping. Unfortunately, these data are always limited in number. It implies that geological information is sparse and subsurface models are thus always built of both subjective interpretation and mathematical interpolation/extrapolation techniques. These models are therefore uncertain and this uncertainty is rarely pointed out in a geological prognosis. Our study proposes to bring a new methodology for the evaluation of geological uncertainties related to 3-D subsurface models and to test its potential use.

The methodology we propose is based on the 3-D subsurface model, which is here considered as the most probable prediction (notion of best guess). The various geological interfaces that compose the subsurface model are handled individually as Gaussian random fields. At each location of an interface, the random function $Z(\mathbf{u})$ describing the position of this interface is composed of a deterministic part $m(\mathbf{u})$ which represents the expected position, and a random part $\sigma(\mathbf{u})\varepsilon(\mathbf{u})$ which describes fluctuations around the predicted position. Then, a model of spatial variability (a variogram function $\gamma(\mathbf{h})$) is proposed in order to condition the random field according to available observations. Several structural constraints, such as the shape of folds and the thickness of layers can also be accounted for in this model. At this point, we are able to estimate the local variance all over the study area by the application of the kriging technique.

Finally, the variability is converted into three-dimensional information by calculating probabilities, this describes the occurrence of the various rock masses that are present in the study area. The probabilities are calculated according to intersection rules that govern the stratigraphic sequence of the subsurface model, and they allow us to build a probabilistic model of subsurface structures in the form of a three-dimensional probability field. All of this has been incorporated in a computer program.

Table of Contents

Chapter 1.
Introduction

1.1. Introduction

During these last decades, geosciences, as other sciences, have taken advantage of the rapid improvement of computer technologies. Computing devices are now widely used and have found many applications in the field of Earth sciences. Among these applications, three-dimensional geological modelling largely profited from these improvements, because handling of 3-D data needs fast computation capabilities and large storage capacities. Hence, since the beginning of the 90's, one can see a fast growing software development devoted to three-dimensional representations.

Three-dimensional geological modelling opens new perspectives and is now applied to many areas such as civil engineering, mining engineering or hydrogeology. Although data digitizing and file computing are always time consuming tasks, 3-D geological modelling still appears interesting for many reasons: it allows to better understand subsurface structures, to evaluate various alternatives, to test structural assumptions, to manage and store large quantities of information (Jones 1990). Finally, it also allows one to solve structural inconsistencies that are sometimes difficult to perceive in conventional two-dimensional representations such as cross sections or geological maps.

Geological models are built with data collected in the field such as boreholes, geophysical measurements, pilot shafts or geological mapping. Unfortunately, these data are always limited because their acquisition is expensive and time consuming. This implies that geological information is sparse and subsurface models are thus always built on large parts of both, subjective interpretation, and mathematical interpolation/extrapolation techniques. But bringing such subjective information (which has not been observed) into the model often leads to imperfect predictions. These models are therefore uncertain and this uncertainty is rarely pointed out in a geological prognosis. Reliable areas are usually not distinguished from those where the model should be read with caution. In general, there are even no references about uncertainties: when building a model, geologists are by experience aware that this model is to a great extent based on an uncertain personal interpretation, but they rarely mention these uncertain predictions. Generally, a geological model is accepted as it is, although it is known that it is a subjective representation that can be wrong in some areas.

For activities that are sensitive to geological risks, such as civil engineering, it is important to know how accurate the geological prognosis can be, in order to be able to manage costs and technical choices. When working on expensive projects, geological uncertainties can induce non-negligible financial risks or misleading strategies. It is thus necessary to locate and quantify these uncertainties as well as possible, so that risks can be correctly evaluated. A tool allowing the estimation of geological uncertainties would therefore be of invaluable

assistance to engineers and geologists trying to estimate advance rate and costs for tunneling, for example. Some decision-making systems already require, as an input, an estimation of geological uncertainties, but these uncertainties are still poorly evaluated. A detailed evaluation of uncertainties can also reveal areas that may require further investigations.

1.2. "State of the Art"

Obviously, it is not only important to build the best geological model possible, but it is also important to know how uncertain this model is. Several researches have been made in order to evaluate uncertainties inherent to geological models. These investigations are often only started if uncertainties can induce important financial losses on projects that are sensitive to geological risks. But they concern various domains of the Earth sciences such as:

- tunneling, where uncertainties about mechanical properties, fracturation, geological structures can affect the construction planning,
- petroleum exploration, where applications uncertainties can have severe consequences on reservoir management and recovery predictions when the estimation of volumes and hydraulic parameters in oil/gas reservoir models are unsure,
- geotechnical engineering, where uncertainties about mechanical properties and soils structures are often encountered,
- mining exploration,
- hydrogeology.

All these sphere of activity are affected by geological uncertainties, but they each have their own specific demand about the information contained in a geological model. So there are quite as many methodologies for the estimation of uncertainties as study cases. We can therefore find in literature a great variety of available approaches and methodologies dealing with geological uncertainties. Unfortunately, we cannot give here an exhaustive review of all existing researches.

However, we can notice that often, when dealing with geological uncertainties, available references in fact describe uncertainties about the estimation of some physical property such as porosity or permeability, but few references really deal with uncertainties about subsurface structures. So our subject is relatively badly documented at present, since references devoted to "structural uncertainties" are rather rare (Royer, Shtuka et al. 1997). In the following paragraphs, we will introduce some of the most relevant contributions to the estimation of these uncertainties, but we will focus our review on available methods applied to tunneling.

Construction performance in tunneling largely depends on the quality of many technical choices such as tunneling methods or equipment. But adverse geological conditions can also affect the performance and, as geological predictions often are uncertain, tunnel constructions are commonly subject to many geological uncertainties. When trying to predict the value under uncertainty, the estimation may differ from the reality, thus implying an estimation

error. To forecast tunnel construction cost and time, it is thus not sufficient to have a reliable estimation of the various parameters, but equally important is the assessment of uncertainties about the value of these parameters is equally important (Einstein, Halabe et al. 1996).

Often when assessing risks in tunneling, the statistical variability of the parameters is not taken into consideration. Best guessed values are used to bring out the most plausible scenario and possible alternatives are then taken into consideration by using some parameters with extreme values (Loew 1997). Such a qualitative technique has no statistical background and does not allow a proper assessment of uncertainties, but it can nevertheless bring a set of plausible pessimistic and optimistic scenarios that can help engineers to understand how tunnel construction can be affected by varying parameters. The disadvantage of this approach is that there is no way to take into account the relative likelihood of occurrence of the various scenarios. Such an approach can be improved by computing *Monte-Carlo simulations* in order to obtain a statistical overview of the various scenarios (Dudt and Descoeudre 1999).

Uncertainties are in fact a measurement of possible inaccuracies in the prediction of an unknown quantity or characteristic. They can only be defined in a probabilistic framework. So one can say that all available methods for the assessment of uncertainties have in common a probabilistic framework and the main distinction we can make between these methods can be done by considering the type of variable under study.

Uncertainties appear when trying to predict the quantity or the characteristic of a variable when the underlying physical process that generates this quantity or characteristic is unknown. One of the most popular examples is the prediction of the number that will come out when throwing a die. If the study variable is a quantity or a characteristic that does not vary in space or time - like the number coming out of the die - it can be considered as *a simple random variable* and in this case, classical statistics are applied.

In a probabilistic framework, the simplest way to evaluate uncertainties related to subsurface models is to consider the study variable as a simple random variable that does not depend on the location and on the spatial distribution of information. As an example, risk analysis techniques for tunneling often take into consideration several geological parameters such as the degree of weathering, the water availability, etc... The value of these parameters is obtained by observations or by empirical methods and uncertainties about these values are often estimated with a subjective description of the possible variability. As the estimation of uncertainties is not spatially dependent, it must then be corrected either by the application of some arbitrary rule, or by the use of more sophisticated techniques such as the *Markov process* (Vick and Einstein 1974), in order to represent changes with depth, for example.

When several simple random variables should be taken into accout, *probability trees* are often used, because they allow to integrate various sources of information into the evaluation of uncertainties. Conditional or Bayesian probability rules are applied in probability trees in order to integrate the various dependent sources of geological uncertainties when events are related. As an example, Bayesian Networks were used by (Holden and Mostad 1997) to evaluate uncertainties in petroleum applications. Probability trees also were successfully applied to tunneling by (Vick and Einstein 1974).

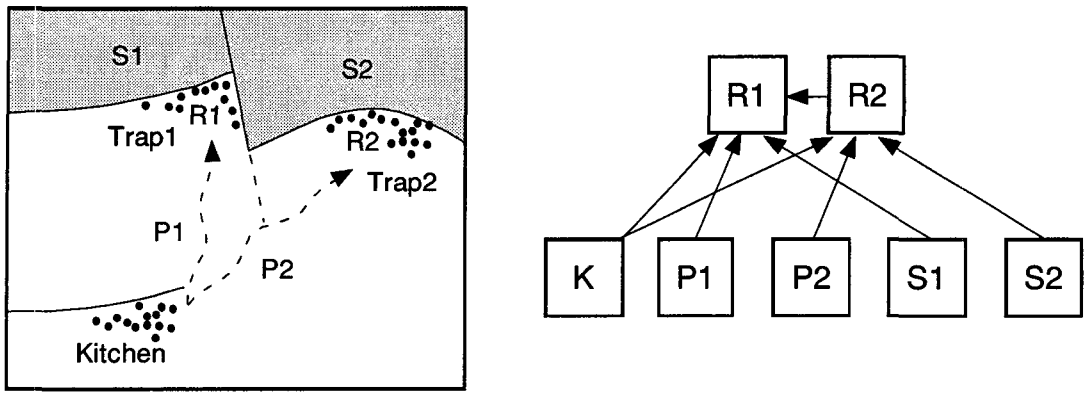


Figure 1-1 A very simple study case with its probability network for oil prospecting (Holden and Mostad 1997). The event K indicates the binary stochastic variable specifying whether the kitchen has an adequate amount of petroleum, P1 and P2 denote the variable for an adequate migration path, S1 and S2 denote whether the trap is adequately sealed and, finally, R1 and R2 are variables specifying whether there is petroleum in the trap.

Probabilities are calculated according to the relations given by the network. So the probability to find petroleum in Trap1 is $P(R1)=P(K)*P(P1)*P(S1)$ and the conditional probability to find petroleum in Trap1 - knowing that petroleum has been found in Trap2 - is given by $P(R1|P2)=P(K)*P(P2)*P(S2)$.

In Earth sciences variables are very often quantities that vary in space. This kind of variable is called a *regional variable*. The variable can be a *continuously varying quantity* in space (which is generally the case when modelling some geochemical or petrophysical parameters such as porosity, permeability, pollutant concentration, ore grade, etc.); or it can be, on the contrary, be a *discrete variable* when considering a set of various rock types or the distribution of fractures in a rock mass.

In a probabilistic framework, when the variable under study is an outcome of a regional phenomenon -as it is the case in our work- uncertainties are occurring when trying to predict the quantity or the characteristic of a variable at a location that has not been observed. Uncertainties in the prediction are thus mainly due to an unsure interpolation between known locations.

Many interpolation techniques are available, but as most of these techniques are purely deterministic, they are not able to access to the prediction errors. The most popular interpolation technique, which has a statistical background and allows the estimation of uncertainties, is the geostatistical *kriging* technique introduced by (Krige 1951) and formalized by (Matheron 1963).

An example of a direct application of the kriging technique for the estimation of geological uncertainties in tunneling is given by the study for the tunnel across the Channel (Blanchin, Margron et al. 1990). In this case, it was planned to bore the tunnel along the Chalk Marl layer, which is a particularly favourable rock for tunnel construction. The problem was that this favourable layer is only 30m thick and that, just below that layer, there is the very unfavourable Gault Clay layer. To prevent any risk of reaching the Gault Clay layer during the tunnel construction, a precise location of the top and bottom boundaries of the Chalk Marl

layer, as well as an estimation of possible uncertainties about the predicted position, were necessary.

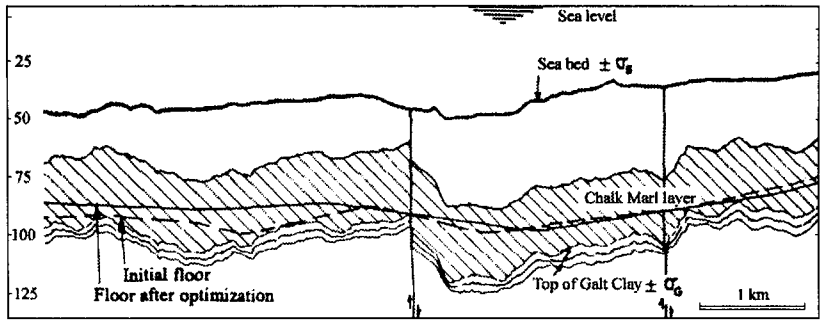


Figure 1-2 Cross section along the profile of the south tunnel (Blanchin and Chilès 1993) where the predicted position of the top boundary of the Gault Clay layer is surrounded by a confidence interval σ containing 84% of chances to encounter the boundary.

In this example, kriging errors could bring, according to some assumptions, to an estimation of uncertainties about the geological prognosis. It was thus not only possible to estimate the position of the various boundaries, but also to calculate a confidence interval around their expected position.

This example shows the application of a formal approach to quantify uncertainties about the position of an interface, but we must be aware that the classical kriging technique requires some strict conditions for a correct application. When using kriging, errors are estimated as the variance around an expected value and the expression of the statistical mean is therefore essential. When the mean value is not known, the simplest way to do is to assume that it is constant throughout the study area and the classical *Ordinary Kriging* technique can then be used. Such an assumption could be taken into consideration in this last example, because no obvious regional drift is noticed in subsurface structures. The case of the tunnel across the Channel was thus particularly favourable for the use of this technique, because geological structures were likely flat and horizontal throughout the Channel.

Unfortunately, a constant mean can rarely be reasonably assumed to describe geological structures and furthermore natural phenomena are often not homogeneous in space. In such cases, some more sophisticated kriging techniques as the *Universal Kriging* are available. Indeed universal kriging allows considering a functional representation of the mean, which can then be deduced from the study variable. The variable is here considered as the sum of a smooth deterministic regional drift and some erratic residuals. A practical example of an application of Universal Kriging for the estimation of subsurface uncertainties is given in (Abrahamsen and Omre 1994).

However, it is often difficult to find a reliable deterministic representation of the regional drift and, when the underlying geological phenomenon is unknown, the choice of a functional representation will always be debatable. On the next figure (Figure 1-3), we notice that an interpolation between the few available samples computed without knowing the underlying phenomenon obviously leads to a wrong description.

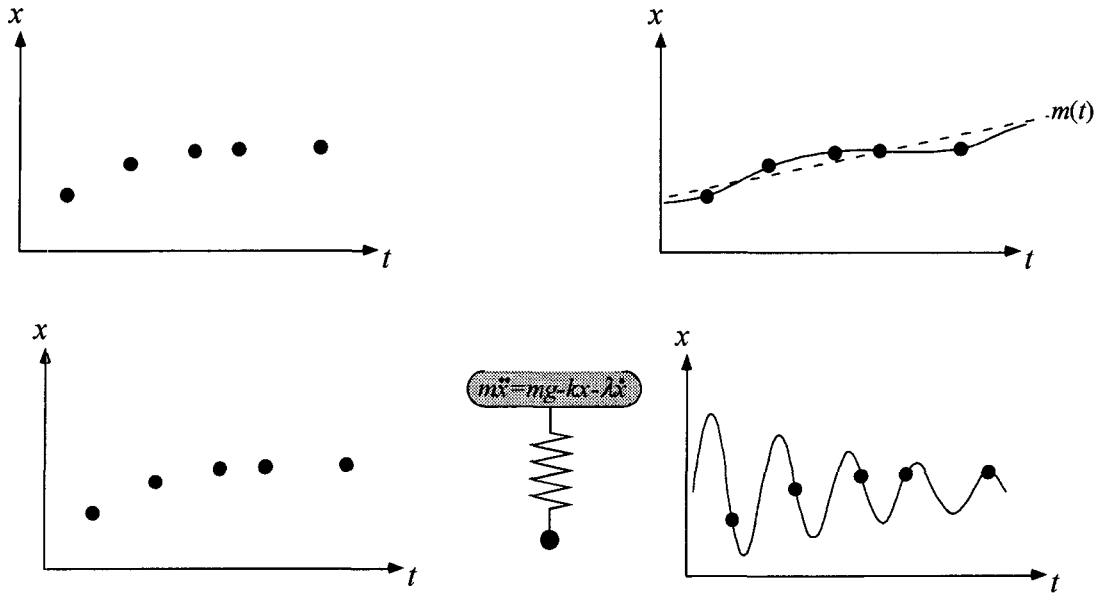


Figure 1-3 Example of an oscillating pendulum. Only few samples $x(t_i)$ of the position of the pendulum at various moments are available. On the upper figure, a linear trend model can be used to estimate the position of the pendulum at unsampled times with Universal Kriging, but this choice is debatable since the underlying phenomenon (bottom figure) is unknown.

Many other "detrending and residual modelling" techniques are available. Among these, the most exotic is the one using *Neural Networks* to find a regional structure model in a sample set. This technique is advantageously applied when the underlying structure is strongly non-linear or out of reach of any description (chaotic structures, for example) or when data are fuzzy or subject to large uncertainties. Generally, neural networks are helping to find a regional trend which is then substracted from the samples so that ordinary kriging technique can be applied on a "stationarized" phenomenon. The application of this technique in environmental sciences is described in (Kanevski, Demyanov et al. 1997). Direct probability mapping, where neural networks are used as an interpolation technique, can also be performed in order to describe uncertainties as illustrated by the geotechnical study case of (Najjar and Basheer 1996).

Since the number of data is often small, additional description of the phenomenon that is not contained in the available sample set will of course advantageously improve the estimation and thus reduce uncertainties. Several methods allow to incorporate any kind of external information as a-priori-knowledge in addition to hard data information. Among these, the Bayesian formulation is the most direct way to incorporate prior knowledge - or qualified guess – into the estimation; so *Bayesian Kriging* techniques can describe uncertainties according to prior knowledge. This technique has the advantage to also considering uncertainties about the qualified guess as being possible (Omre 1987). A practical application of Bayesian Kriging for the estimation of subsurface uncertainties is given in (Abrahamsen and Omre 1994).

The above mentioned Kriging techniques have the disadvantage that they can only provide one outcome for a given sample set. This outcome is the optimal solution for the kriging system, which minimizes the estimation variance. The consequence of this minimization is that estimations always are smoother than in reality. When wishing to keep some particular statistics or spatial pattern, the special technique of *random simulation* is more suitable. Random (or stochastic) simulation is the process which consist in building up alternative, equally probable models. With this technique, uncertainties can also be estimated; they can either be directly through the *error simulation* algorithm (Deutsch and Journel 1998), or indirectly obtained by a statistical description of several equally probable models coming out of jointly simulations. There are many simulation algorithms, but none of them is enough flexible to allow the reproduction of the wide variety of features and statistics that are encountered in nature; and often, hybrid approaches are proposed in order to simulate different types of features. Simulations were applied to estimate uncertainties in (Haldorsen and Damsleth 1990) for petroleum applications or by (Wingle 1997) for hydrogeology.

All these above-described techniques for the quantification of uncertainties show various ways of approaching the problem of geological uncertainties. Depending on what our need is, one technique will be preferred to the other, but each of these approaches has its limits and none of them allows a fully satisfying description of the features encountered in practice. They are either bring a too simplistic description or have to verify some constraining assumptions and cannot be correctly applied to any geological features. As a conclusion, we can say that the perfect technique does not exist. It is essential to first clearly specify what our demands on a geological model are and which technique will be best suited to satisfy these demands. The choice mainly depends on the quantity or the characteristic that is of interest for the study.

As said before, none of the available techniques is flexible enough to bring out a fully satisfying description of natural phenomena. So hybrid technique that can consider several features are certainly more suitable for complex features; but it remains a challenge for the future to include more precise deterministic descriptions of the phenomena in order to minimize the random parts of the model (Abrahamsen and Omre 1994).

1.3. Method Overview

The aim of this work is to set up a methodology allowing the estimation of uncertainties related subsurface structures in 3-D geological models and to test its potentialities on several cases. The basic idea of this methodology is to start from a best-guessed geological model. Then, according to the available observations and by making geological inferences, we build a model of variability, which describes how may differ the reality from the best guess. This leads to a statistical description of the phenomenon under study. Finally, a probabilistic model of the subsurface can be deduced from these statistics. It describes everywhere in the study area the probability to encounter a given rock type.

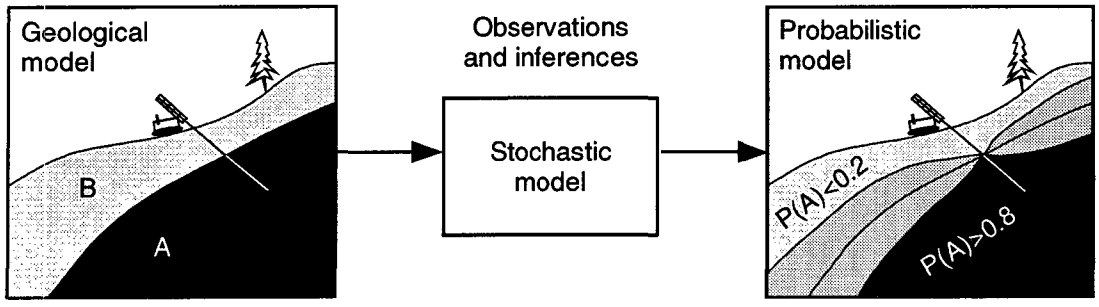


Figure 1-4 A basic scheme of the methodology. We start from the best guess (left). Studying available observations and doing geological inferences, we build a statistical description of the phenomenon under study. Finally, the probabilistic model is carried out, which will provide a geological prognosis in terms of probabilities.

The methodology can be summarized by the three following stages:

1° Our methodology needs a model as a starting point, which is here considered as the most probable realization (best guess). It is made up of both hard data and subjective geological interpretation. In the following example (Figure 1-5), the model contains only one geological interface, which is known through two observed points, and the boundary to a great extent part guessed.

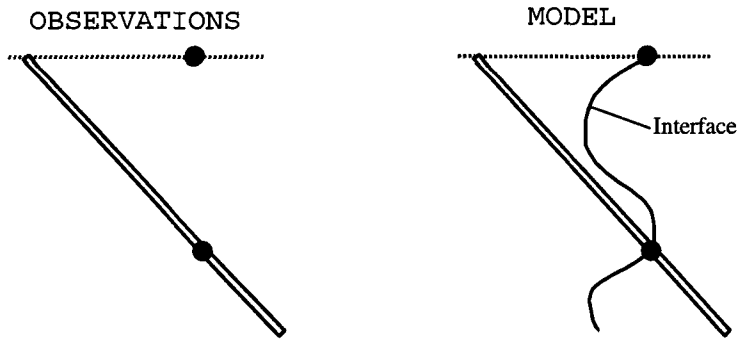


Figure 1-5 Simplified cross section showing two locations at which the interface is observed (left) and the model for the expected position of this interface (right).

2° The subsurface model is composed of several geological interfaces bounding the various rock masses. At this stage, each of these interfaces is handled as a random field. The random fluctuations are centred on the expected position of the interface and are described by a model of spatial variability (a variogram function $\gamma(\mathbf{h})$). This function expresses how increases the width of the uncertainty bounds when one moves away from the observed points (see Figure 1-6). The variogram may be different for each interface.

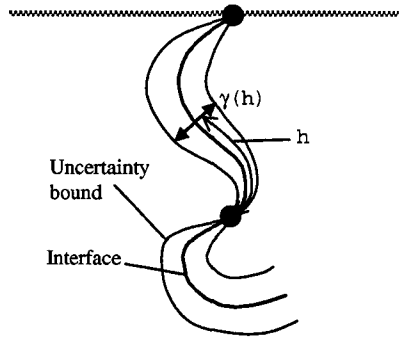


Figure 1-6 The same interface as in last figure (Figure 1-5) with its uncertainty bounds.

3° Finally, the variability is converted into a volumic information thanks to the calculation of probabilities describing the occurrence of the various rock masses that are present in the study area. The uncertainty about the positioning of the interfaces is converted into a 3-D probability field. Then, these probabilities are updated according to intersection rules that govern the stratigraphic sequence of the subsurface model. The process supplies a probability field for each rock mass to be encountered everywhere in the study area.

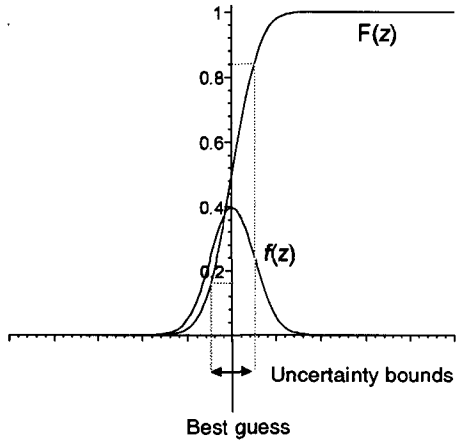


Figure 1-7 The probability of encountering a given rock type is obtained by integrating the probability density function $f(z)$. Uncertainty bounds here represent a confidence interval of $\pm\sigma$ which likely contains the real position of the interface

As the uncertainties are centred around the best guess (see Figure 1-7), the meaning of the results depends on its quality. Thus, the accuracy of the initial geological study is essential. When subsurface structures are expected to be very complex, one should be aware that the best guessed model can be very far from the reality. At the extreme, the best guess could be completely wrong. This case is not addressed in this work.

In the following sections, detailed explanations of these steps are first given in 1-D, then 2-D and 3-D. Although the principle remains unchanged whatever the dimension of the study area is, the following points must be considered:

1-D: The study area represents a stack of parallel layers. The interfaces are dots. There is no variogram, the uncertainty bounds are directly transferred.

2-D: The interfaces are line in a plane, geological bodies are areas. Since the distances are measured along the interfaces (Figure 1.3 and Chapter 3), the variogram is 1-D.

3-D: The interfaces are surfaces in the 3-D space, geological bodies are volumes. Since the distances are measured along the interfaces, the variogram is 2-D (Chapter 3).

In any case, the dimension of the probability fields is the one of the study area.

Chapter 2.
Preliminary

Before starting the description of the methodology, we will introduce in this chapter some important notions about how geological models are built and what the sources of uncertainties associated with these models are.

2.1. The Sources of Information

When building a subsurface model with limited information, geologists always take account of any additional information that can give a best knowledge of subsurface structures. As a consequence, the information used to build a geological model is often heterogeneous. Any useful information, whatever its natures, is taken into consideration in order to build the most accurate model possible. We will therefore describe in the following paragraphs the various sources of information available to build a subsurface model.

Lithological information is qualitative information about the type of rock present at a given location. This information is obtained either by *direct observations* of outcropping rocks, boreholes and pilot shafts or by indirect observations such as geophysical measurements. These can also indicate the types of rock present in the subsurface by measuring physical properties of subsurface rocks such as their electrical resistivity or their sound velocity, for example.

Direct observations of outcropping rocks are the most accurate information we can get about the geology (they are also referred to as "hard data"), but these observations are often too few and too scattered in practice, so that additional information is required. In this case, structural information such as *dip measurements* of rock layers is often accounted for, because it brings out invaluable indications about the geometry of structures in the study area. However, depending on the context, surface data cannot be considered as hard data. In layered deposits, *constant thickness* - if present - is also a very valuable information, because it gives a constraint and reduces the number of possible choices for the description of subsurface structures. In some folded terrains, one can make the hypothesis that folds are *cylindrical*; this is another precious information when building a subsurface model, because it indicates that structures are invariant along the direction of the fold axis, so that observations can be translated along this direction; even, data collected outside the study area can, by such a translation, be introduced into the model as additional information.

We can distinguish between data which is considered as sure (called "hard data") and inaccurate observations (called "soft data"). We usually consider that *hard data* are acquired when the information is directly accessible through observation. Such data are assumed to be

free of errors, but in reality, we often notice that they are not always accurate. Even direct outcrop observations can sometimes be subject to acquisition errors; for example when misidentifying the type of rock or when making a wrong localization of the observation in the field.

All these sources of information are often insufficient to achieve an accurate description of subsurface structures. Most of the time, there often still remains a certain number of arbitrary choices the geologists have to make. In their decision-making process, geologists refer to their experience and to their personal knowledge of geological phenomena. When supposing some similarities between previous experiences and the study case, they bring their own prediction into the model as a *qualified guess*. This is often case when drawing by hand a geological interface according to a given fold shape, for example.

2.2. The Sources of Uncertainty

Uncertainty in engineering geology can be grouped into three major sources: 1) measurement errors 2), innate spatial variability of geological formations, and 3) inaccuracies caused by modelling physical behaviour (Einstein and Baecher 1982). When modelling subsurface structures, the same sources of uncertainty can be considered. They are listed below:

1) Measurement Errors

Uncertainties due to measurement errors can be introduced by *sample disturbance, random procedural effects, bias, model inadequacy, or statistical fluctuations. These each have random and systematic components, and many are well studied. So-called random errors are presumed to have an average of zero (...). Bias errors introduce deviations which are systematically of the same trend. Model inadequacy is caused by the requirements of interpreting observed behaviour according to a physical model (...). Finally, statistical fluctuations are caused by finite sample size (...) and the variation of properties from one set of measurements to another* (Einstein and Baecher 1982).

Geophysical data are to a very large extent subject to measurement errors, often because of both random procedural effects and model inadequacy. However, direct observations, which are the most likely to be certain, are also inaccurate: Outcrop observations may be inaccurate because of a wrong positioning or a misidentification; observations made along boreholes can also be inaccurate, because the direction taken by the drilling head is not precisely known and information collected along the drilling track is thus difficult to locate with precision. The inaccuracy increases with the length of the borehole.

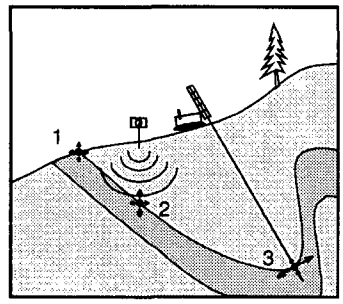


Figure 2-1 Possible acquisition errors in the field: 1) Outcrop location error, 2) geophysical measurement error and 3) borehole localization error.

2) Spatial Variability

Subsurface structures are spatially variable due to the fact that they are composed of different materials which are folded, stratified, truncated and, in other ways, separated into more or less discrete zones. When the number of available observations is sufficient, this variability can be precisely characterized. However, the number of observation is usually limited. Between available observations, subsurface structures are uncertain.

3) Model Uncertainty

Limited information on subsurface structures is a direct cause of the geological uncertainty (Alabert 1994), but wrong interpretation of the available information is another source of uncertainties. When building a subsurface model with limited data, geologists have to make some choices based on theory or empirical relations. To make up for poor information, they have to bring in their own interpretation and to build up their own subjective model. At this point, methodological misconceptions as well as inadequate choices and important factors left out of the interpretation model can be expected. Unfortunately, uncertainty resulting from human judgment are often difficult to detect and to quantify.

In addition to these three major sources of uncertainties, there is uncertainty due to omissions. The real world has properties and interrelationships that can never entirely be included in an analysis (Einstein and Baecher 1982). In our case, if subsurface structures such as faults or wedges are not observed, they will be left out of the geological model and the quality of the prognosis will thus seriously be affected. *Unless conditions are hypothesized, they cannot be included in the prediction. Many of the major failures of constructed facilities have been attributed to omissions* (Einstein and Baecher 1982).

2.3. 3-D Geological Modelling with EarthVision®

In practice, we will use data files created with the EarthVision® geological modelling software. It is thus important to describe some basic notions about the modelling techniques required by the software.

2.3.1. Input Data

Input data for the building of 3-D geological models with EarthVision® are simple scattered points coming from digitized cross sections, geological maps and boreholes. These points are then interpolated to complete the 3-D model (see Dynamic Graphics, 1997) for more details). These points are coming from direct observations as well as from interpretation.

2.3.2. Surfaces

Surfaces are the basic elements when building structures with EarthVision®. They represent the geological boundaries of the various rock masses in the model. These surfaces are obtained by a 3-D interpolation between available input data throughout the study area. There are two fundamentally different ways of building a 3-D geological model with EarthVision®, depending on the complexity of the structures we are aiming to model. Simple 2-D boundaries only require a direct interpolation, but complex structures may have boundaries that cannot be directly interpolated and that require a particular 3-D processing. More details on the building of geological structures with EarthVision® are given in (Mayoraz 1993).

2.3.3. The 3-D Model

When building a model of subsurface structures with 2-D data grids, only points belonging to the various geological interfaces are needed, because each rock mass is defined by its upper and its lower boundary. Each interface is contained in a single data file. The geological model is built by organizing and combining these data files all together into a 3-D model according to some stratigraphic relations. The EarthVision® modeller accounts for two main types of stratigraphic relation, which are:

- depositional surfaces and
- erosion surfaces.

By organizing and combining these two basic objects into a specific sequence, it is possible to build any kind of geological structure, from the simplest to the most complex one (Mayoraz 1993).

The various rocks are pilled one on top of the other starting from the bottom to the top of the stratigraphic sequence. This process of building a geological model is similar to the superposition law mentioned by Steno (1669)[1]

[1] The Danish physicist Nicholas Steno (1638-1686) gave two fundamental principles: the ***Law of Original Horizontality*** states that most sediments, when originally formed, were laid down horizontally. The ***Law of Superposition*** states that sedimentary rocks are formed bed by bed, and the layers are piled one on top of the other. Thus, in any sequence of layered rocks, a given bed must be older than any bed on top of it.

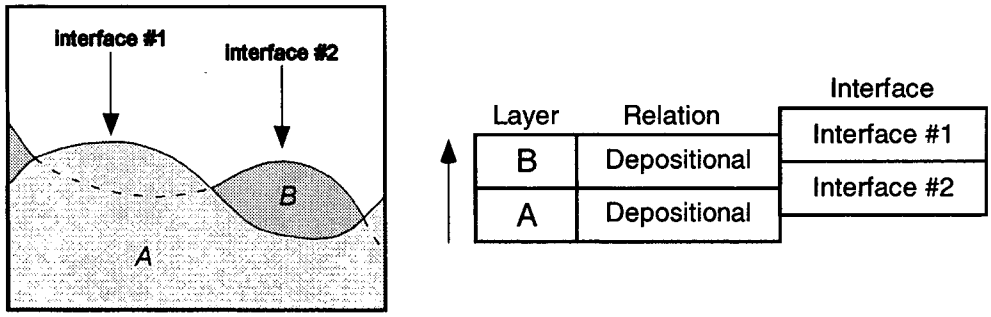


Figure 2-2 Illustration of the basic principle for the building of geological structures with EarthVision®. On the right, the stratigraphic sequence shows the chronology of the construction: First deposition of the body *A*. Then deposition of the body *B* over the body *A*. The Interface#1 remains unchanged, but Interface#2 disappears when it is located below the Interface#1.

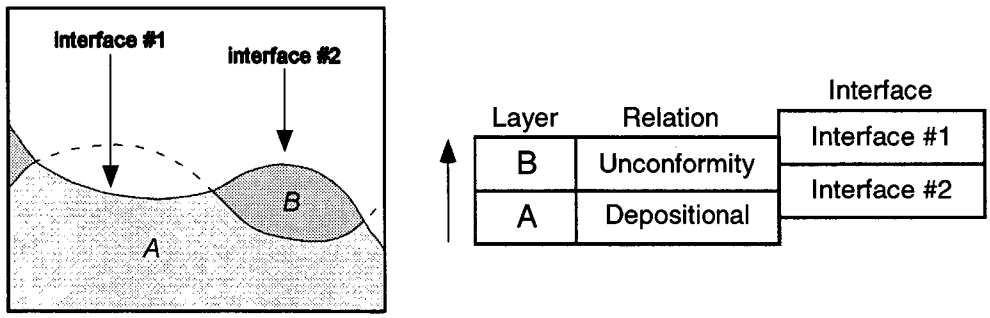


Figure 2-3 The same basic principle for the building of geological structures, but this time the Interface#2 is erosive. In this case, Interface#2 exists everywhere and the Interface#1 disappears when it is located above the Interface#2.

In a topological point of view, depositional surfaces and erosion surfaces are merely needed to build any kind of structure, but in reality, from a geological point of view, the classification should be more accurate. For example, fault surfaces are simply considered as erosion surfaces by the modeller.

2.3.4. Dealing with Overturned Folds

Since this methodology has to work in intensely deformed terrains, complex structures as those encountered in the Alps should be taken into account. Unfortunately, we will see that complex structures often induce geometrical problems which must be solved for a correct application of our methodology. In fact, our methodology can only work with simple 2-D surfaces and cannot account for 3-D data grids. This limit can however be solved if input files are correctly prepared and we will see that boundaries of complex geological structures can be represented with simple 2-D surfaces if they are correctly handled.

Interfaces are defined as surfaces in the 3-D space of our study area. Three coordinates are then needed to locate the points belonging to this surface. When using the Cartesian coordinate system, any point P_i is located by three coordinates x_i, y_i and z_i. With surfaces, the z coordinate can be expressed as a function of the two horizontal coordinates x and y, hence as $z(x,y)$, but in this case, when folds are for example overturned, it is possible to find multiple z coordinates at the same location (x,y) on the same interface.

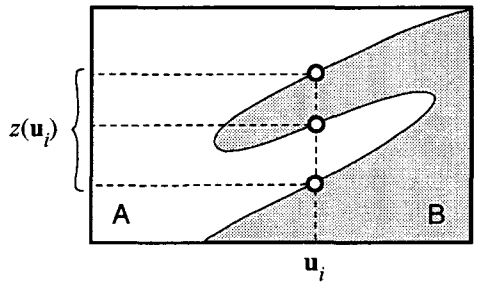


Figure 2-4 Illustration of overturned folds. The bounding surface between rock type A and B shows here multiple z at the horizontal location $\mathbf{u}_i$.

As our methodology can only consider simple 2-D surfaces, it requires that interfaces are free of multiple z. Unfortunately, complex 3-D structures such as overturned folds are often encountered in strongly deformed terrains. We could obviously not leave this problem apart without limiting our methodology to very simple and thus rare cases. At this point, the solution we suggest consists in the separation of the different limbs of an overturned fold into distinct interfaces.

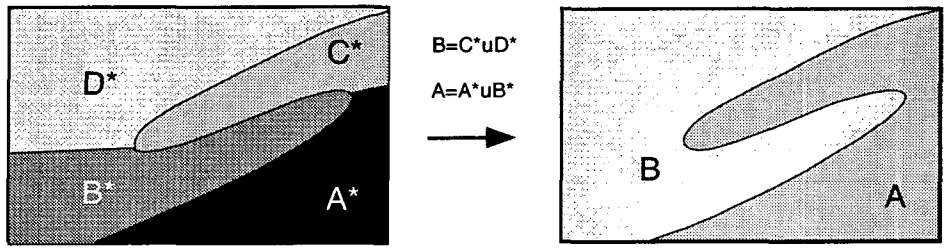


Figure 2-5 Illustration of how the problem of overturned folds can be solved. In this example, each limb of the fold is handled as a distinct interface. An intermediate model containing four bodies A^*, B^*, C^* and D^* is built. Superfluous rock masses are thus created, but the real structure can be obtained by considering that $A=A^* \cup B^*$ and $B=C^* \cup D^*$.

When building a geological model, we must therefore modify all surfaces that induce multiple z by finding a correct combination of additional simple surfaces - as in the example presented in Figure 2-5 - in order to avoid the presence of multiple z.

By doing this, each limb of a fold becomes an entity and will then be handled as an individual interface in the further calculation of the uncertainties.

2.3.5. Lateral Discontinuities

The same treatment should be applied when lenticular bodies occur.

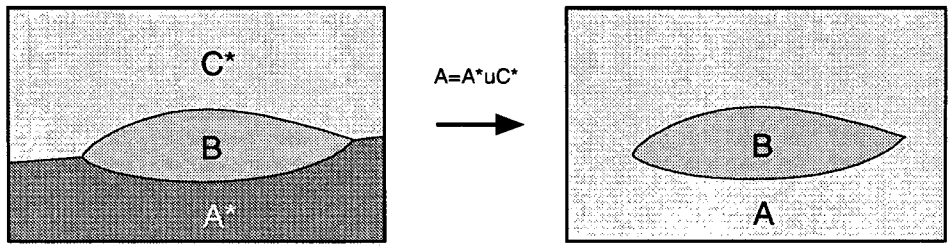


Figure 2-6 Illustration of how a lenticular body is processed. In this example, each side of the rock lens is handled as a distinct interface. An intermediate model containing three rock types A^*, B and C^* is therefore generated. As A^* and C^* are in reality the same rock mass, the original model can be obtained by considering $A=A^* \cup C^*$.

When modelling a faulted domain, multiple z do not occur yet the same treatment should be applied in order to ensure that rock masses are the same on each side of the fault.

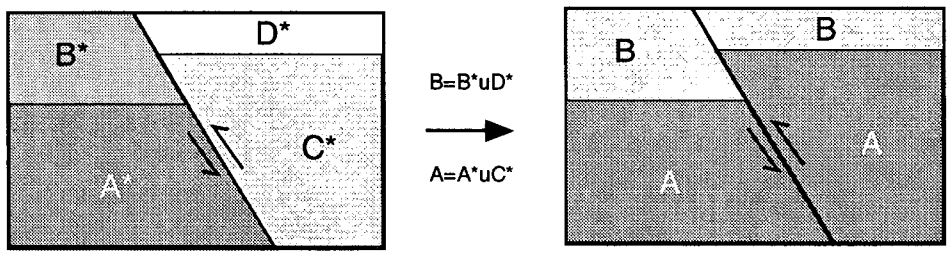


Figure 2-7 Illustration of how a faulted domain is processed. Interfaces on one side of the fault are not related to those on the other side. We must therefore say that A=A*∪ B* and B=B*∪ D*.

In the Alpine domain, like in all strongly deformed rocks areas, overturned folds and faults are often encountered. The geological model near the southern portal of the Lötschberg tunnel gives a good example, where this solution was applied in order to calculate the structural uncertainty (see Chapter 5).

Chapter 3.
Assessment of the Stochastic Model

Obviously, subsurface structures are intrinsically deterministic in the sense that they are not the result of a pure random process and that there is only a single description for them: the reality. But such a full deterministic description cannot be taken into consideration since, in most cases, we are not able to describe the phenomenon through a functional representation. An accurate description would only be possible if the underlying process that generated the phenomenon were accessible. We are thus just able to bring an uncertain prediction about subsurface structures.

In this chapter, we will introduce a way to describe the spatial variability of subsurface structures. We will describe the structures as being outcomes of a stochastic process and then build an appropriate stochastic model that describes the spatial variability. As the stochastic model is particularly important for a correct assessment of uncertainties, this chapter will thus constitute the most relevant part of our work.

3.1. Choice of the Study Variable

As we said in our introduction chapter, there are many ways of describing geological uncertainties and the choice of an appropriate technique mainly depends on what our demand about the geological model under study is. The expression of uncertainties will also depend on this demand. This is why, before describing the statistical model itself, we should pay a particular attention to the choice of an adequate study variable for the description of uncertainties about subsurface structures.

1° Properties data

Rock masses can be described as a set of points belonging to the study area $D \subset \mathbb{R}^3$ where each point has a property characterizing the type of rock that can be found at this location. We therefore have, at each location $\mathbf{x}_i = (x,y,z)$ of the study area a corresponding property p_i. For such a description, study variables take the following form:

$$p(x,y,z)$$

In a 3-D study area, three components are needed to locate the samples and another component, which here is the property p describing the type of rock, is required for the study variable. Such data are thus four-dimensional.

As rock masses are elements of a limited set of mutually exclusive categories, the property p

is not a continuously varying quantity. Such a variable is rather what we call a *categorical variable* (Deutsch and Journel 1998). Categorical variables are useful to describe geological features, such as terrains composed of a limited number of rock types, because rock masses can obviously not be regarded as a continuously varying quantity from a given rock type to another.

Each location of the study area is identified as belonging or not to one of the various rock types present in the study area. In this case, the most common categorical variable is the binary indicator $i(\mathbf{x})$ which is obtained by the application of a particular transformation to the available data.

Let us suppose, for example, that two different rock masses A and B are present in the study area, then the binary transformation can then be expressed as:

$$i(\mathbf{x}_i) = \begin{cases} 1, & \text{if location } \mathbf{x} \text{ belongs to the rock mass } A \\ 0, & \text{otherwise} \end{cases}$$

This kind of variable succeeds to describe uncertainties in subsurface models. An example of application of such transformation is given in (Kinnikutt and Einstein 1996).

Such an indicator variable is well suited for a discrete variable that describes a limited set of facies types. Other transformations may also allow multiple categories, but they are more difficult to handle in a statistical description and the physical meaning of multiple indicator thresholds are difficult to justify. The interest of a binary indicator variable is that it allows to directly provide a level of uncertainty about whether a random variable falls below or over a certain threshold value. The question is to know whether a transformed variable such as the binary indicator is still able to describe the phenomenon under study (Luo 1996).

2° Location data

Subsurface structures can also be described as an assemblage of bounding surfaces. In this case, the variable of interest is not the quantity or the characteristic that one can observe at a given location of the study area, but the location itself. One or more geographical coordinates are in this case considered as the random variables. A good illustration of an approach is also (Abrahamsen and Omre 1994).

Such a variable could be the z coordinate of points belonging to an interface. In a three-dimensional study area, this variable can be expressed as a function of two-dimensional coordinates vector $\mathbf{u} = (u,v)$, hence as $z(\mathbf{u})$.

Accounting for location instate of property as the study variable is more comfortable, since this approach has the advantage of requiring only the location of the various bounding surfaces and not the whole 3-D volume of the study area, hence less dimension. For this reason, we will prefer this variable as the study variable for our methodology.

As mentioned in the previous chapter, bounding surfaces are the basic elements of 3-D geological models built with the modelling software EarthVision®. At this point, the various interfaces should be extracted from the 3-D model and handled as individual objects. The first manipulation is thus the extraction of these boundaries.

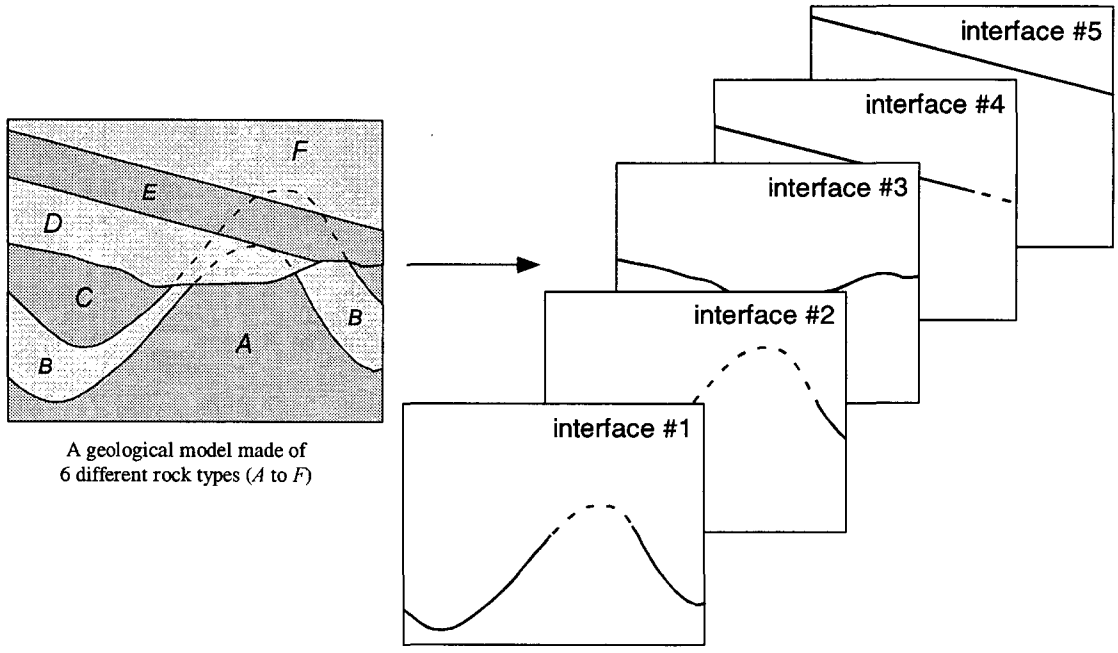


Figure 3-1 Extraction of the various interfaces. The geological model (on the left) is composed of 6 different rock types (domains *A* to *F*) and 5 bounding surfaces. These surfaces are extracted (on the right) and will then be handled as independent objects.

Using this approach, we account for information about the position of the interfaces and will thus only obtain uncertainties about an uncertain positioning of the various interfaces. This way we will not have a direct access to the uncertainty about the occurrence of the various rock types, as it is the case when accounting for property data, but we will see in the next chapter (Chapter 4) that such a description of subsurface structures can be regarded as equivalent to the previous one in the sense that the volumes containing the various rock types can now be obtained in this case by taking into account the domains limited by the various bounding surfaces.

3.2. The Random Function Concept

Now that we have chosen a suitable variable for the description of subsurface structures, we aim to describe the uncertainties, we should work in a probabilistic framework and consider the study variable as a random quantity. In a probabilistic framework, the process is stochastic and the available observations $z(\mathbf{u}_1)$, $z(\mathbf{u}_2)$,.., $z(\mathbf{u}_n)$ are outcomes of a random process, which has now to be defined.

We have chosen the coordinate z as being random and, in this case, the random process is defined in the two-dimensional study area $D \subset \mathbb{R}^2$ as the set of random variables defined in

the same probability field:

$$\{Z(\mathbf{u});\ \mathbf{u}\in D\}^2 \qquad (3\text{-}1)$$

The samples $z(\mathbf{u}_1)$, $z(\mathbf{u}_2)$,.., $z(\mathbf{u}_n)$ collected at n different locations $\mathbf{u}_1$, $\mathbf{u}_2$,.., $\mathbf{u}_n$ of the study area, are thus considered as outcomes of a *regionalized stochastic process* and are written with capital letters as $Z(\mathbf{u}_1)$, $Z(\mathbf{u}_2)$,.., $Z(\mathbf{u}_n)$.

As the set $\{Z(\mathbf{u}), \mathbf{u}\in \text{study area}\}$ is an outcome of a regionalized random process, the notion of random function comes by itself. By definition, $Z(\mathbf{u})$ is a *random function.*

Subsurface structures are not the result of a pure random process, because only few observations are available and geological models are built on both subjective knowledge and observations. Such models are thus mixed representations of precise information and uncertain predictions. Allowing for that, we chose to consider the stochastic model as also being a mixed representation of both deterministic and random information. In this case, the geological prognosis is the most probable description of subsurface structures and uncertainty is added to this deterministic description by introducing the notion of possible variability around the best-guessed prognosis. A similar approach can be found in Abrahamsen and Omre (1994) and Pilz, Spoeck et al. (1996).

The random function $Z(\mathbf{u})$ can be expressed as the sum of a local drift and of a local variability. The random function is then composed of two distinct parts: the first part $m(\mathbf{u})$ represents a deterministic component describing what we can consider as the most probable value for $Z(\mathbf{u})$ and the second part $\sigma(\mathbf{u})\varepsilon(\mathbf{u})$ represents a random term that expresses the fluctuating random process. This random function model enriches the deterministic description given by the geologist by introducing a random component that expresses the possible dispersion around this prediction.

Such an expression of the random field is given by Abrahamsen and Omre (1994):

$$Z(\mathbf{u}) = m(\mathbf{u}) + \sigma(\mathbf{u})\varepsilon(\mathbf{u})\ ; \qquad \mathbf{u}\in D\subset\mathbb{R}^2 \qquad (3\text{-}2)$$

This expression possesses two apparently contradictory characteristics. The first term represents a general, structured aspect that requires a certain functional representation and the second term expresses a local, random, erratic aspect (Journel and Huijbregts 1978).

- *The deterministic term $m(\mathbf{u})$*
  $m(\mathbf{u})$ is the expected value of $Z(\mathbf{u})$ at the location $\mathbf{u}$. It thus represents the most probable value. By using this expression, we suppose that the local mean $m(\mathbf{u})$ is known and that it describes the regional trend/drift of the random function. This functional description is what we call the "best guess" and it is known since we

[2] We will see in a later section that the usual x,y,z Cartesian coordinate system is not suitable to locate complex geological structures. In what follows, we will prefer using a general notation $\mathbf{u}=(u,v)$ for the position vector in order to avoid any confusion. The random function is thus written, without loss of generality, according to these particular coordinates, $Z(\mathbf{u})$; $\mathbf{u}\in D$.

can assume that the prognosis given by the geologist is be the best choice among all possible alternatives. The term $m(\mathbf{u})$ is thus given by the predicted position of an interface.

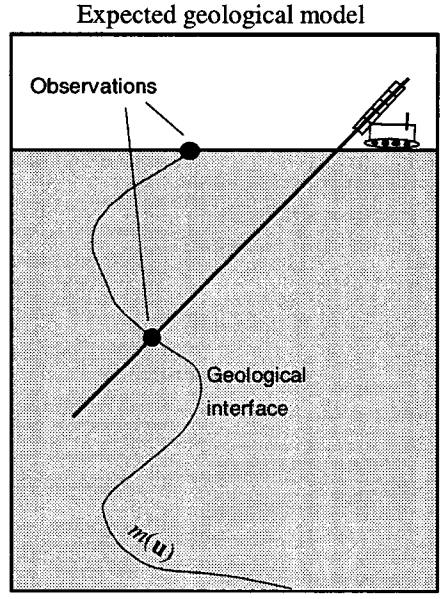


Figure 3-2 The predicted position of a geological interface. This prediction is regarded as the most probable one and thus represents $m(\mathbf{u})$.

- *The spatially dependent term* $\sigma(\mathbf{u})$
  The term $\sigma(\mathbf{u})$ is a standard deviation which expresses the variability at the location $\mathbf{u}$. This parameter is said to be spatially dependent, because it only depends on the distribution of available information. In a way, it can be regarded as a weighting parameter which provides the amplitude of possible fluctuations of $Z(\mathbf{u})$ at the location $\mathbf{u}$ according to available information.

  This parameter will be carefully studied, because it describes the variability of the phenomenon and thus provides us with a measurement of uncertainties at unsampled locations.

- *The random term* $\varepsilon(\mathbf{u})$
  The term $\varepsilon(\mathbf{u})$ is a random component expressing erratic fluctuations. The term $\varepsilon(\mathbf{u})$ is not spatially dependent, but it is a random function with $E\{\varepsilon(\mathbf{u})\}=0$ and $Var\{\varepsilon(\mathbf{u})\}=1$, which describes the way in which the random fluctuations are distributed around the mean position.

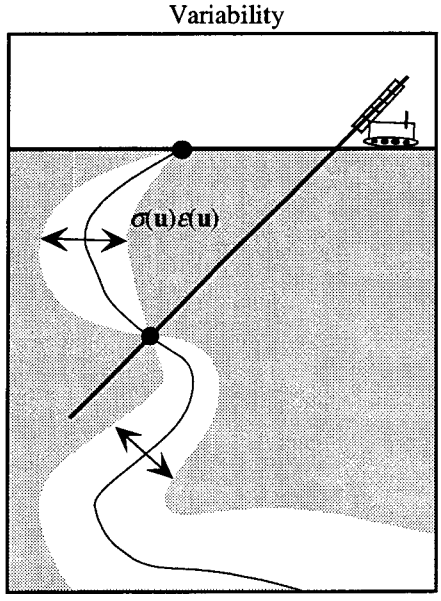


Figure 3-3 A model of variability $\sigma(\mathbf{u})\varepsilon(\mathbf{u})$ is added around the expected position of the geological interface.

The random function Z(**u**) (equation 3-2) is therefore defined such that $E\{Z(\mathbf{u})\} = m(\mathbf{u})$ and $Var\{Z(\mathbf{u})\} = \sigma^2(\mathbf{u})$.

A smooth regional trend model is often chosen as the mean position $m(\mathbf{u})$ of an interface and the observations $z_{obs}(\mathbf{u})$ are then compared to this model. This confrontation allows the measurement of the variability, since the residues $(z_{obs}(\mathbf{u})\text{-}m(\mathbf{u}))$ are regarded as random outcomes. But with this method, it is difficult to determine what part of the observed variability must be ascribed to the regional trend (often referred to large scale structures) and what part has to be imputed to local random fluctuations (small scale structures) (Pilz, Spoeck et al. 1996). Whatever this choice is, it always is debatable.

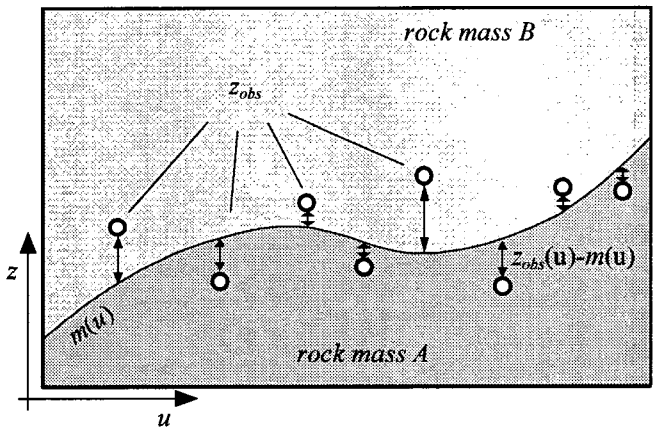


Figure 3-4 Profile of a geological interface given by a regional trend model. The measured distances between observed points z_{obs} and the trend model are considered as random fluctuations around a mean value.

The problem in correctly distinguishing between the contribution of the regional trend model and the contribution of local fluctuations in the random function often depends on the choice of the observation scale.

With regard to our concern, the subsurface model is considered as the most probable outcome of the random process. The deterministic component $m(\mathbf{u})$ is thus totally known, since it is given by the position of the various boundaries of the model. There is, in this case, no ambiguous choice of a general trend model.

In return, we cannot observe any fluctuation around this best guess, because we only have the deterministic description of $m(\mathbf{u})$ in the subsurface model. We thus cannot compare any alternative random outcomes to the prediction in order to measure a statistical variability.

As we aim to quantify errors related to a geological prognosis, particular attention should be paid to the evaluation of the random component expressing the possible dispersion around the best guessed model. In particular, the local standard deviation $\sigma(\mathbf{u})$ introduced in the equation 3-2 should be carefully assessed, because this parameter is the key to the estimation of uncertainties.

One way to get to a description of the variability would be the repeated measurements of discrepancies between prediction and reality. This would provide a direct observation of the variability around the mean position given by the geological prognosis. This approach is of course not really correct, because reality is unique and cannot be considered as an outcome of a random process. If we collect many successive measurements of the position of an interface at the same location, they will always be the same. As we said at the beginning of this chapter, subsurface structures are intrinsically deterministic. However, by making the assumption that the measured discrepancies at different locations are outcomes of the same random process, we can regard the measured discrepancies as a set of statistical values.

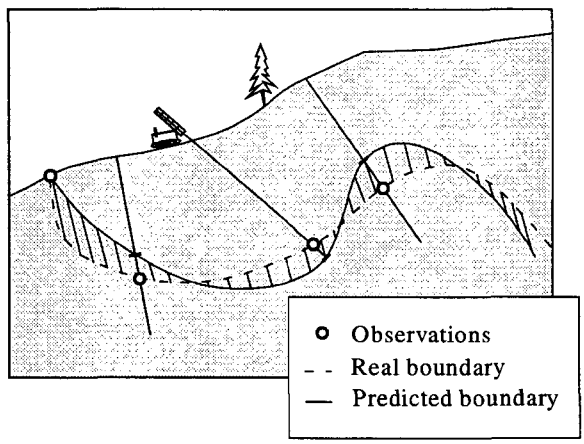


Figure 3-5 Direct subsurface observations, like boreholes, are the only way to directly compare the prediction to the reality. These observations can provide a measure of the discrepancies between the predicted position of an interface and its real position. They can thus provide a description of the variability around an average position.

Boring and observations in pilot shafts or in tunnel tracks allow a direct access to the measurement of the dispersion around the predicted position of an interface. Only a repeated

sampling of such data can provide a statistical measure of the discrepancies between prediction and reality (Figure 3-5), but unfortunately such costly information is rarely numerous enough to be representative. In some rare cases however, it is possible to have sufficient experimental data to allow an evaluation of the errors made in the prediction. The case of the Channel tunnel, which is presented below, is one of these few available study cases:

The Channel tunnel study case
During the first stage of a geological exploration for the excavation of the Channel tunnel, a first subsurface geological model based on a hundred of boreholes and 83 seismic profiles was created. This model could be considered as the best guess. In a second survey, as the project required a higher level of precision, some further seismic profiles were realized. This new information could be compared to the first prediction and hence gave measurements of the errors committed in the geological prognosis by comparing the observations obtained in the second survey to the prediction of the initial model (Blanchin and Chilès 1993 147).

These measurements provide a good illustration of the shape that can be expected for the distribution of errors. On the next figure, these measurements are represented as a histogram of the measured discrepancies between prediction and reality (Figure 3-6). When predictions is near to reality, discrepancies tend to zero. Most of the prediction are thus near the reality as the maximum frequency is located around zero. We can also notice that the distribution is quite symmetrical around zero.

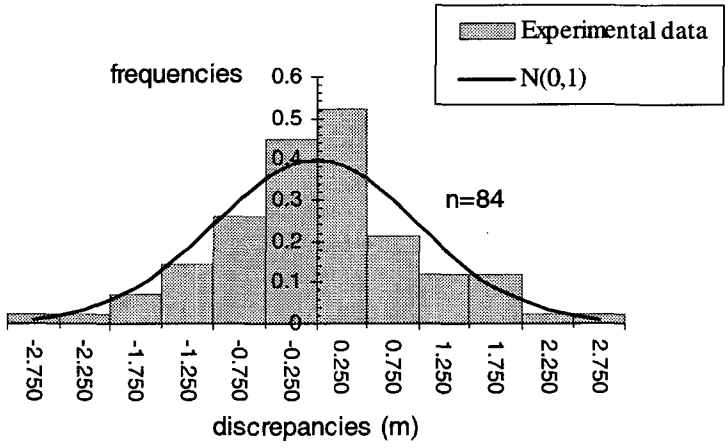


Figure 3-6 Histogram of the normalized discrepancies measured between prediction and reality according to (Blanchin and Chilès 1993 147).

Additional study can be done on these data. The χ^2 test gives an idea of how well this error distribution fits to a normal distribution. Comparing the normalized distribution of discrepancies to the distribution $N(1,0)$ and ranking the results in 12 probability classes, the χ^2 test gives a significance level greater than 10%, which can be considered as sufficient to assert that the distribution of errors is significantly normal.

Exhaustive data are usually not available and insufficient information often compels us to make arbitrary choices about the distribution function. In this case, we will assume that the distribution of $\varepsilon(\mathbf{u})$ is Gaussian and centred at zero (no systematic errors $E\{\varepsilon(\mathbf{u})\} = 0$).

This assumption is strengthened by the fact that, in practice, many random phenomena are following a normal distribution law (Journel and Huijbregts 1978), (Ross 1987). Another reason that leads us to assume a Gaussian distribution is the fact that such a distribution has handy properties. Indeed, when assuming that the distribution of $\varepsilon(\mathbf{u})$ is Gaussian also implies that the random function $Z(\mathbf{u})$ is Gaussian.

$Z(\mathbf{u}) = m(\mathbf{u})+\varepsilon(\mathbf{u})$, $\mathbf{u}\in D$ can therefore be considered as a Gaussian random field with a deterministic component $m(\mathbf{u})$ and a stationary zero mean random field $\sigma(\mathbf{u})\varepsilon(\mathbf{u})$.

$$\text{We can write } f(\varepsilon) = \frac{e^{\left(\frac{-\varepsilon^2}{2\sigma^2}\right)}}{\sigma\sqrt{2\pi}};\ -\infty < \varepsilon < \infty \qquad (3\text{-}3)$$

At this point, we are able to give an estimation of the uncertainty around a given position. We then have the most probable position, which is $m(\mathbf{u})$, and the dispersion around this average position given by $\varepsilon(\mathbf{u})$. However, this description of the variability so far only consists in a constant value which does not depend on the location: $Var\{\varepsilon(\mathbf{u})\} = \text{const}\ \forall \mathbf{u}\in D$.

3.3. Some Elements of Geostatistics

Geostatistics designate the statistical study of terrestrial phenomena. Initially, geostatistics were simply built to estimate ore grade for mining applications and the first geostatistical estimator was set up by (Krige 1951). But, as classical statistics fail to describe spatial phenomena, geostatistics were then further formalized and they opened the way to the theory of regionalized variables.

In this chapter we will introduce some elements of geostatistics that are necessary to elaborate a statistical model of uncertainties related to subsurface structures. However, the aim of this section is not to give a complete description of the geostatistical theory. Further literature can be found in (Journel and Huijbregts 1978) and (Isaaks and Srivastava 1989).

3.3.1. The Notion of Spatial Continuity and Spatial Variability

The description of the random process given in the preceding paragraph is so far incomplete as it does not accounting for the fact that the random process is spatial. We should now introduce the notion of spatial continuity in order to take into consideration the spatial pattern of the phenomenon.

A rock mass is physically continuous and elements near one another might thus be expected to have properties that are somewhat similar. On the average, the smaller the separation between

elements of a rock mass is, the more similar the properties are (Einstein and Baecher 1992). An estimation is obtained according to what has been observed at known locations and the quality of this estimation thus depends on the spatial distribution of the available information. The estimation mainly depends on the distance that separates the estimated point from a known location. We can thus expect that as the distance increases, the confidence on the geological prognosis will decrease.

In other words, we can expect that the local variance $\sigma^2(\mathbf{u})$ of the random function $Z(\mathbf{u})$ is small around known locations and that it increases when the distance from an observation increases. The notion of spatial continuity implements the previous expression of the random function by accounting for the distance that separates the estimation from the available observations.

The statistical inference requires a repeated sampling of the random variable. But, in our case, only one outcome - given by the geological model - is available at each location. So at the location $\mathbf{u}_i$, we only have one available outcome $Z(\mathbf{u}_i)$ of the random process and we thus do not have the required sampling in order to make an inference. At this point, the *stationarity assumption* is very often made in geostatistics. A random process is stationary, if samples collected at other locations $\mathbf{u} \neq \mathbf{u}_i$ can be assumed as being outcomes of the same random process. In this case, the random process is invariant by translation all over the sampling area.

$$F(\mathbf{u}_i, z_i) = F(\mathbf{u}_i + \mathbf{h}, z_i) \text{ for all } i = 1, .., n \text{ and all translation vector } \mathbf{h}. \qquad (3\text{-}4)$$

It implies that all statistical moments of the random process also are invariant by translation.

The spatial correlation

Covariance is a statistical moment of a bivariate distribution. If we have two independent random functions X and Y, the covariance is by definition:

$$Var\{X+Y\} = Var\{X\} + Var\{Y\} + 2Cov\{X,Y\}$$
$$\text{thus } Cov\{X,Y\} = E\{|X - E\{X\}| \cdot |Y - E\{Y\}|\} \qquad (3\text{-}5)$$

As far as we are concerned, the two variables are two measures of the same random function, but taken at two different locations $\mathbf{u}_i$ and $\mathbf{u}_j$. The covariance is then:

$$Cov\{Z(\mathbf{u}_i), Z(\mathbf{u}_j)\} = E\{Z(\mathbf{u}_i) \cdot Z(\mathbf{u}_j)\} - E\{Z(\mathbf{u}_i)\} \cdot E\{Z(\mathbf{u}_j)\}; \ \forall \mathbf{u}_i, \mathbf{u}_j \in D$$

Under the second order stationarity assumption, $E\{Z(\mathbf{u})\} = m = \text{const.}$, the covariance does not depend on the location $\mathbf{u}$, but depends only on the translation vector $\mathbf{h}$.

$$C(\mathbf{h}) = E\{Z(\mathbf{u}+\mathbf{h}) \cdot Z(\mathbf{u})\} - [E\{Z(\mathbf{u})\}]^2$$
$$\text{Therefore } C(\mathbf{h}) = E\{Z(\mathbf{u}) \cdot Z(\mathbf{u}+\mathbf{h})\} - m^2; \ \forall \mathbf{u}, \mathbf{u}+\mathbf{h} \in D \qquad (3\text{-}6)$$

The correlation function $C(\mathbf{h})$ is a measure of the spatial continuity and describes a two-points

connectivity pattern on all realizations of $Z(\mathbf{u})$ (Journel and Gomez-Hernandez 1993).

When the random field $Z(\mathbf{u})$ is Gaussian, it can be entirely defined by the spatial covariance function as $\varepsilon(\mathbf{u}) \propto N(0,C)$ (Deutsch and Journel 1998). So the expression of the random component $\varepsilon(\mathbf{u})$ of the random function can now be represented by $C(\mathbf{h})$.

Virtually all properties of a Gaussian random function are determined by the expectation, the variance and the correlation function (Abrahamsen and Omre 1994).

The variogram

An alternative to the covariance is the semi-variogram $\gamma(\mathbf{h})$. In a probabilistic framework and under the assumption of stationarity, the variogram[3] can be defined as the mean of a squared difference (Isaaks and Srivastava 1988)

$$2\gamma(\mathbf{h}) = E\left\{\left[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})\right]^2\right\} \qquad (3\text{-}7)$$

or as the variance of the increments

$$2\gamma(\mathbf{h}) = Var\left\{Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})\right\} \qquad (3\text{-}8)$$

The semi-variogram expresses the variability as a function of the distance between known locations. A remarkable property of the variogram is that it can be viewed as an estimation variance, the variance of error committed when the value at **u** is estimated by the value at the location **u+h** (Journel and Huijbregts 1978), which is particularly interesting in our case, as we aim to describe uncertainties.

If the assumption of stationarity is verified, the variogram is linearly related to the previous covariance as follows:

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \quad \forall\, \mathbf{u} \in D \qquad (3\text{-}9)$$

Development:

$$\begin{aligned}
&2\gamma(\mathbf{h}) = E\left\{\left[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})\right]^2\right\} = E\left\{\left[Z(\mathbf{u}) - m - Z(\mathbf{u}+\mathbf{h}) + m\right]^2\right\} \\
&= E\left\{\left[Z(\mathbf{u}) - m\right]^2\right\} - 2E\left\{\left[Z(\mathbf{u}) - m\right]\cdot\left[Z(\mathbf{u}+\mathbf{h}) - m\right]\right\} + E\left\{\left[Z(\mathbf{u}+\mathbf{h}) - m\right]^2\right\} \\
&= 2\sigma_Z^{\,2} - 2Cov\left\{Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})\right\} \\
&\Rightarrow \gamma(\mathbf{h}) = \sigma_Z^{\,2} - Cov\left\{Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})\right\}
\end{aligned}$$

The next figure (Figure 3-7) shows the relation existing between the variogram and the covariance function for a stationary random process.

[3] There often is a confusing language about the variogram. The function $\gamma(\mathbf{h})$ should be called "semi-variogram" as $2\gamma(\mathbf{h})$ is the function called variogram.

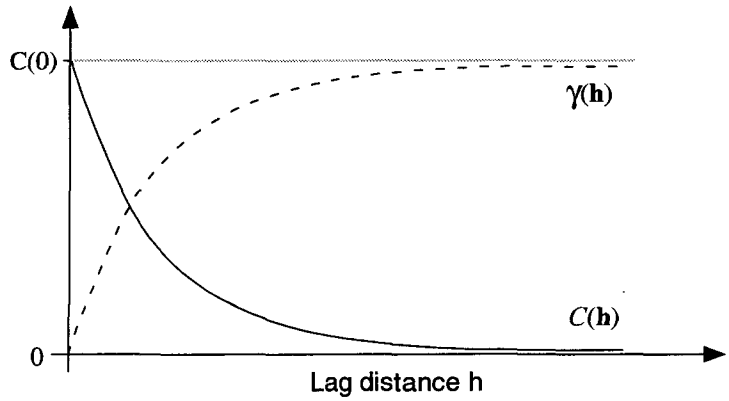


Figure 3-7 Relation between the covariance function (full) and the semi-variogram (dashed) for a stationary random process. The gray line represents the threshold value $\gamma(\infty)= C(0)$.

Very often, in practice, the semi-variogram stops increasing beyond a certain distance and becomes more or less stable around a limit value. Under the strong stationarity assumption this threshold value is simply the *a priori* variance of the random function (Journel and Huijbregts 1978). For large separation distances, the variogram equals $\gamma(\infty) = C(0)$.

In others words, the strong stationarity assumption means that all probability distribution functions of the random functions $Z(\mathbf{u}_i)$ are invariant by translation for a single sample $z(\mathbf{u}_i)$ as well as for the entire set of n samples $\{ z(\mathbf{u}_1), z(\mathbf{u}_2), .., z(\mathbf{u}_n) \}$.

When the correlation function exists, $C(0) = E\{(Z(\mathbf{u}) - m)^2\}$ is by definition the variance σ_Z^2 of the data set $\{Z_i\}$.

3.4. Choice of a Model of Variability

As was said previously, a remarkable property of the variogram is that it can be viewed as an estimation variance. This means that the variogram is able to bring out a model of variability for the studied phenomenon.

3.4.1. The Variogram Model

The experimental variogram

The estimation of the "true" variogram is usually made through an experimental variogram calculated by using the available sample set $\{ z_1(\mathbf{u}), z_2(\mathbf{u}), .., z(\mathbf{u}) \}$. The experimental variogram is obtained by calculating the mean of squared differences between all sample pairs $z(\mathbf{u}_i)$ and $z(\mathbf{u}_j)$ and it is a function of the separation distance $\mathbf{h}$ between the two sampled locations $\mathbf{u}_i$ and $\mathbf{u}_j$.

As the variogram is a measure of a statistical moment, each pair of samples $\{z(\mathbf{u}_i); z(\mathbf{u}_j)\}$ is ordered in classes of about the same distance lag and each distance class contains $n(\mathbf{h})$ pairs.

The experimental variogram is usually calculated with:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} \left[\left(z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h}) \right) \right]^2 \qquad (3\text{-}10)$$

Generally, the variogram model is obtained by adjusting a *function model* to the experimental variogram. Adequate parameters are needed to build a function that correctly models the variogram. In practice, these parameters are usually obtained by a visual or a least-square fitting of the function model to the experimental variogram.

This function is one of the few positive definite functions known (linear model, power model, spherical model, Gaussian model, etc...) and needs the three following parameters:

c = the sill value,
r = the range value,
s = the nugget,

where $a, c, s \geq 0$

The *sill* is defined as $c = \lim_{|\mathbf{h}| \to \infty} \gamma(\mathbf{h}) = C(0)$.

This sill is reached when the distance separating two points is greater than the distance of correlation. The variogram then reaches a maximal threshold value which is equivalent to the variance of the sampling set $Var\{Z(\mathbf{u})\}$ under the assumption of stationarity.

The *range* is the distance $r = h_0$ at which the model function reaches the sill. The range represents the limit distance of the spatial dependence and, beyond this limit, there is no more correlation.

If needed, a *nugget s* can also be added to the model. The nugget brings in an initial variability to the model. This can be due to sampling errors, for example.

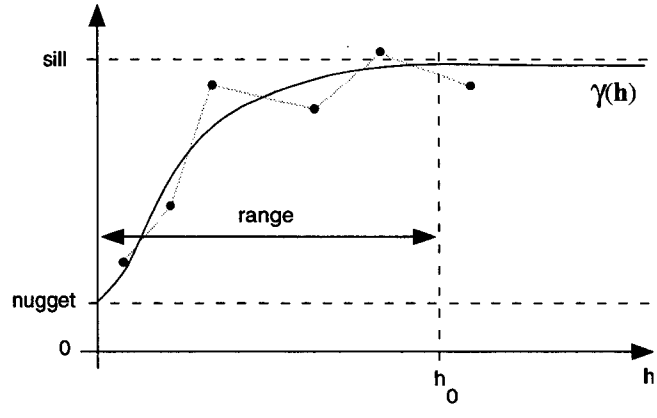


Figure 3-8 Example of a semi-variogram model fitting on experimental data (black dots). In this case, the Gaussian model $s + c \cdot \left(1 - e^{-\frac{|\mathbf{h}|^2}{r^2}} \right)$ is here used. The sill, the range and the nugget are also indicated on this figure.

Several issues arise when considering an experimental approach for the assessment of the variogram model:

- Common geostatistical methods usually assume that the variogram or the covariance function can be specified exactly by making the *ergodic assumption* that observations are coming out of a random process and that they constitute an exhaustive sampling set. The sample set $\{ z_1(\mathbf{u}), z_2(\mathbf{u}), .., z_n(\mathbf{u}) \}$ is supposed to be large enough to be representative of the random process. However, due to the fact that the sample size is often limited, the experimental variogram always is an estimation of the exhaustive variogram; it hence is only an uncertain model of the variogram. In order to preserve the statistical inference, it is necessary to compute the variogram with a sufficiently large number of data pairs. To consider the experimental variogram as statistically representative of the mean of the squared differences, (Journel and Huijbregts 1978) have suggested that the number of experimental data pairs should be higher than 30 for each distance lag. With n available samples, we can build $C_n^2 = n!/(n-2)!n!$ sample pairs, but they will not be equally distributed in each distance class. Generally the number of available pairs in a distance class decreases as the lag distance increases. The experimental variogram is therefore not significant for large lag distances and it is recommended to neglect the experimental variogram for distance lags that exceed than the half-size of the study area.

- Another problem is that many occurring spatial distributions show a marked tendency toward clustering (Davis 1973), so we must not only take into account the number of data, but also their spatial distribution. This is the case with for samples which are usually not uniformly distributed through the whole area, but rather clustered on the surface and along boreholes.

- Finally, the stationarity is often not verified. When examining the example of Figure 3-9, we notice that the variogram varies according to the extension of the study area. We thus are in presence of an obvious non-stationarity and, as a consequence, such an experimental variogram does not allow the access to the true variance.

The significance of an experimental variogram does not only depends on the quantity of available samples, but also on the statistical significance of the sampled area (Figure 3-9). It is therefore important to consider the extension of the sampling area in comparison with the dimension of the structures under study.

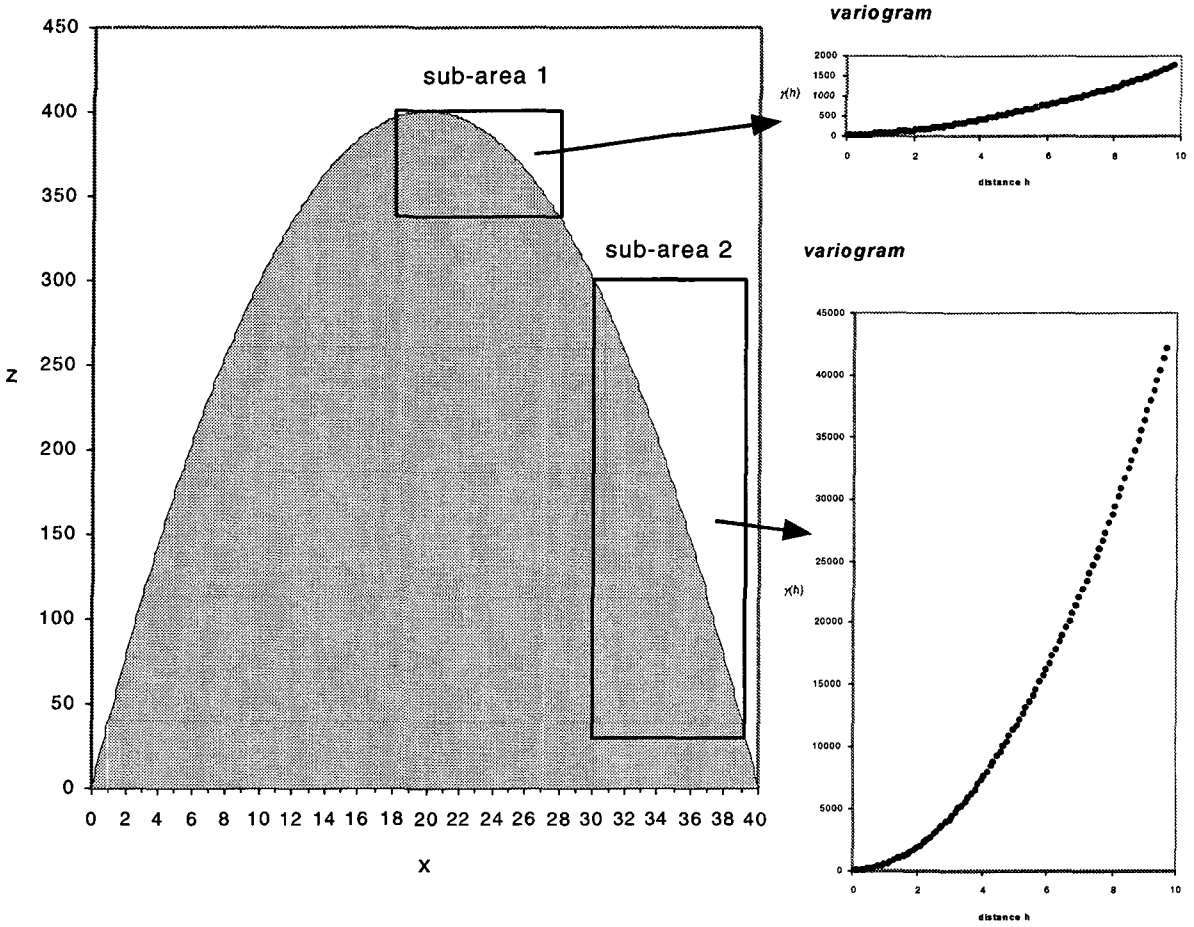


Figure 3-9 Example of experimental variograms calculated over two different sub-areas of a folded surface. These two variograms are very different and illustrate the fact that an experimental variogram is not representative of the phenomenon if the size of the sampled area is too small compared to the size of the studied phenomenon.

Some strict conditions are necessary to calculate a reliable experimental variogram but, as we pointed it out, these conditions are unfortunately hardly honoured in our case. This is mainly due to the fact that too few observations are available and that the study area is, in general, too small in comparison with the underlying phenomenon.

Attempting to calculate a reliable experimental variogram is not reasonable; we have therefore considered the possibility of finding an adequate variogram by other means than the usual experimentally calculated variogram. However, not taking account of the available samples (whose acquisition have required a lot of time and money) could be perceived as disrespectful towards the work that was necessary to collect these data. The usual procedure is to honour available samples by fitting a variogram model as best as possible to the experimental variogram.

Some people think that although it is occasionally possible to estimate the correlation

function/variogram, it must usually be on the basis of the knowledge which we have on the phenomena (Abrahamsen and Omre 1994). Knowing the problems we will encounter if we try to calculate an experimental variogram with the available set of data, we would rather suggest to choose an empirical variogram model. The variogram must have a physical meaning and the parameters must be related to the geology (Jaquet 1989). This empirical model should therefore be chosen according to what we know about the structures under study, rather than by taking into account available observations. As we already mentioned, the variogram gives some indications about the pattern of the studied phenomenon and can provide a description of the spatial structures present in the studied area.

Following these remarks concerning the problems that arise when using the experimental variogram, we will attempt to bring a satisfying description for an empirical variogram.

When studying subsurface structures in deformed rocks like those encountered in the Alps, folds are perhaps the most common tectonic structure (Ramsay 1987). So, complex structures such as folds should be correctly accounted for in our methodology, in order to bring a reliable estimation of uncertainties.

Folds in rocks can be found over a wide range of scales and shapes. Very small folds, such as crenulation cleavage, only measure some millimeters, but large folds, such as nappes, are structures that are often more than ten kilometers long (Figure 3-10).

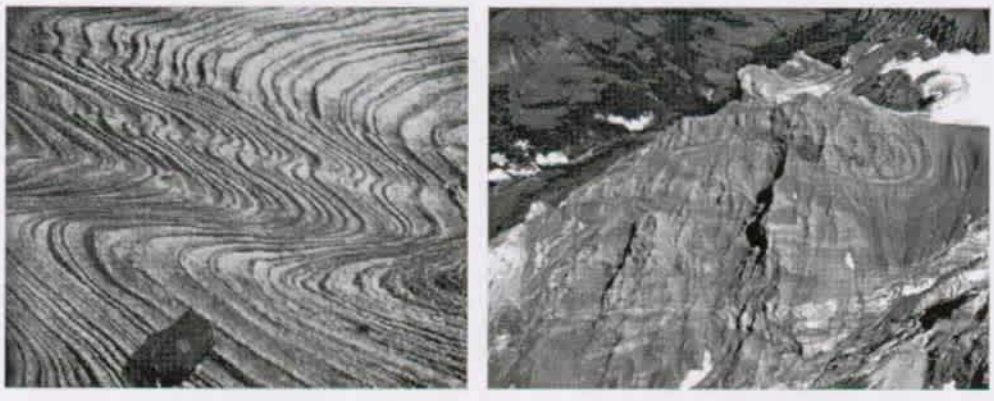


Figure 3-10 Two pictures of folded rocks showing the wide range of scales for such features. On the left, a folded sequence of calc-silicate and marble in (Ramsay 1987). On the right, a view of the Diablerets nappe (Swiss Alps).

3.4.2. The 1-D Model of Variability

If we suppose that the fold shape has a periodic regularity, it can be described as simple sinusoidal waves for example. This is shown by the periodical structure of following figure (Figure 3-11), where the boundary is represented by the function $z(x) = a \cdot \sin(\omega \cdot x)$.

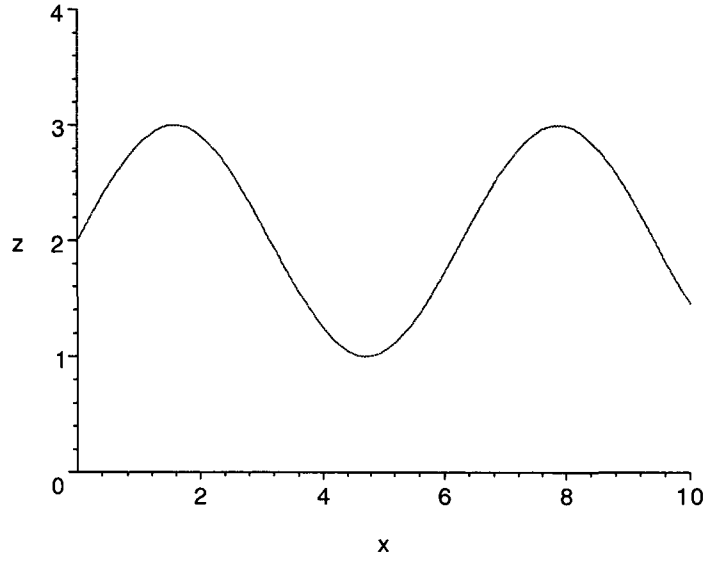


Figure 3-11 A folded interface is described by the function $z(x)=a\sin(\omega x)$ where $a=1$ and $\omega=1$.

This functional representation introduces the mathematical concepts of amplitude and wavelength, which are given by the parameters a and $1/\omega$ respectively. In this case folds can be described in terms of amplitude and wavelength.

The experimental variogram is calculated according to equation (3-10) and as $z(x)$ has a functional expression; the variogram can also be expressed by an integral. This variogram, which is referred to as the *regional variogram* by (Chilès 1999), can thus be calculated by an integral over the study area as follow:

$$2\gamma_D(h)=\frac{1}{(x_{max}-x_{min})}\int_0^{x_{max}-x_{min}}\left[a\sin(\omega x)-a\sin(\omega(x+h))\right]^2 dx \qquad (3\text{-}11)$$

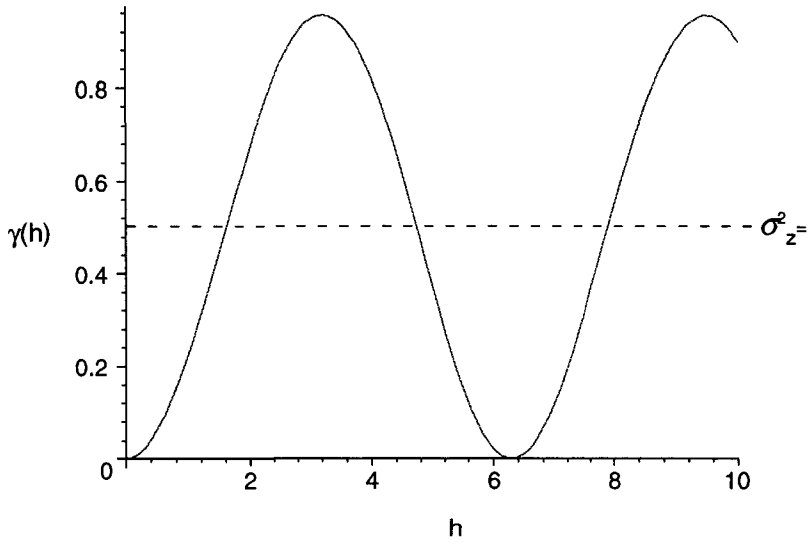


Figure 3-12 The semi-variogram obtained by integrating a sinusoidal signal. The semi-variogram is also periodic, with the same period as the signal. The maximum is reached when h equals a multiple of the half wavelength of the periodic function $z(x)=a\sin(\omega x)$.

This representation of the variogram is identical to the experimental variogram we would calculate with an exhaustive data set $\{z(x_i)\}$.

Such a description of fold shapes is of course too simplistic, but (Ramsay 1987) suggested that the wide variety of fold shapes could be represented by a series of trigonometric functions with following form:

$$f(x) = a_0 + a_1\cos(x) + a_2\cos(2x) + .. + a_n\cos(nx) + b_1\sin(x) + b_2\sin(2x) + .. + b_n\sin(nx) \quad (3\text{-}12)$$

With a combination of trigonometric functions it is possible to simulate every kind of fold shapes and to find their corresponding variograms. By comparing the variographic response of the various fold shapes, we notice that the shape does not have a great impact on its variogram (see Figure 3-13).

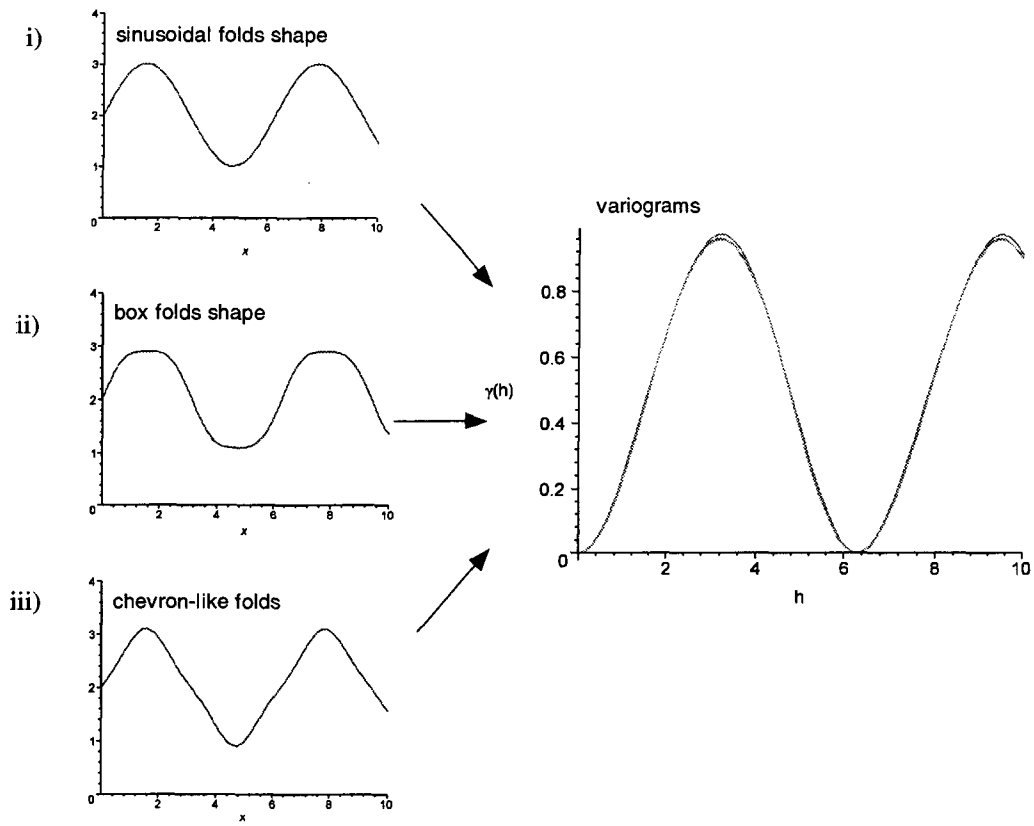


Figure 3-13 Calculated semi-variograms for various fold shapes that can be found in nature. i) Simple sinusoidal folds $z(x)$=sin(x), ii) box fold shapes $z(x)$=sin(x)+0.1sin(3x), and iii) chevron-like folds $z(x)$=sin(x)-0.1sin(3x). The resulting variograms are all grouped in the same plot (see figure on the right).

As we can see in the last figure, the various shapes of folds do not produce very different variograms. The most important parameters for the variogram are the amplitude and the wavelength of folds. In fact, the amplitude will determine the maximum value of the

variogram and the wavelength will determine the distance(s) at which this maximum is reached.

According to the principle of spatial continuity, two locations are better correlated when they are close one to the other. But the in above case, we are confronted with the problematic situation where, for some of the short distances, two points are less correlated than elsewhere two points separated by a larger distance (see Figure 3-12). In this case, the variogram is obviously not stationary and has no statistical meaning. However, such a perfect periodic regularity of fold shapes can of course not occur in reality. During the folding process, heterogeneities in mechanical properties of rocks may have induced perturbations in the supposed regularity. Therefore, we always observe fluctuating amplitudes and wavelengths from one fold to another in nature (see Figure 3-14). In opposition to the classical "dominant wavelength" description, some evidences of fractal self-similarity in fold shapes were described by (Budd and Peletier 2000).

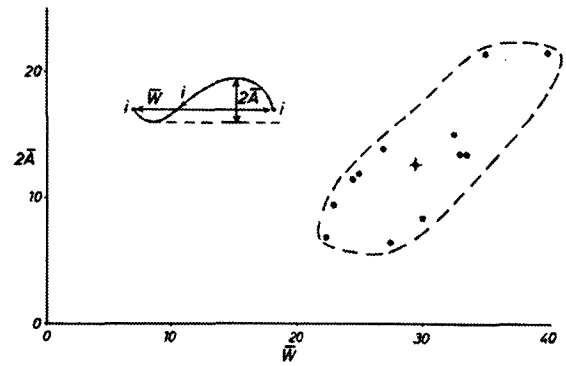


Figure 3-14 Wavelength-amplitude data from a sample of metasandstones and pelites from Mull, Scotland (from (Ramsay 1987)).

The full deterministic expression of the variogram given in equation 3-3 has no statistical usefulness, because it results from only one possible structure among all existing shapes. If we regard the amplitude and the wavelength as randomly fluctuating parameters (as shown by the data provided in the last figure, where varying amplitude and wavelength are measured on the same structure), the variogram can then be related to a random process. The previous sinusoidal function becomes $A\sin(\Omega x)$; where A and Ω are two independent random variables, the distribution of which is $f(A)$ and $f(\Omega)$ respectively.

The previous function is now a random function $Z_i(A_i, \Omega_i, x) = A_i\sin(\Omega_i x)$.

Its mean is: $\mu_{Z_i} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_i \sin(\Omega_i x) f(A_i) f(\Omega_i) dA_i d\Omega_i = 0$ (3-13)

and its variance is: $\sigma_{Z_i}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(A_i \sin(\Omega_i x)\right)^2 f(A_i) f(\Omega_i) dA_i d\Omega_i$ (3-14)

Unfortunately, mean and variance depend on the location x and the covariance not only depends on the separation distance h, but also on the location x. So such a representation may not be useful to define a correlation function or a variogram. However, when assuming that the same random process is invariant by translation along x, $Z(x)$ can be considered as a stationary random function and the mean variance along x can be regarded as the variance of this random process.

$$\overline{\sigma_{Z_i}}^2 = \int_0^{2\pi n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\left(A_i \sin(\Omega_i x)\right)^2 f(A_i) f(\Omega_i) dA_i d\Omega_i}{2\pi n} dx \qquad (3\text{-}15)$$

Let us suppose that the amplitude and the frequency of each mode are random variables that are uniformly distributed as $A \in [\, a_{\min}, a_{\max}]$ and $\Omega \in [\omega_{\min}, \omega_{\max}.]$. The mean variance can then be approximated by:

$$\overline{\sigma_{Z_i}}^2 \cong \frac{\overline{a}^2}{2} \qquad (3\text{-}16)$$

where $\overline{a}$ is the average amplitude of the signal.
Note that the influence of the varying frequency Ω can be neglected.

In contrast to the periodical signal described by a sine function, the noisy signal $Z(x)$ has no regularity. Each "wave" has a shape different from the others and there is no repetition of the same structure along x, so the perfect correlation from one wave to another is now lost. At long distances (longer than one wavelength) the variability reaches its maximal value and the variogram therefore reaches a sill value. As we assumed that the random process is stationary, this sill value is equivalent to the variance of the random process.

The parameters c and r of the model variogram can thus be approximated by:

$$c \cong \frac{\overline{a}^2}{2} \text{ and } r \cong \frac{1}{\omega_{\max}} \qquad (3\text{-}17)$$

Numerical example
As shown in Figure 3-15, a noisy signal can be produced by a sum of N "noisy" modes with random amplitude and wavelength (like a spectral representation).

$$Z(x) = \sum_{i=1}^{N} A_i \sin(\Omega_i x)$$

The mean and variance are respectively $\mu_Z = \frac{1}{N}\sum_{i=1}^{N}\mu_{Z_i}$ and $\sigma_Z{}^2 = \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{Z_i}{}^2$.

The resulting signal is shown in the next figure (Figure 3-15):

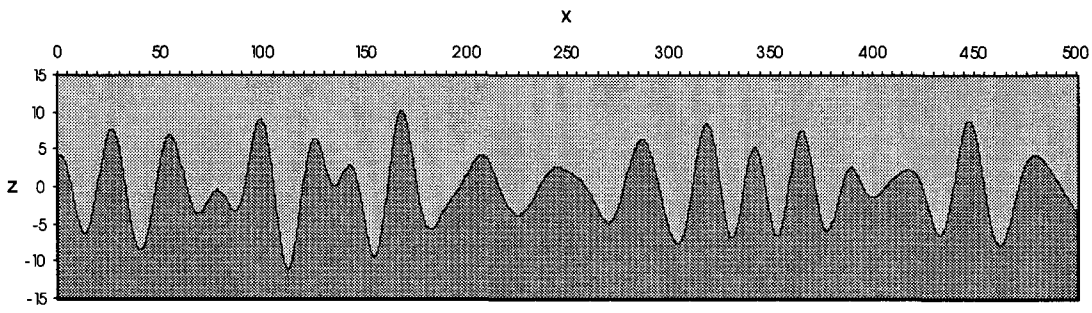


Figure 3-15 A simulated profile of a random signal resulting from the sum of hundred noisy modes $Z(x) = \sum_{i=1}^{100} A_i \sin(\Omega_i x)$, where amplitude and wavelength are random variables with uniform distributions such as $A \in [55, 65]$ and $W \in [20, 40]$.

In this case, the parameters used for this simulation can be used to approximate a semi-variogram model.

The sill can be approximated by $c = \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{Z_i}{}^2 \cong \frac{1}{100^2}\cdot 100 \cdot \frac{60^2}{2} = 18$,

the range can be approximated by the minimum half-wavelength $r = \frac{w_{\min}}{2} = 10$,

and the semi-variogram model we suggest is thus the following Gaussian model:

$$\gamma^*(h) = c\left(1 - e^{\left(\frac{h^2}{r^2}\right)}\right) = 18 \cdot \left(1 - e^{\left(\frac{h^2}{100}\right)}\right).$$

This semi-variogram is compared with the experimental semi-variogram calculated on the basis of the simulated data in the next figure (Figure 3-16).

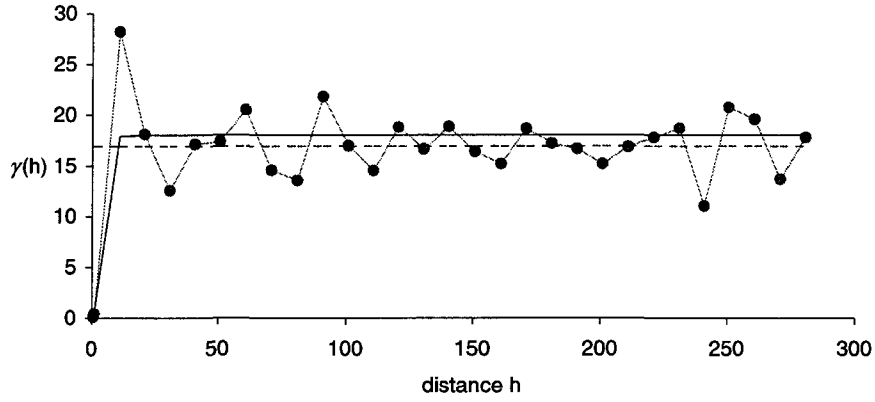


Figure 3-16 Comparison between the semi-variogram model (line) and the experimental semi-variogram (black dots) calculated with the simulated data of the Figure 3-15. The estimated sill c=18 is near the sample variance σ^2_Z=16.92 value (dashed line).

In a more practical way, the parameters for the variogram model can also be directly measured on the simulated profile. We can measure the mean amplitude and a minimum wavelength directly from the samples. In the case presented in Figure 3-15, the mean amplitude is $\bar{a} = \overline{2A}/2 = 5.58$ and the minimum half-wavelength is $w_{min}/2 = 16/2 = 8$. So it is an easy way to find adequate variogram parameters, which are $c = 15.58$ and $r = 8$. These two values are close to those calculated before.

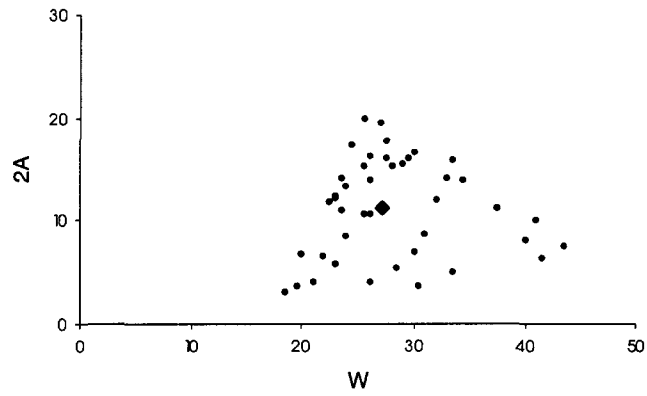


Figure 3-17 Wavelength-amplitude data measured on the simulated profile.

In nature, when a multilayered sequence is deformed, it often happens that a wide variety of geometrical forms and scales of folds is produced due to complex interactions between rock layers of various thickness and mechanical properties. The most remarkable geometrical feature is the so-called polyharmonic folding, where layers are folded with more than one wavelength. Structures are composed of multiple folding phases stacked together.

Such a folding shape can also be expressed as a combination of trigonometric functions. They can be expressed as a sum of various structures of various scales.

$z(x) = z_1(x) + z_2(x) + ... + z_n(x)$, where the $z_i(x)$ are n structures observed at various scales.

As an example, the figure below (Figure 3-18) shows two superposed structures of different size.

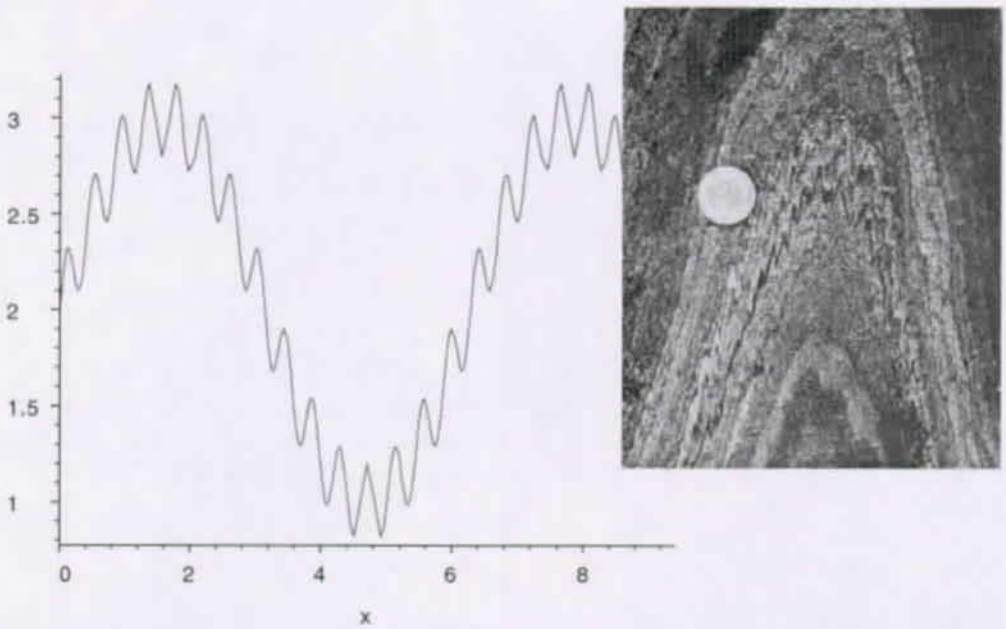


Figure 3-18 A polyharmonic folding composed of two superposed folding phases. The large-scale structure is described by the function $z_1(x)=2+\sin(x)$ and the small-scale structure is described by $z_2(x)=0.1\sin(15x)$. The resulting structure is then $z(x)=z_1(x)+z_2(x)=2+\sin(x)+0.1\sin(15x)$.

The same remarks can be made about the regularity of this functional expression. So, when adding again some random noise to each component, we have the random expression $Z_i(x) = A_i \sin(\Omega_i x)$ for each harmonic.

If we assume that the various harmonics $Z_i(x)$ are statistically independent random functions[4], the resulting variogram describing the whole structure is simply the sum of the variograms $2\gamma_i^*(h)$ of each individual harmonic. We can therefore build a so-called *nested semi-variogram* as:

$$\gamma^*(h) = \gamma^*_1(h) + \gamma^*_2(h) + ... + \gamma^*_n(h) \qquad (3\text{-}18)$$

In this case, we can also estimate the parameters of the various semi-variograms when we know the mean amplitude and the wavelength of each harmonic to obtain a semi-variogram model for the whole structure (see Figure 3-19).

[4] This assumption can be verified if the structures observed at different scales in polyharmonic folds are not correlated.

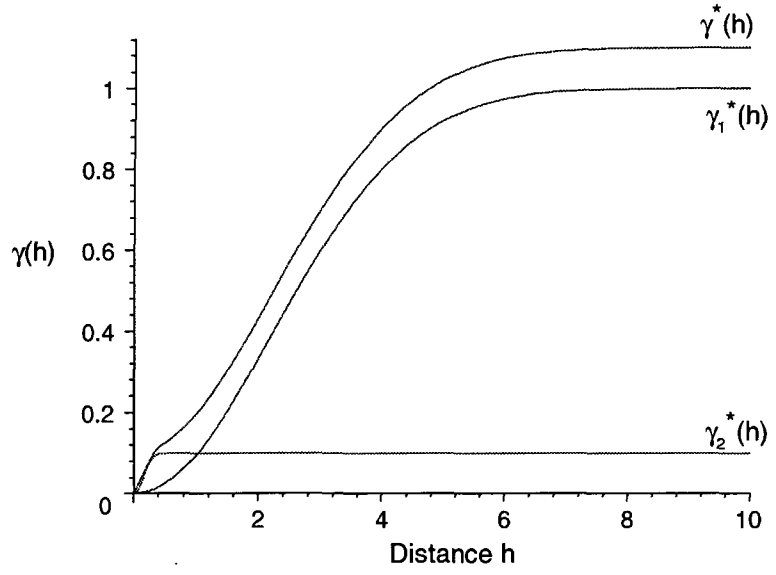


Figure 3-19 The three semi-variograms calculated with the polyharmonic signal presented in Figure 3-18. The semi-variogram $\gamma_1^*(h)$ corresponds to the small-scale structures (c=0.1, r=0.2) and the semi-variogram $\gamma_2^*(h)$ to the large-scale structures (c=1, r=3.14). The resulting nested semi-variogram $\gamma^*(h)= \gamma_1^*(h)+\gamma_2^*(h)$ represents the whole polyharmonic signal.

Empirical variograms can advantageously avoid the issue of both non-stationarity and non-ergodic variogram. But in return, we should be able to justify our choice to neglect experimental data. From a statistical point of view, an experimental variogram is certainly more meaningful, but when the behaviour of a specific variable really matters, a regional variogram could be more representative of the underlying process. As shown in Figure 3-9, an experimental variogram will not show the characteristics of the entire process, but only those underlying the observations available in the limited sampling area. Choosing a regional variogram amounts to represent the essential rather than the anecdote. This is mainly a question of philosophy.

3.5. Bringing Additional Constraints to the Stochastic Model

In addition to the notions of folds' shape, we will now introduce in this section some tectonic notions that can be taken into consideration for the description of spatial variability.

So far, we are able to estimate uncertainties related to each geological interface independently, but it is not convincing to see geological boundaries as single isolated objects (Figure 3-20). We also must consider the way in which each interface interacts with the others. This leads us to account for the various stratigraphic and tectonic constraints that impose interactions between the different interfaces.

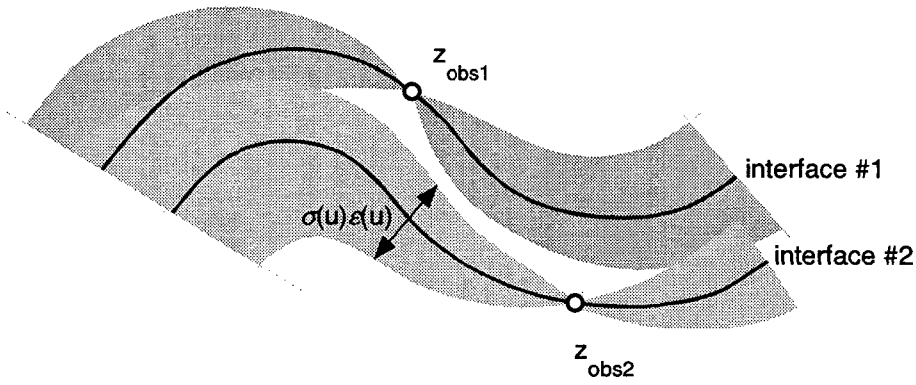


Figure 3-20 Example of a folded layer with a constant thickness and bounded by two interfaces. In this case, the observation P_1 available on interface#1 is not taken into account on the interface#2, and similarly the observation P_2 available on interface#2 is not taken into account on the interface#1.

The most constraining relation between the various interfaces is the constant thickness constraint. If a given layer has a strictly constant thickness (the layer is *isopach*), any information concerning one boundary can be transposed to the other boundary of the layer. Uncertainties about the position of a given interface are therefore not totally independent of what we know about the other interfaces. There must be an interaction between the various random fields, so that the information available about a given interface can influence the uncertainty about the others.

If we know that the thickness of a layer is constant in a given direction, the uncertainty about the position of one edge interface has to influence directly the uncertainty about the position of the other edge (Figure 3-21).

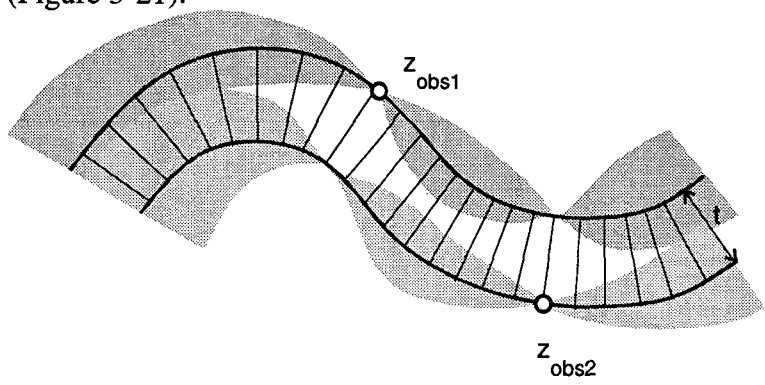


Figure 3-21 The same isopach layer as in the previous figure (Figure 3-20). The layer has a constant thickness t and its bounding surfaces are thus related one to the other. Any information (observed points z_{obs1} and z_{obs2}) located on one interface can thus be transposed to the other one.

In nature, perfectly isopach layers are hardly ever observed and the constraints should be applied less strictly. If the thickness of a layer is not exactly constant throughout the studied area, we then can determine the variance of the thickness and add it as a nugget effect to the model of variability. In this case, information is transposed from one interface to another as if it were soft data (Figure 3-22).

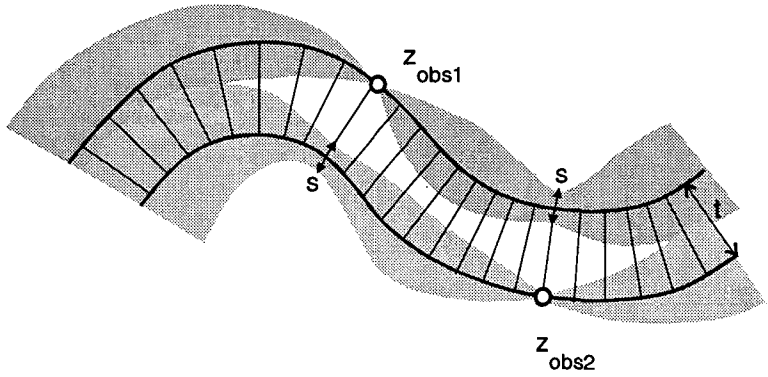


Figure 3-22 A folded layer. The constraint of constant thickness is applied less strictly. Observations (z_{obs1} and z_{obs2}) made along an interface are transposed to the other interface as if it were soft data by adding a nugget s to the model of variability.

3.5.1. Fold Classification

To apply the constraint of constant thickness, one has to know the direction along which information can be transposed from one interface to another. This direction essentially depends on the style of folding. The way in which the thickness of a layer varies around a fold depends on the fold style. This leads us to the fold classification.

So, in order to account for the criterion of constant thickness, we have to distinguish between several tectonic features that can be found in nature.

Folds can be classified in only two types: the *parallel folds*, on the one hand, and the *similar folds* on the other hand. This classification is only based on the measure of constant thickness around the fold, but does not represent the whole variety of the folds' geometry. Although this classification restricted to only two types of folds is simplistic, it has the advantage that the two above-mentioned types of folds are the most remarkable ones. This simple classification was held by structural geologists for many years until it was shown that there could exist folds that have characteristics ranging in-between these two models (de Sitter 1957). A new fold classification using layer thickness was then proposed by Ramsay ((Ramsay 1967)).

Parallel Folds
In a parallel fold, the layer thickness measured orthogonally across the layer is constant throughout the fold (see Figure 3-23). Folds with an almost constant curvature and thus with circular boundaries are usually termed circular folds.

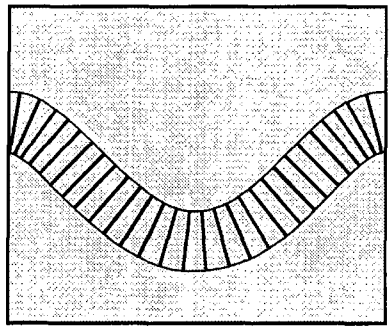


Figure 3-23 Illustration of parallel folds.

In parallel folds, the information will be transposed along a direction perpendicular to the bedding surfaces (such was the case in Figure 3-22).

Similar folds.
In similar folds, the layer thickness measured anywhere in the fold in a direction parallel to the axial surface is constant (see Figure 3-24). A remarkable property of such folds is that all bounding surfaces of similarly folded beds are identical.

Similar folds are formed by simple heterogeneous shear deformations in a direction parallel to the axial surface.

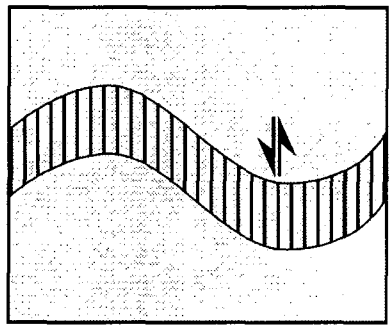


Figure 3-24 Illustration of similar folds produced by heterogeneous simple shear.

In similar folds, the information will be transposed along the shear direction. The same model of variability can be applied to each interface, because these interfaces are of the same shape.

3.5.2. The 2-D Model of Variability

As the title of this section could be misunderstood, we should first be more specific about what we mean by a "geological coordinate system". When introducing this coordinate system, we do not attempt to bring a coordinate transformation in order to detrend or unfold subsurface structures. This coordinate system will rather setup in order to take into account particular directions for the description of the variability.

Often in geostatistics the so-called *omnidirectional variogram* is considered. In this case, the usual Euclidian distance separating two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is taken into consideration:

$$h = \delta(\mathbf{x}_i; \mathbf{x}_j) = \| \mathbf{x}_i - \mathbf{x}_j \| \qquad (3\text{-}19)$$

The omnidirectional variogram is in fact the mean variogram calculated along all directions of the study area. Such a variogram will be well suited if the studied phenomenon is isotropic.

However, most natural phenomena are not so ideally isotropic and the omnidirectional variogram will thus fail to bring an accurate description of such phenomena. If the variogram is inferred from the original Euclidian distance, data pairs will be compared without any regard to the geometry of these structures and the main structural features will be neglected. A detailed description should account for the direction and the ratio of this anisotropy. It would therefore be preferable to bring out a different description along the main directions of anisotropy. So, we should rather consider several *directional variograms*, which are calculated individually along the main directions of a Cartesian coordinate system oriented according to the main directions of anisotropy. This way, we can bring along a more detailed description than the one we would obtain with the omnidirectional variogram.

Folded surfaces present two distinct anisotropic directions along which the variability is fairly different. Along the fold axis, little fluctuations are generally observed and often, when building 3-D geological models, a perfect continuity is assumed along this particular direction. In this case folds are said to be *cylindrical*. On the contrary, the largest fluctuations can be observed perpendicularly to the fold axis. These two directions are the main tectonic directions. Particularly contrasted features can thus be observed along each tectonic direction. It is therefore obvious that we should consider these directions individually in order to describe the variability (or the spatial continuity) of the structures with precision.

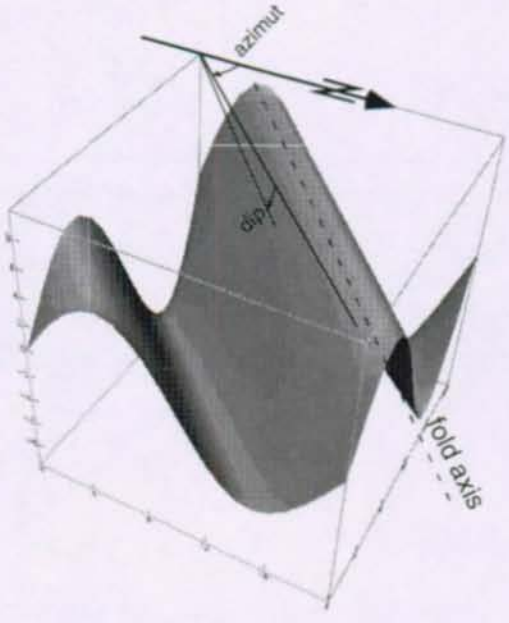


Figure 3-25 An example of cylindrical folds. A perfect continuity can be observed along the fold axis. An azimuth and a dip angle usually define the direction of a fold axis.

This leads us to the idea of considering a new coordinate system, which allows taking into account the anisotropic structures such as those encountered in folded terrains. This approach is close to the "natural coordinate system" introduced by Dagbert, David et al. (1984), McArthur (1988) or mentioned in Deutsch and Journel (1998). Such a "natural coordinate system" may be taken into consideration for mainly two reasons. First, because we want to bring a detailed description of the structures under study, and second, because we must consider the directions along which the probability law is invariant in order to honour the stationarity assumption.

As it is obviously not advantageous to consider the usual (x,y) Cartesian coordinate system, we will henceforward take into account a geological coordinate system $\mathbf{u} = (u,v)$ that is oriented along the main tectonic directions. This particular coordinate system has already been accounted in our notation.

The vector $\mathbf{u}(u,v)$ is oriented tangentially to the geological interfaces, and z being oriented in a perpendicular direction to the interface. The separation distances are not obtained by the usual Euclidian distance measurement, but they are measured as a path length on the interfaces.

This model of variogram has so far been used to describe the variability of structures with a direction perpendicular to the fold axis. However, this approach can also be envisaged to describe the variability along the fold axis. In this case, other parameters should be taken into consideration in order to bring a variogram which is representative of the little fluctuations that are generally observed along this direction. A large range and a small sill value are expected for this variogram. These parameters correspond to large-scale fluctuations of small amplitude.

If different variogram models are applied along these two main tectonic directions, the anisotropy will then be expressed as the ratio $r_{//} / r_{\perp}$ of the range along folds $r_{//}$ and the range perpendicular to fold axis $r_{\perp}$.

In this case, the resulting directional Gaussian semi-variogram model will be expressed as:

$$\gamma(\mathbf{h}) = \gamma(h_{//}, h_{\perp}) = c \cdot \left[1 - \exp\left(-\left(\sqrt{\left(\frac{h_{//}}{r_{//}}\right)^2 + \left(\frac{h_{\perp}}{r_{\perp}}\right)^2} \right)^2 \right) \right] \qquad (3\text{-}20)$$

The directional variogram gives a two-dimensional model of the spatial variability which is oriented along the two main tectonic directions: parallel an perpendicular to fold axis.

If folds are not cylindrical, a model of variability which accounts for large fluctuations along the fold axis can be taken into consideration. In this case, the same previous reasoning (see section 3.4.2) can be made about the wavelength and the amplitude of the fluctuating direction of fold axis.

3.6. Estimating the Local Variance

We now have provided a satisfying variogram model that correctly accounts for the complex geometry of subsurface structures. We are now ready to give a description of uncertainties by estimating the local variability of the structures under study.

At this point, we are able to provide an estimation of the three following statistical moments of the random function that have been previously defined (section 3.2).

$$E\{Z(\mathbf{u})\} = m(\mathbf{u})$$

$$Var\{Z(\mathbf{u})\} = \sigma_Z{}^2(\mathbf{u})$$

$$Cov\{Z(\mathbf{u}_i), Z(\mathbf{u}_j)\} = \sigma_z(\mathbf{u}_i)\sigma_z(\mathbf{u}_j)\rho(\mathbf{u}_i, \mathbf{u}_j)$$

where the covariance is $Cov\{Z(\mathbf{u}_i), Z(\mathbf{u}_j)\} = C(0) - \gamma(\mathbf{h})$ under the stationarity assumption.

At this point, the Gaussian random field $Z(\mathbf{u})$ can be fully described by these statistical moments.

As previously said (see section 3.2), the key for the description of these uncertainties is the estimation of the local standard deviation $\sigma_Z(\mathbf{u})$ - or the local variance $\sigma_Z^2(\mathbf{u})$. At this point, the kriging technique can allow us to obtain an estimation of this parameter.

The kriging technique was set up by Krige (1951) and further formalized by Matheron (1963). This technique originally intended to reconstruct a phenomenon over a domain on the basis of values observed at a limited number of points. The kriging can thus be regarded as an interpolation technique, but with the remarkable difference, compared to other interpolation techniques, that it starts from a statistical model of the phenomenon under study.

As mentioned in the introduction chapter (Chapter 1), there are many different kriging techniques that one can choose. The choice of an adequate kriging technique will mainly depend on whether the mean is known or not. In our case, the mean is known, because we are considering that the geological model is the best guess and thus provides us with the mean $m(\mathbf{u})$ (see section 3.2). The *simple kriging* (SK) is therefore the most adequate technique.

The basic scope of kriging is to provide an estimation Z^* of an unknown value by a linear combination of the n available observations:

$$Z^* = \sum_{\alpha=1}^{n} \lambda_\alpha Z_\alpha \qquad (3\text{-}21)$$

By solving the linear simple kriging system of n equations, we can get the value of the n unknown weights λ_α.

Thanks to the statistical framework in which the kriging technique is defined, it does not only provide us with an estimation of unknown values. It also provides us with the estimation error. This error is often referred to as the kriging variance, which is:

$$\sigma^2_{SK} = E\left\{\left[Z^* - Z_{obs}\right]^2\right\} = C(0) - \sum_{\alpha=1}^{N} \lambda_\alpha C_{\alpha 0} = \sum_{\alpha=1}^{N} \lambda_\alpha \gamma_{\alpha 0} \qquad (3\text{-}22)$$

The kriging can thus bring to an estimation of the sample variance $\sigma_Z^2(\mathbf{u})$ by its error estimation σ^2_{SK}, but the equality between the sample variance and the kriging variance is only true if there is no evidence of a *proportional effect*.

The proportional effect is induced by a relation of proportionality between the sample variance and the value of the sample itself. Typically, ore grade or pollutant concentrations are subject to this proportional effect, because the higher the value is and the more variable the measurement can be.

Fortunately, in our case, no proportional effect should be expected. There is no relation between the value of our study variable and its variance. The variance only depends on the distance towards observations. However, one could point out that, when we study subsurface structures, uncertainties usually increase with depth as if variance was in fact related to the value of the variable by a proportional effect. We can explain this behaviour by the fact that available information is mainly clustered near the surface.

The kriging variance is able to provide an estimation of the variance we need to describe the variability around the predicted position of a geological interface. The kriging technique can thus be directly applied, so that no further development is required. However, more details about kriging can be found in (Journel and Huijbregts 1978), (Isaaks and Srivastava 1989) and (Chilès 1999).

3.7. Numerical Example

The methodology for the estimation of uncertainties about the position of an interface is now completed and this simple numerical example shows how uncertainties are computed according to the methodology described above.

In Figure 3-26, a hypothetical cross section into the subsurface is proposed by a geologist. This model is a circular fold with a bounding interface between two different rock types *A* and *B* (the simplistic - therefore not realistic - shape of this fold was chosen in order to easily measure distances along the boundary).

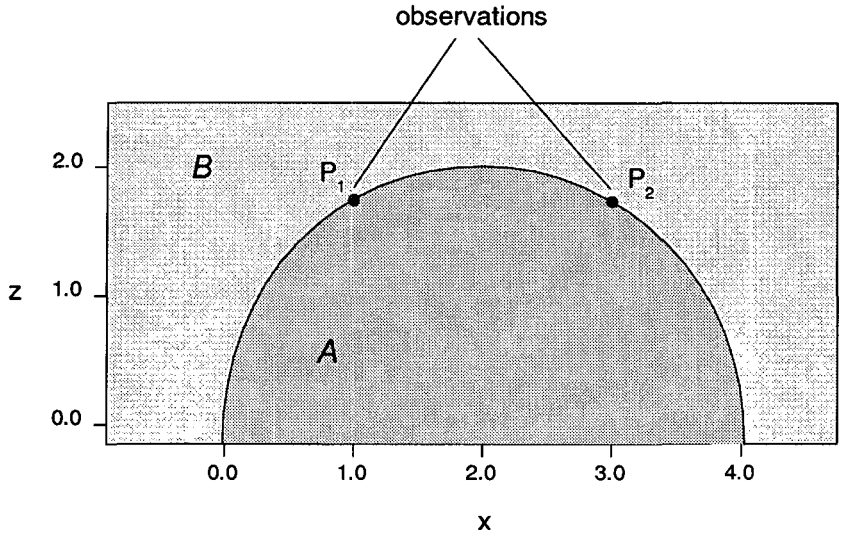


Figure 3-26 A geological model showing two rock types *A* and *B* separated by an interface. Two observations at x=1 and x=3 are available on the boundary.

According to our methodology, this interface is regarded as the best-guessed model and the aim is now to estimate uncertainties about the position of the predicted model.

The fold proposed in this model describes only one possible shape. Other neighboring folds can have different shapes, but in this model we cannot measure the variability of all possible fold shapes.

By comparing the same folds in the field, we introduce an artificial variability, since we know that such folds can have an amplitude ranging from 3.5 to 4.5 and a half-wavelength that can vary between 3.5 and 4.5.

We can therefore estimate the sill and the range of the Gaussian model of a regional variogram. As said in section 3.4, the sill $c = 2$ is given by the variance of the phenomenon and the range $a = 2$ is the mean half-wavelength. The regional variogram, which is a measurement of the variability of the fold shapes, is thus given by:

$$\gamma(h) = 2 \cdot \left(1 - e^{\frac{-h^2}{2^2}} \right)$$

With two available observations P_1 and P_2, the kriging system is the following:

$$\begin{pmatrix} \gamma(h_{11}) & \gamma(h_{12}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & 1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(h_{1p}) \\ \gamma(h_{1p}) \\ 1 \end{pmatrix}$$

where the separation distances h_{ij} are the lengths of the path from i to j along the boundary. We can thus calculate the distances and, as we know the values at the two locations $x_1 = 1$ and $x_2 = 3$, the kriging matrix becomes:

$$\begin{pmatrix} 0 & 0.482 & 1 \\ 0.482 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ 1 \end{pmatrix}$$

At each unknown location p the kriging system can provide the values of the weights λ_1 and λ_2 and the value of the Lagrange parameter μ. For example, at $x = 0$ we calculate $\gamma(h_{1p})$=0.480 and $\gamma(h_{2p})$=1.332. The kriging system will therefore provide:

$\lambda_1 = 1.463$
$\lambda_2 = -0.463$
$\mu = 0.685$

and we can obtain the estimation variance at $x = 0$ by calculating:

$$\sigma^2_{OK} = \lambda_1 \cdot \gamma(h_{1p}) + \lambda_2 \cdot \gamma(h_{2p}) + \mu = 0.769$$

With all locations, from $x = 0$ to $x = 4$, the same calculation is computed and we obtain and estimation of the variability around the predicted position of this interface throughout the study area (Figure 3-27).

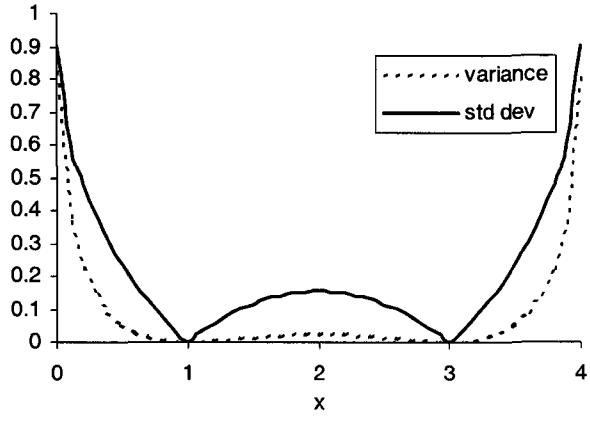


Figure 3-27 Calculated kriging variances and standard deviations along x. We can notice that this variance is equal to zero at known locations x=1 and x=3.

Finally, in the above figure (Figure 3-28), the calculated variability is represented as standard deviation intervals. As the fold is similar, errors are propagating perpendicularly to the interface line and standard deviations are thus placed along these perpendiculars.

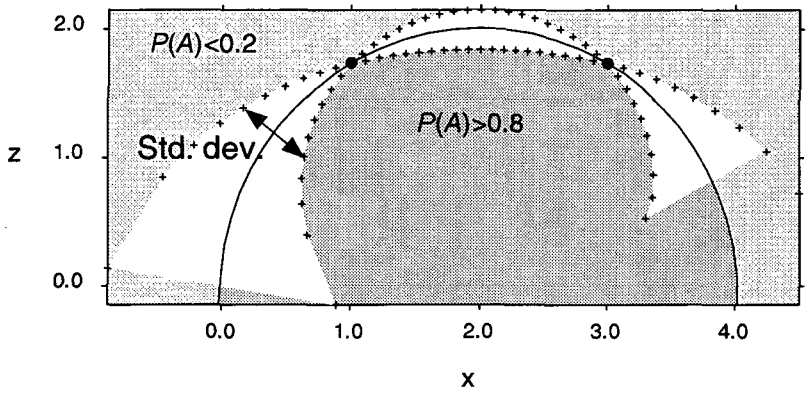


Figure 3-28 The same structure as the one shown in Figure 3-26. Uncertainty bounds are calculated around the best-guessed interface and are represented by the little crosses, which are located at a distance of $\pm\sigma(x)$, perpendicular to the best-guessed interface.

These intervals represent uncertainty bounds, which delimit a $[-\sigma; \sigma]$ confidence interval centred on the predicted position of the interface.

Chapter 4.
Assessment of the Probabilistic Model

In the previous chapter, we described the variability of position for the various interfaces in a subsurface model. This variability describes the characteristics of "past" issues deduced from the study of available information. We are now able to predict possible events by building up a probabilistic model according to the variability of the phenomenon. The probabilistic model is predictive in the sense that it can express all possible future outcomes in terms of probabilities.

The probabilities describe the chances that one has to encounter a certain type of rock at a given location of the study area. The term *occurrence probabilities* will sometimes be used to refer to these probabilities.

At this point, it is necessary to recall the *three classical axioms* of probabilities.

> We admit that for all event E_i of the fundamental set Ω, there exists a value $P(E_i)$ called the probability of the event E_i. For each event of Ω, it is assumed that the value $P(E_i)$ exists and satisfies the three following axioms:
>
> 1) $0 \le P(E_i) \le 1$
> 2) $P(\Omega) = 1$
> 3) $P(\bigcup_i E_i) = \sum_i P(E_i)$, where $E_1, E_2, \ldots$ are mutually exclusive events. (4-1)

In our case, the occurrence of the various rock types are mutually exclusive events because it is obvious that two different rock types cannot coexist at a same location. Therefore, if E_i is the event "Rock type A is found at $z(\mathbf{u})$" and E_j is the event "Rock type B is found at $z(\mathbf{u})$", then the joint event $E_i E_j = \varnothing$ if $i \neq j$.

4.1. Principle

As our study variable $Z(\mathbf{u})$ is a continuous random function, there exists a function f, which has, for all sets of real B, the following property:

$$P(Z \in B) = \int_B f(z)dz \qquad (4\text{-}2)$$

This function is called the *probability density function* (pdf) of the random function $Z(\mathbf{u})$.

In other words, the probability that Z takes a value of B is obtained by integrating this function over B. This means that every problem related to probabilities can be treated by f, when it can be known.

In the previous chapter, we proposed a description for the random function (equation 3-2) describing the position of an interfaces in a structural model. According to the assumptions we made in this chapter about this random function, we know the statistical parameters at any location on the interface:

$$E\{Z(\mathbf{u})\} = m(\mathbf{u}) \text{ and } Var\{Z(\mathbf{u})\} = \sigma^2(\mathbf{u})$$

where $m(\mathbf{u})$ is the predicted position of the boundary and $\sigma^2(\mathbf{u})$ is the local variance describing the variability at the location $\mathbf{u}$.

As $Z(\mathbf{u})$ is assumed to be Gaussian[5], we need only to know its expectation $m(\mathbf{u})$ and its variance $\sigma^2(\mathbf{u})$ for a full account of the searched function. According to these assumptions, we can thus define the following probability density function of the random function $Z(\mathbf{u})$:

$$f(Z(\mathbf{u})) = \frac{e^{-\left(\frac{(z(\mathbf{u})-m(\mathbf{u}))^2}{\sigma^2(\mathbf{u})}\right)}}{\sigma(\mathbf{u})\sqrt{2\pi}}; \quad -\infty < z(\mathbf{u}) < \infty \,. \qquad (4\text{-}3)$$

Probabilities are obtained by the primitive function F(z):

$$\mathrm{F}(z) = \int_{-\infty}^{z} f(y)dy = \frac{1}{2}erf\left(\frac{z}{\sqrt{2}}\right) + \frac{1}{2}\Bigg|_{-\infty}^{z} = P\{Z \in [-\infty; z]\} = P\{Z \le z\} = P\{Z < z\} \qquad (4\text{-}4)$$

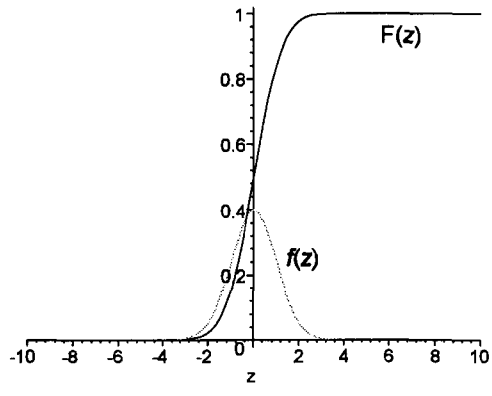


Figure 4-1 Example of calculated probabilities F(z) along $z \in$ [-10; 10]. The probability density function $f(z)$ is Gaussian with m=0 and σ^2=1.

[5] This assumption has been introduced in the section 3.2. This choice will also be discussed in the section 6.2.

When considering a boundary separating two different rock types A and B (as illustrated in Figure 4-2), the probability that $Z(\mathbf{u})$ belongs to one of the rock types A or B at the location $\mathbf{u}$ is thus obtained by integrating the density function as follows:

If $P(A)$ denotes the probability that Z belongs to the rock type A, then this probability is:

$$\mathbf{F}_A(\mathbf{u};z) = \int_{-\infty}^{z} f_A(y)\,dy = \text{Prob}\{Z(\mathbf{u}) \le z\} = P(A)$$

and similarly, the probability $P(B)$ that $Z(\mathbf{u})$ belongs to rock type B is given by:

$$P(B) = \text{Prob}\{Z(\mathbf{u}) \ge z\} = \int_{z}^{\infty} f_A(y)\,dy$$

As A and B are mutually exclusive events and $\Omega = A \cup B$, $P(B)$ is therefore the complementary of $P(A)$, because

$$P(B) = 1 - \int_{-\infty}^{z} f_A(y)\,dy = 1 - P(A) = P^C(A)$$

Both $P(B)$ and its complementary $P(A)$ are presented in the same figure (Figure 4-2):

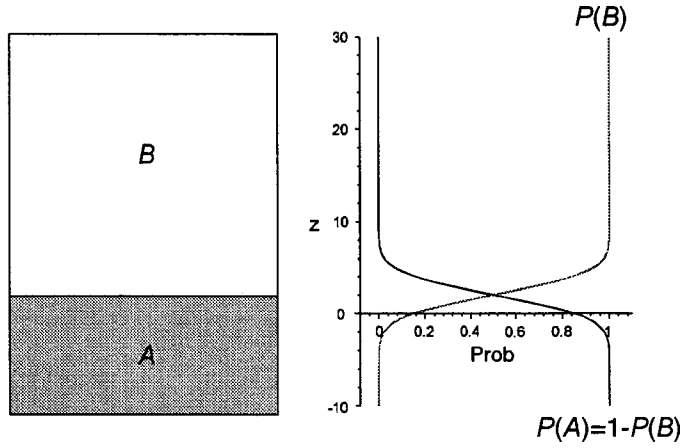


Figure 4-2 Simple model composed of two rock types A and B and their respective probabilities $P(A)$ and $P(B)$ along the vertical axis z. The distribution $f(z)$ is Gaussian with m=2 and σ^2=4.

In this last figure, the expected position of the interface is given by the value $m(u) = 0$, but we can notice that the boundary between two rock types A and B can also be defined as the location where there are equal chances to be in A or in B. Hence, the boundary is located

where occurrence probabilities are such as $P(A) = P(B) = 0.5$ whatever the values m and σ^2 are.

4.1.1. Conditional Probabilities

As we previously saw, spatial continuity implies that elements near one other are expected to have properties that are somewhat similar. Thus knowing the property at a given location implies that uncertainties around this observation are diminishing. The variability of $Z(\mathbf{u})$ is thus conditioned by the n available observations $\{ z_1(\mathbf{u}), z_2(\mathbf{u}), .., z_n(\mathbf{u}) \}$ and the local variance $\sigma^2(\mathbf{u})$ (or the standard deviation $\sigma(\mathbf{u})$) thus depends on the neighbouring observations according to the spatial correlation.

Probabilities are hence conditioned by available information and are in fact *conditional probabilities*:

$$P\{Z(\mathbf{u})|(n)\},\ n=1,..,N \qquad (4\text{-}5)$$

In this case, the probabilities are contrasted around observed locations (Figure 4-3).

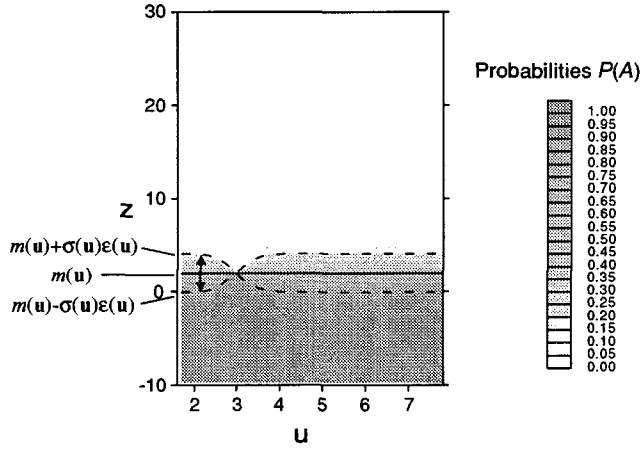


Figure 4-3 Two-dimensional probability field of occurrence of the rock type A conditioned by the available observation at u=3 calculated with our CREER program. The distribution $f(z)$ is the same Gaussian function centred at $m(u)$=2. Local variances are obtained by ordinary kriging using a Gaussian semi-variogram with a sill c=4 and a range r=0.5. Dashed lines represent uncertainty bounds of $\pm\sigma$ around the expected position of the interface.

Due to conditioning, uncertainty bounds tighten up near available observations. Any observation lying within the correlation range will reduce the local variability. On the contrary, if the separation distance is greater than this range, then the variability is maximum. In the previous figure, for example, the value of the range was r = 0.5. This means that locations beyond this distance have the maximum magnitude of uncertainties.

4.1.2. Representing Uncertainties

We can calculate exactly the probabilities describing the occurrence of the various rock types for any location **u** of the study area *D*. It is now possible to build a two- or three-dimensional probability field.

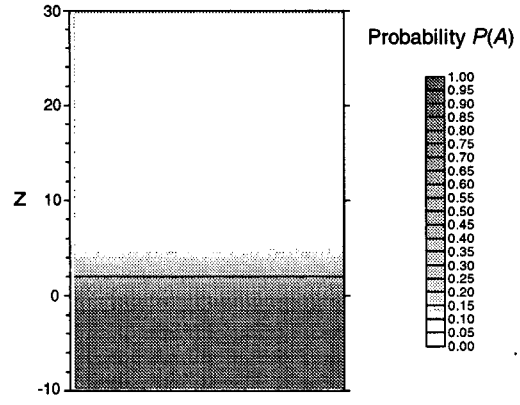


Figure 4-4 Two-dimensional probability field calculated with our CREER program. As previously, probabilities are calculated with the same simple model composed of two rock types *A* and *B* and the distribution $f(z)$ is the same Gaussian function with m=2 and σ^2=4. Uncertainties are here presented as a gray scale colour filling. Dark colours represent areas where the rock *A* is the most probable rock type and, on the contrary, light colours represent areas where there are few chances to encounter the rock type *A*.

However, the results can be more easily read when the probabilities are represented with uncertainty bounds. These bounds are defined by a confidence interval of probability. We can choose any threshold value for this interval, but the standard deviation usually is a commonly chosen threshold value. The uncertainty bounds are then defined as:

$$z_{\text{upper bound}}(\mathbf{u})=m(\mathbf{u})+\sigma(\mathbf{u}) \text{ and } z_{\text{lower bound}}(\mathbf{u})=m(\mathbf{u})-\sigma(\mathbf{u}) \qquad (4\text{-}6)$$

In terms of probabilities, these uncertainties bounds are defined as:

$$\text{Prob}\{Z(\mathbf{u}) \le m(\mathbf{u})+\sigma(\mathbf{u})\} \le 0.84$$
$$\text{and } \text{Prob}\{Z(\mathbf{u}) \ge m(\mathbf{u})-\sigma(\mathbf{u})\} \ge 0.16$$

These uncertainty bounds are cut off limits and only probabilities contained in the interval [0.16; 0.84] are taken into consideration. This interval represents 84% of all possible realisations.

The next figure illustrates how the result can be represented with uncertainty bounds.

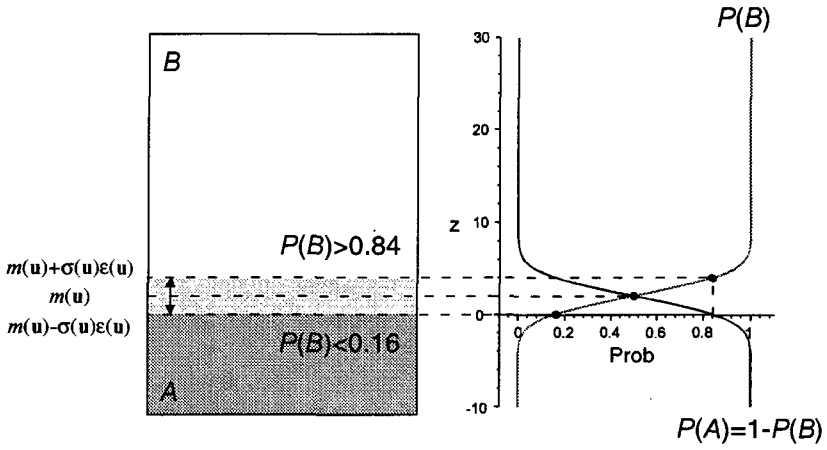


Figure 4-5 Still the same simple model composed of two rock types A and B. The distribution $f(z)$ is the same Gaussian function with m=2 and σ^2=4. The uncertainty bounds are here represented as two dashed lines at $\pm\sigma$, hence at z=0 and z=4.

The uncertainty bounds are threshold values for an arbitrary confidence interval. They should therefore not be interpreted as strict limits for uncertainties, but should rather be regarded as a mean of identifying areas with a large or small variability. The thickness of the confidence interval is a direct measure of uncertainties and can therefore indicate areas where the geological model is reliable and areas where the model is uncertain.

Uncertainties can be represented either by what could be called "fuzzy interfaces", such as shown in Figure 4-4, or by "thick interfaces", as shown in Figure 4-5. In fact these representations are two equivalent ways of representing the same result. The choice of a fuzzy or a thick interface representation mainly depends on the complexity of the structures under study. Thick interfaces are well-suited to represent uncertainties related to simple structures, but fuzzy interfaces will be preferred for the representation of uncertainties related to complex structures such as those encountered in the Lötschberg study case (see Chapter 5).

4.2. Combining Uncertainties According to Stratigraphic Relations

In the previous section, we introduced the basic principle of how to calculate probabilities of occurrence when a single interface separates the domain into two sub-areas A and B. We must now build up a method for the calculation of probabilities with one, two or more interfaces according to geological constraints.

The basic notions of the building of a geological model with a succession of depositional and erosion surfaces were introduced in section 2.3. As the probabilistic model is closely related to the construction process of the geological model, we will consider the same process to calculate the probabilities of encountering the various rock types. We will thus account for the stratigraphic relations and recall the notions of depositional and erosion surfaces.

4.2.1. Depositional Surfaces

A sequence of deposited rocks is formed by piling layers one on top of the others. In the same way, probabilities are first calculated starting from the bottom layer to the upper layer by adding successive layers to the model. As we rise in the sequence, the probabilities are updated at each addition in order to account for the presence of the new layer. This is thus an *iterative process* of updating probabilities as piling a sequence of depositional layers.

Such an iterative process was first introduced by (Tacher and Parriaux 1997) and we propose here an improved version of this process, which can be described as follows:

Let us consider a geological model made of 4 deposited layers (Figure 4-6).

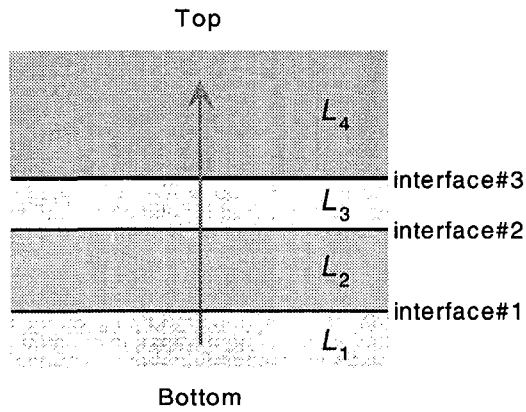
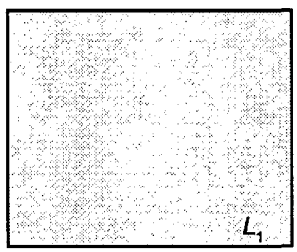


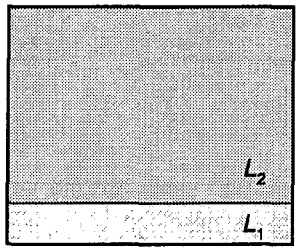
Figure 4-6 A sequence of four deposited layers.

The probability $P_j(L_i)$ is the probability to find the layer L_i, where $i = 1,..,N$, which is calculated at the j^{th} iteration. Each step in the iterative process is detailed as follows:



First iteration ($N = 1$)
(one layer L_1 and no boundary)

$P_1(L_1) = 1$



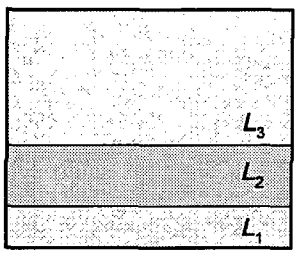
Second iteration ($N = 2$)
(two layers L_1 and L_2, one interface)
Starting from the bottom of the model, the first interface separates the layer L_1 from layer L_2. This interface is the upper boundary of layer L_1 and, for instance, L_2 has no upper boundary.

$F_1(\mathbf{u};z)$ is the cdf related to the first interface at the location **u**.

$$P(L_2) = F_1(\mathbf{u};z)$$

The probability to find the layer L_1 is updated according to the presence of the new layer L_2. New probabilities are:

$$P_2(L_1) = P_1(L_1)*[1- P(L_2)]$$
$$P_2(L_2) = P(L_2)$$

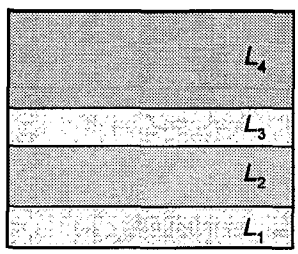


Third iteration ($N = 3$)
(three layers L_1, L_2 and L_3, two interfaces)
A second interface is added to the model. This interface is the upper boundary of layer L_2.

$$P(L_3) = F_2(\mathbf{u};z)$$

The probability to find the layer L_2 are updated according to the presence of the new layer L_3. New probabilities are:

$$P_3(L_1) = P_2(L_1)$$
$$P_3(L_2) = P_2(L_2)* [1- P(L_3)]$$
$$P_3(L_3) = P(L_3)* [1- P_2(L_1)]$$



Fourth iteration ($N = 4$)
(Four layers L_1, L_2, L_3 and L_4, three interfaces)
A third interface is added to the model. It is the upper boundary of layer L_3.

$$P(L_4) = F_3(\mathbf{u};z)$$

The probability to find the layer L_3 are updated according to the presence of the new layer L_4. New probabilities are:

$$P_4(L_1) = P_3(L_1)$$
$$P_4(L_2) = P_3(L_2)$$
$$P_4(L_3) = P_3(L_3)* [1- P(L_4)]$$
$$P_4(L_4) = P(L_4)* [1- P_3(L_1)- P_3(L_2)]$$

The generalized algorithm for the iterative process can be written as follows:

$$\begin{cases} P_N(L_i) = P_{N-1}(L_i) & \text{where } i = 1, \ldots, N-2 \\ P_N(L_{N-1}) = P_{N-1}(L_{N-1}) \cdot [1 - P(L_N)] & \\ P_N(L_N) = P(L_N) - \sum_{j=1}^{N-2} P(L_N) \cdot P_{N-1}(L_j) & \end{cases} \quad (4\text{-}7)$$

4.2.2. Erosion Surfaces

From a geological point of view, erosion surfaces do not belong to a sedimentary process, but are rather perceived as a discontinuity in a sedimentary process and indicate the end of a sedimentary sequence. Erosion surfaces are thus singular objects and should not be handled in the same way as depositional interfaces. In our point of view, erosion surfaces do not delimit any layer, but separate tectonic or stratigraphic blocks.

Typically, there are two geological features which can be represented as erosion surfaces when modelling subsurface structures with EarthVision®. These features are: faults and discordance surfaces (see also section 2.3.3).

Faults are modeled using an erosion surface as shown in Figure 4-7.

When the $(N\text{-}1)^{th}$ interface is an erosion surface, the probabilities are updated as follow:

$$\begin{cases} P_N(L_i) = P_{N-1}(L_i) \cdot [1 - P(L_N)] & \text{where } i = 1, \ldots, N-1 \\ P_N(L_N) = P(L_N) & \end{cases}$$

Updating uncertainties with an erosion surface does not require to carefully follow the succession of interfaces used when building the geological model. If the i^{th} interface is an erosion surface of a series of N interfaces, we can obtain the same result if we consider this interface at the end of the process. In this case, the updating procedure is:

$$\begin{cases} P_N(L_j) = P_{N-1}(L_j) \cdot [1 - P(L_N)] & \text{for all } j < i, \ j = 1, .. N \\ P_N(L_k) = P_{N-1}(L_k) \cdot P(L_N) & \text{for all } k > i, \ k = 1, .. N \end{cases} \quad (4\text{-}8)$$

4.2.3. Verifications

The algorithms for the calculation of probabilities resulting from a sequence of depositional or erosion surfaces are now defined. It remains to be checked if the axioms of probabilities (equations 4-1) are honoured.

- The first axiom assert that $0 \leq P(E_i) \leq 1$. In our case, this axiom is always honoured, because the updating process is computed either by multiplying two probabilities i.e.:

$$0 \leq P_{i-1}(L_{i-1}) \cdot [1 - P(L_i)] \leq 1$$

or by subtracting residual probabilities. In this case:

$$0 \leq P(L_N) - \sum_{j=1}^{N-2} P(L_N) \cdot P_{N-1}(L_j) \leq 1 \text{ , because } P(L_N) \geq \sum_{j=1}^{N-2} P(L_N) \cdot P_{N-1}(L_j)$$

- The second and the third axiom state that $P(\Omega) = P(\bigcup_i E_i) = \sum_i P(E_i) = 1$.

In our case, we can check the resulting sum of probabilities for $N = 3$, for example:

$$\begin{aligned} P(\Omega) &= P_3(L_1)+P_3(L_2)+P_3(L_3) \\ &= P_2(L_1)+P_2(L_2)*[1- P(L_3)]+P(L_3)*[1- P_2(L_1)] \\ &= 1-P(L_2)+P(L_2)*[1-P(L_3)]+P(L_3)-P(L_3)*[1-P(L_2)] = 1 \end{aligned}$$

This shows that the second and the third axiom are also honoured. This result can also be verified with any number of iteration N.

4.2.4. The Case of Discontinuous Rock Masses

What we call discontinuous rock masses are bodies interrupted in their lateral extension.

As each interface separates a rock mass in our modelling process, a geological model containing N interfaces will thus always contain N+1 rock types (Mayoraz 1993). Some discontinuous structures such as overturned folds or lenticular bodies need more than one boundary. So modelling such structures will generate more rock types than the number that it can be observed in reality (see Figure 2-5 and Figure 2-6).

This artefact must be suppressed and the case of discontinuous rock masses thus needs a particular, but simple operation to get rid of any artificially generated rock body. This purpose is reached as follows:

Let us suppose that a model consists of N interfaces. There will thus be N+1 rock masses. Let us now suppose that the i^{th} body is a lenticular rock-mass surrounded by the same rock type. Constraints in building complex structures will force us to create two different rock masses on both sides of the i^{th} body. Rock masses $(i\text{-}1)^{th}$ and $(i\text{+}1)^{th}$ will thus be two distinct rock masses surrounding the discontinuous body.

As the i^{th} body is in reality enclosed within the same rock mass, the whole enclosing rock mass L^*_{i-1} (arbitrarily indexed as i-1) can thus be considered as $L_{i-1} \cup L_{i+1}$. Calculated probabilities for the whole rock mass is therefore obtained with:

$$P^*_N(L_{i-1}) = P_N(L_{i-1}) + P_N(L_{i+1}). \qquad (4\text{-}9)$$

In fact, we can sum any probability in order to have the joint probability of a bulk rock mass as following:

$$P(\bigcup_i L_i) = \sum_i P(L_i) \qquad (4\text{-}10)$$

This operation is illustrated by the two following examples:

In the next figure, the two compartments are separated by a fault. When building such a model with EarthVision, the two compartments are completely disconnected in the sense that rocks that are present in both sides of the fault are not related one to the other.

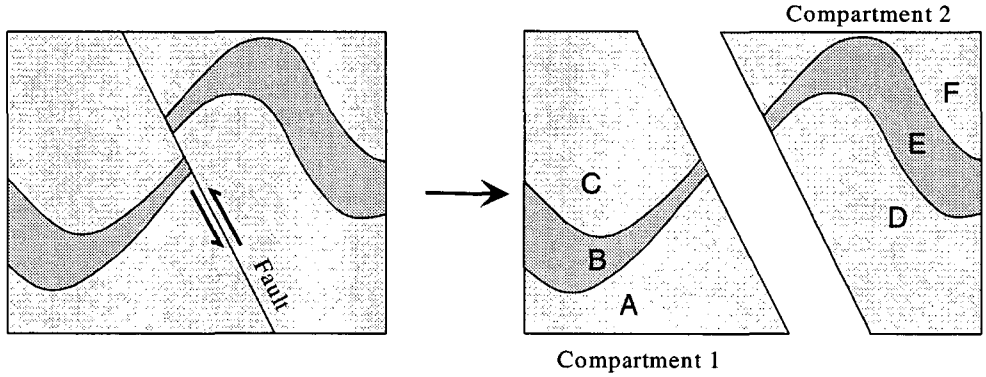


Figure 4-7. A faulted model. When building such a model with EarthVision, the two compartments are completely disconnected in the sense that rocks that are present in both sides of the fault are not related one to the other.

In this case, the occurrence of the layer is obtained by summing the probabilities as follows:

$$P^*(A) = P(A) + P(D),$$
$$P^*(B) = P(B) + P(E),$$
$$P^*(C) = P(C) + P(F).$$

The similar operation can also be computed in the case of discontinuous rock masses such as overturned folds (Figure 2-5) and rock lenses (Figure 2-6).

4.2.5. Numerical Example

In this section, an application illustrates how probabilities are calculated with our iterative process. A simple 2-D geological case is represented by four horizontal interfaces, which are separating five different rock types (A to E).

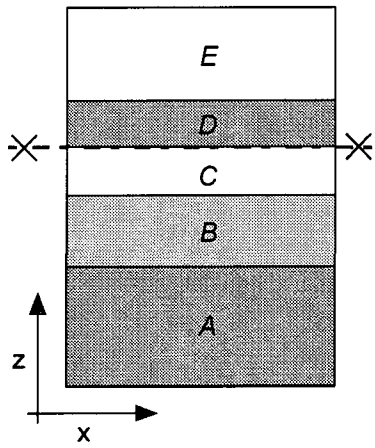


Figure 4-8 A simple 2-D geological model made of four interfaces separating five layers.

The stratigraphic sequence, alike the one we would create with the EarthVision® modeller, is presented in the next figure:

Rock	Relation	Interface
E	Depositional	Interface#4
D	Unconformity	Interface#3
C	Depositional	Interface#2
B	Depositional	Interface#1
A		

Figure 4-9 The stratigraphic sequence of the geological model presented in the last figure. The sequence starts from the bottom layer (layer A) and successive layers are deposited one on top of the others. The stratigraphic relation related to each interface is mentioned in the middle column.

We will illustrate the calculation of probabilities in three stages:

1° The first deposition series,

Three horizontal layers A, B and C are deposited during the same sedimentary process in the same stratigraphic block.

- First deposition

Interface#1, separating A from B, is described by a horizontal line at $z = 2$. Around this interface a Gaussian probability density function with parameters $m = 2$ and $\sigma = 2$ expresses the variability.

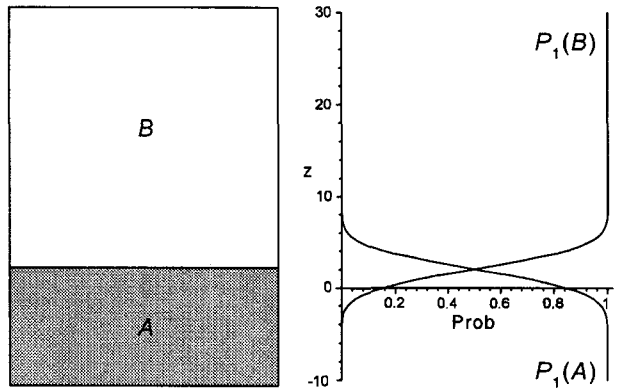


Figure 4-10 Illustration of the first step of the probability calculation. On the left, the conceptual model of the two layers *A* and *B*. On the right, the probabilities calculated for the two layers model according to the density function described above.

- Second deposition

In a second step, the interface#2, separating *B* to *C*, is considered. This interface is described by a horizontal line at $z = 10$. Around this interface a Gaussian pdf with parameters $m = 10$ and $\sigma = 3$ expresses the variability.

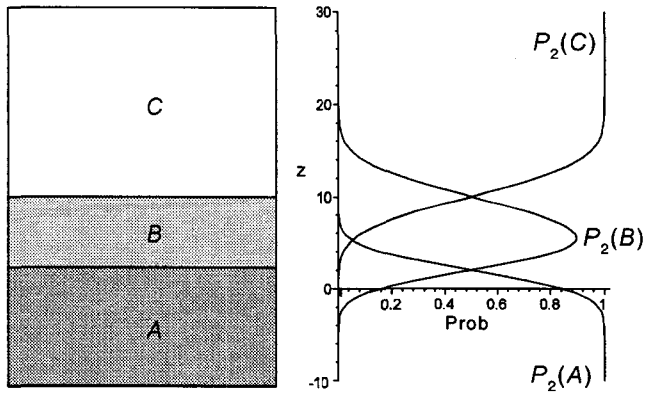


Figure 4-11 Illustration of the second step of the probability calculation. An additional layer *C* is pilled on top of layer *B*. Probabilities are updated so that they can account for the presence of the new layer *C*. On the right, a profile describes the evolution of the probabilities *P(A)*, *P(B)* and *P(C)* describing the occurrence of *A*, *B* and *C* along the vertical coordinate.

2° Erosion

The top of the sedimentary sequence is eroded by a horizontal erosion surface (interface#3). Uncertainty about the position of this interface is expressed by a Gaussian density function with parameters $m = 10$ and $\sigma = 3$. This erosion surface is the upper boundary of the first sedimentary sequence and it separates the layer *C* from the body *D* which lies above the erosion surface.

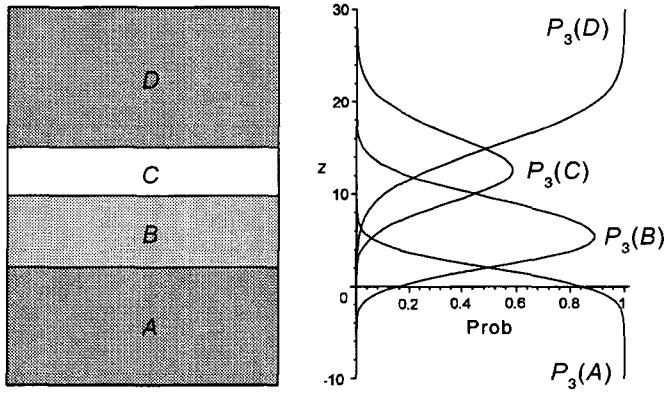


Figure 4-12 Illustration of the resulting calculation for the second stage. The top of the stratigraphic block is eroded along the erosion surface and the following stratigraphic block *D* is deposited above the erosion surface.

3° Deposition

this stage, the layer *E* is deposited on top of layer *D* in the second stratigraphic block, during a second sedimentary process.

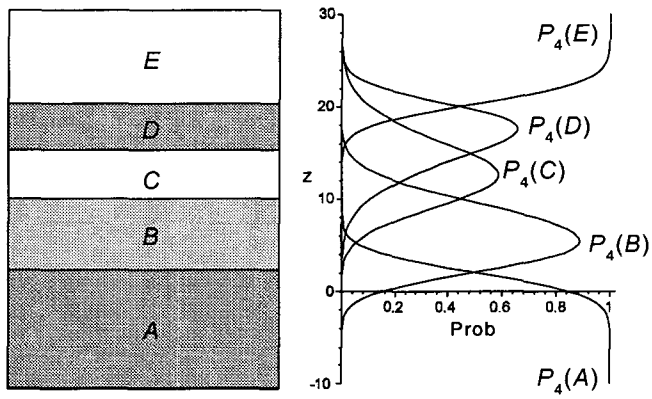


Figure 4-13 Illustration of the resulting model. All layers *A*, *B*, *C*, *D* and *E* have been deposited and the corresponding probabilities, which are shown in the profile on the right, are the resulting probabilities for our uncertainty model.

In this example, we have considered each interface in the order in which it appears in the model, from the bottom to the top:

Interface#1 (depositional), interface#2 (depositional), interface#3 (erosion), interface#4 (depositional).

But, as previously said, it can also be considered at any moment of the calculation. Like in following sequence:

interface#1 (depositional), interface#2 (depositional), interface#4 (depositional), interface#3 (erosion).

As the study area is made of a set of rock masses ranging from *A* to *E*, The joint probability $P(A)+P(B)+P(C)+P(D)+P(E)$ should then be equal to 1, according to the first axiom of probabilities. This is an easy way to check the validity of the calculation.

Chapter 5.
Testing the Methodology

5.1. Applying the Methodology to a Simple 2-D Case

The methodology is first tested on a simple 2-D case. This study case is made of a sequence of three deposited rock layers, which are then eroded by a channel erosion (Figure 5-1).

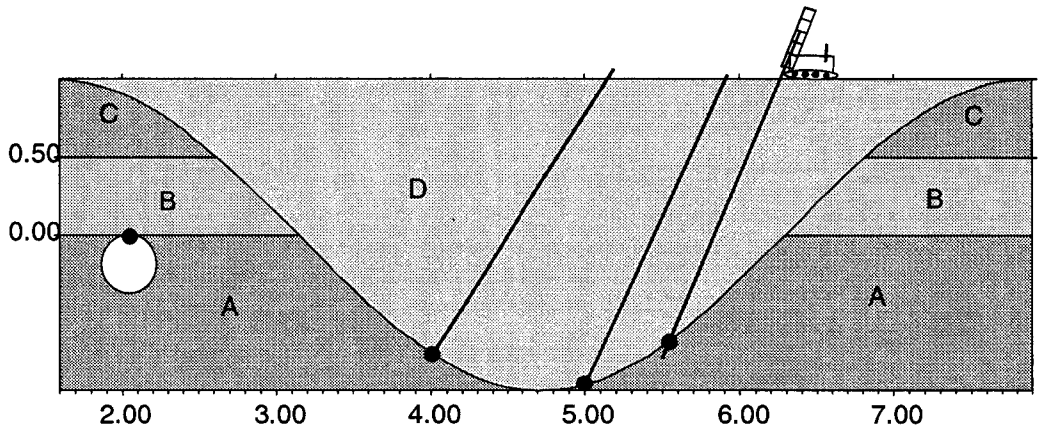
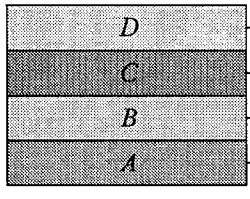


Figure 5-1 A simple 2-D study case. A series of horizontally deposited layers (rock *A*, *B* and *C*) and a layer above the unconformity (rock *D*). Four observations are available (black dots): Three observations through boreholes (center) and one in a tunnel (left)

The stratigraphic sequence of this study case is detailed in Figure 5-2.



Rock	Relation	Interface
D	Unconformity	Interface#3
C	Depositional	Interface#2
B	Depositional	Interface#1
A		

Figure 5-2 The stratigraphic sequence of the above simple study case.

A simple model of variability was taken into consideration for this example. The variability was assumed to be the same for all of the three interfaces of the geological model. The same variogram model was thus applied to each interface with the following parameters:

a Gaussian function model, a sill $c = 1$, and a range $a = 1$. The resulting model is $\gamma(h) = 1 - e^{-h^2}$.

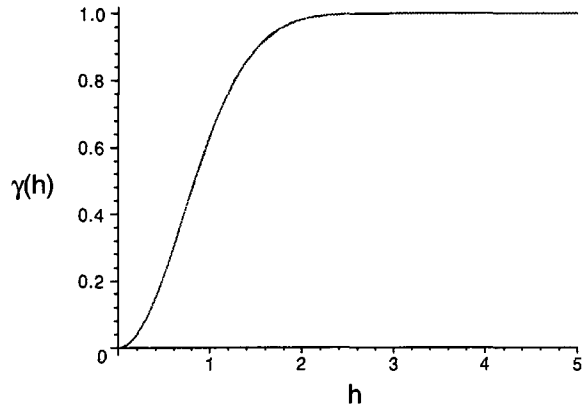


Figure 5-3 The semi-variogram model $\gamma(h) = 1 - e^{-h^2}$.

The above sill parameter c was chosen in order to represent a maximal fluctuation of 1 unit around the best guess. This value corresponds to the amplitude of the spatially varying structures of this study case. The range parameter $a = 1$ was chosen in order to account for 2-units-long structures, which is actually smaller than the size of the channel in Figure 5-1.

Four observations are available in Figure 5-1. All these observations were considered as hard data (thus free of error), except one which was considered as a soft datum: this observation was available at x = 4 with an initial standard deviation s = 0.3.

The thickness of layer *B* was considered as constant. The observation available at the interface between rock *A* and rock *B* can thus be transposed to the opposite interface of this isopach layer.

With the above information, one has all required parameters to describe a model of variability. One can calculate the probability fields related to this model of variability. These probabilities describe the occurrence of the various rocks in the study area.

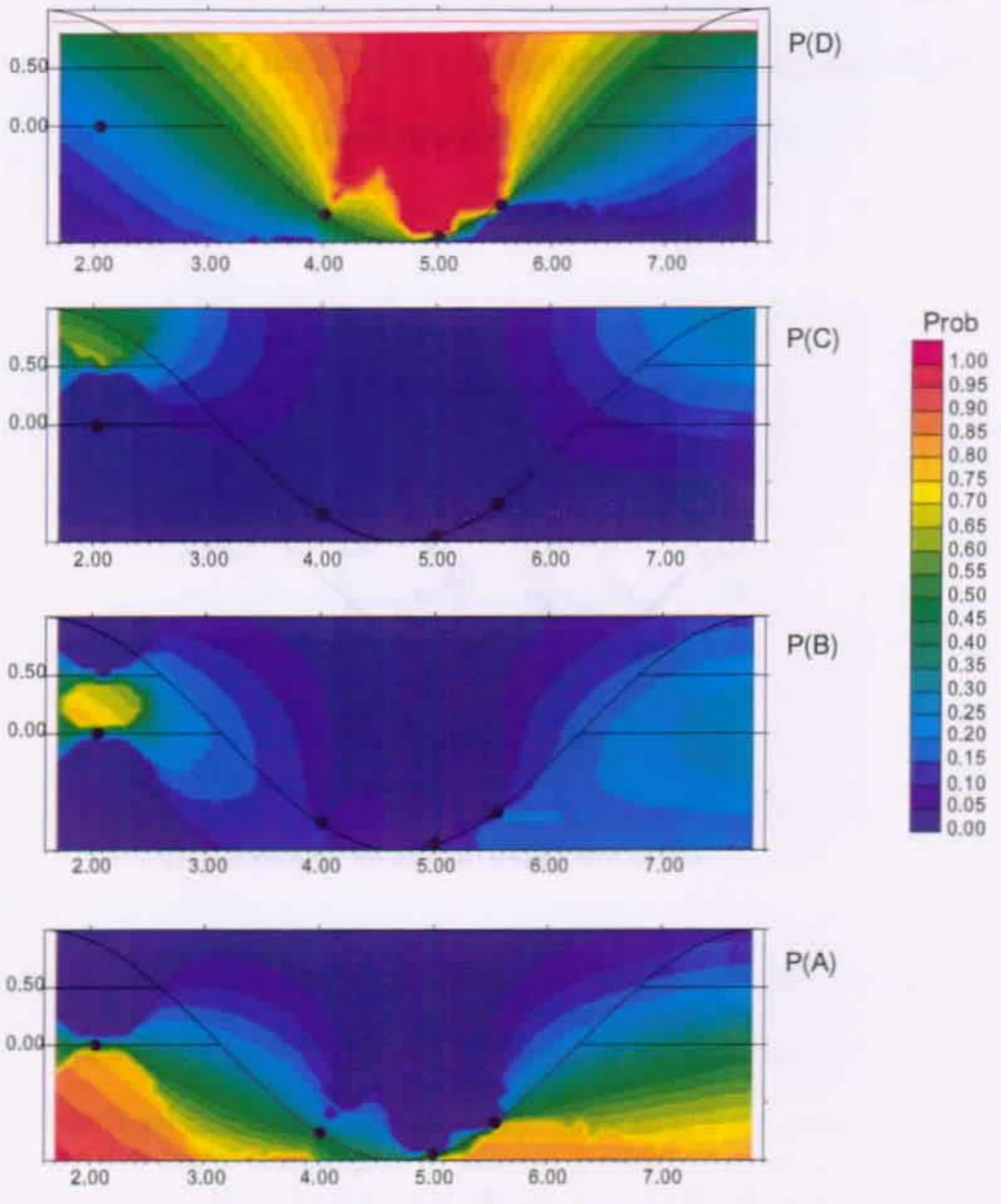


Figure 5-4 Resulting probabilities calculated with the CREER program on a 20x64 grid.

In this last figure (Figure 5-4), resulting probabilities are calculated according to the best-guessed geological model (Figure 5-1) and according to the available observations. These probabilities are related to a model of variability that takes into consideration both the size of the subsurface structures as well as the various geological constraints such as isopachy.

5.2. Applying the Methodology to a 3-D Real Case: The Lötschberg Tunnel (Switzerland)

5.2.1. Introduction

On November 29th 1998, the Swiss voters decided that the Helvetic Confederation should found two new north-south railway projects through Switzerland and thus approved the NEAT/AlpTransit decision. The planned north-south railway axes, which are the Lötschberg and the Gotthard base lines, will substantially contribute to reduce journey times for long-distance rail traffic.

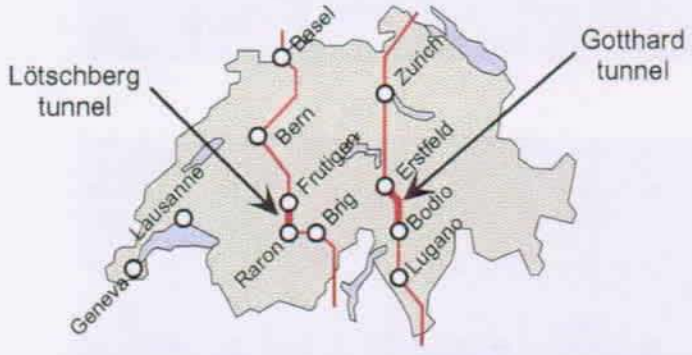


Figure 5-5 A map of Switzerland with the two Gotthard and Lötschberg-Simplon axes.

The construction of these two tunnels, which will reach a total length of 92 km, is the most important civil work ever undertaken in Switzerland. These two giant constructions, scheduled to be completed in 2012, represent a considerable technical challenge.

The Lötschberg base tunnel will be about 35km long and will connect the cities of Bern and Brig by crossing the entire Lötschberg massive. A single-track tunnel will be constructed between the northern portal of Frutigen and the two southern portals of Raron and Steg. A second single-track tunnel will be constructed in a second stage between Miholz and Ferden alis.

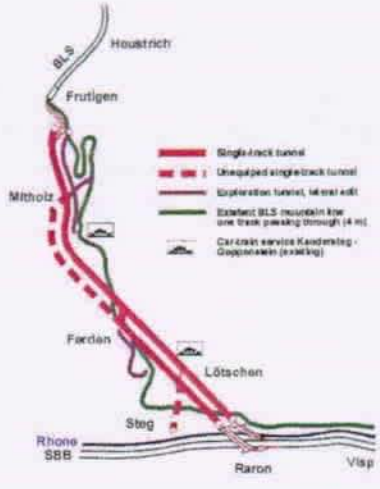


Figure 5-6 The Lötschberg project overview from BLS + Partners.

The simplified stratigraphy in the 3-D model of the area near the southern portal is composed of seven main rock types (Figure 5-7). In fact, it also corresponds to the stratigraphy used in the preliminary geological survey for the Lötschberg tunnel (Kellerhals and Isler 1998).

The Triassic rocks and the St.-German's landslide are the most problematic formations for tunnel construction in the area near the southern portal. It is thus important to determine the depth of the St.-German's landslide to be sure that the tunnel drive will not encounter this terrain. It is also important to have an accurate idea of the location of the Triassic rocks along the planed tunnel drive.

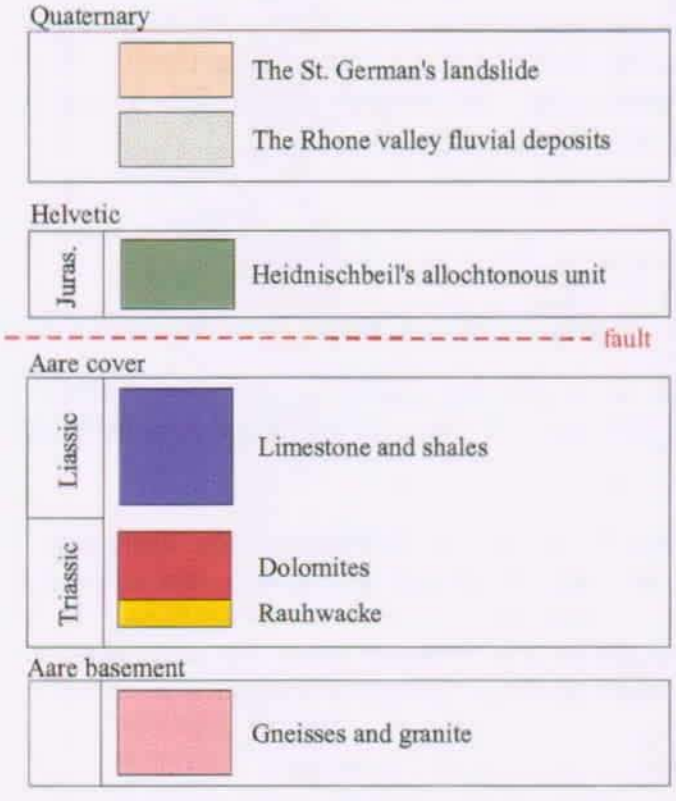


Figure 5-7

The simplified stratigraphy of the main rock types in the area near the southern portal of the Lötschberg tunnel.

The dominant structures in the basement and the cover in the southwestern Aare massive result from alpine folding on the kilometer scale (Dolivo 1982). Folds are nearly cylindrical in the study area and their axes are gently plunging to the SW with an azimuth of 230° and a dip angle of 20°. In this area, the direction of the fold axes are parallel to the measured mineral lineation (Dolivo 1982), as it is shown in the next figure (Figure 5-8).

5.2.2. Geology Along the Lötschberg Tunnel Drive

Relatively good rocks are expected for much of the drive throughout the Aare massive. From the northern portal, the tunnel will cross a variety of successive sedimentary deposits of the Helvetic nappes, after which it will go through the entire Aare massive basement - mainly composed of gneisses and granite -, to finally reach the southern autochthonous Liassic and Triassic sediments.

Despite generally favourable conditions, some particular zones may cause technical problems. The Jungfrau wedge located near the Ferden lateral adit and the Triassic evaporite rocks of the southern portal are the most problematic formations for construction. For these two zones, the structural prediction is critical and unfortunately uncertainties about subsurface structures are important for such a deep construction (Dudt and Descoeudres 1999). These two areas, therefore, represent good study cases to test our methodology on. This is why we chose to consider one of them, namely the southern portal, for the test.

5.2.3. Geology of the Zone Near the Southern Portal of Raron

The area near the southern portal shows complex geological structures. For this reason, we chose to test our methodology on this zone. This way, we can have a complete overview of all possible structural problems.

A first geological study was done in the area by (Schenker 1946). Besides an exhaustive description of the various rock types that are present in this area, this study gives the first description of subsurface structures through a series of cross sections. Then, the study conducted by (Dolivo 1982) brought new information about the area. This study mainly concerns structural and tectonic observations and gave a new subsurface interpretation in the light of modern structural geology. Later on, between 1991 and 1997, a preliminary geological survey for the Lötschberg tunnel project was realized. Two dozen boreholes of 1.3 to 2.6 km long and a 9.5 km pilot tunnel were driven along the planned track of the base tunnel. This survey also brought new information about the subsurface structures near the southern portal thanks to three boreholes that were located in this area (Ziegler 1997).

The stratigraphy

The southern portal is located in the SW part of the Aare massive. In this area, the subsurface is mainly composed of gneisses of the crystalline basement and of its Triassic and Liassic sedimentary covers.

We focused our attention on the rock masses that are the most problematic for tunnel construction. The model is built according to some stratigraphic simplifications, that is to say:

- No distinction between the Marls of the upper Lias and the lower Liassic Limestone.
- Extremely thin layers of Triassic Quartenschiefer and Quartzite have been neglected.

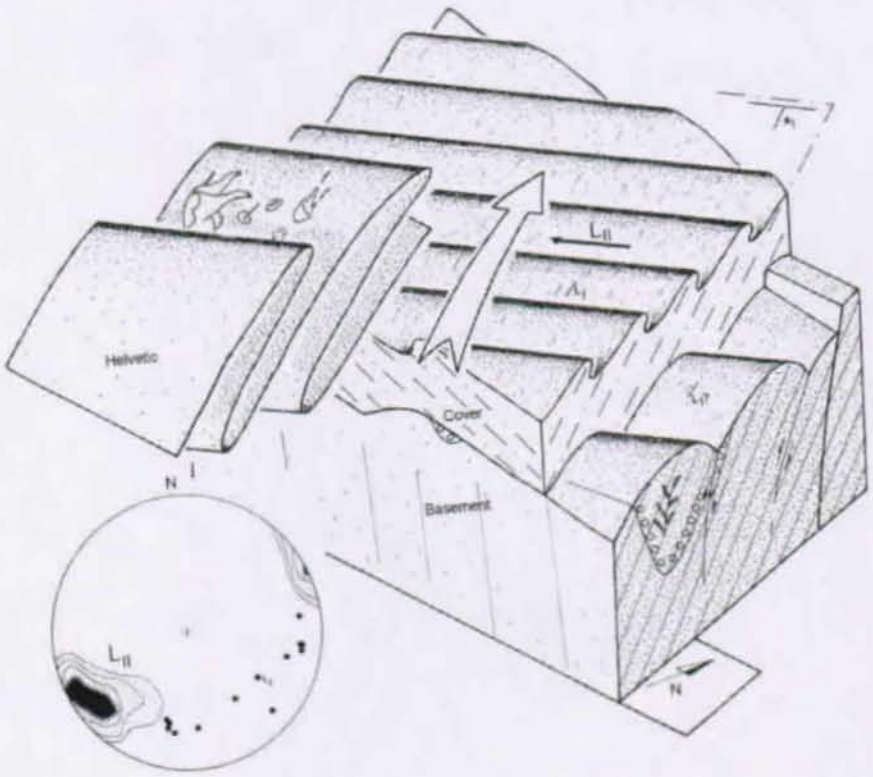


Figure 5-8 A tectonic scheme of the Southwestern Aare massive. The lineation L_{II} is parallel to the fold axes and thus gives the direction of these axes. The measured directions L_{II} are reported in the stereogram at the bottom (after Dolivo 1982).

In this area, subsurface structures are strongly stretched along the direction of mineral lineation (see Figure 5-8). It is thus obvious that these structures are anisotropic and the main directions of this anisotropy are parallel and perpendicular to the fold axes.

5.2.4. The 3-D Geological Model of the Zone Near the Southern Portal

The 3-D subsurface model was built based on five cross sections. These cross sections are mainly inspired by those presented in Figure 5-9 and are oriented perpendicularly to the direction of fold axes and regularly arranged along this direction.

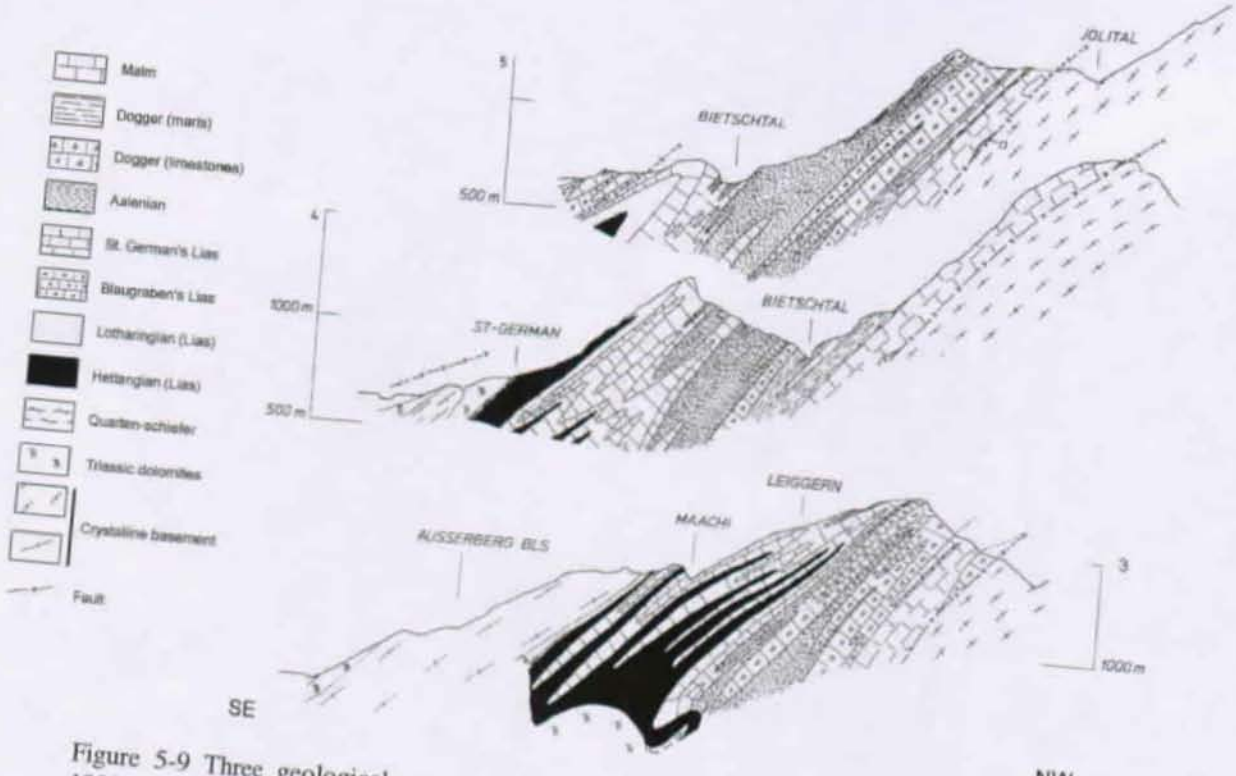


Figure 5-9 Three geological cross sections in the southwestern Aare massive (Dolivo 1982).

These cross sections are not detailed enough in the zone of interest. Additional information coming from four boreholes made near St.-German was thus taken into account in the subsurface model. These boreholes provided detailed information about the presence of Liassic and Triassic rock lenses near the southern portal of the Lötschberg tunnel (Kellerhals and Isler 1998).

The 1:10'000 geological map of the southwestern part of the Aare massive (Dolivo unpublished) also provides some observations about the location of some geological interfaces. However, these observations were scattered, because the interfaces were in large part hidden by the Quaternary cover.

The fact that folds are likely cylindrical in this area (see Figure 5-8), allows us to easily interpolate the structures from one cross section to another along the direction of fold axis. The interfaces of the 3-D geological model were created by interpolating the digitized cross sections and all the available outcrop observations with a 2-D minimum curvature algorithm.

The model is composed of only seven different rock types, but despite this stratigraphic simplification, subsurface structures are complex in the study area and not less than 15 interfaces were thus necessary to build the geological model. Some of these additional interfaces were necessary to build several overturned folds according to the method we have mentioned in the second chapter. The others were necessary to build several Triassic and

Liassic rock lenses. Once all the required interfaces were created, they were arranged into a stratigraphic sequence according to the stratigraphic relations of these rock masses (Figure 5-10).

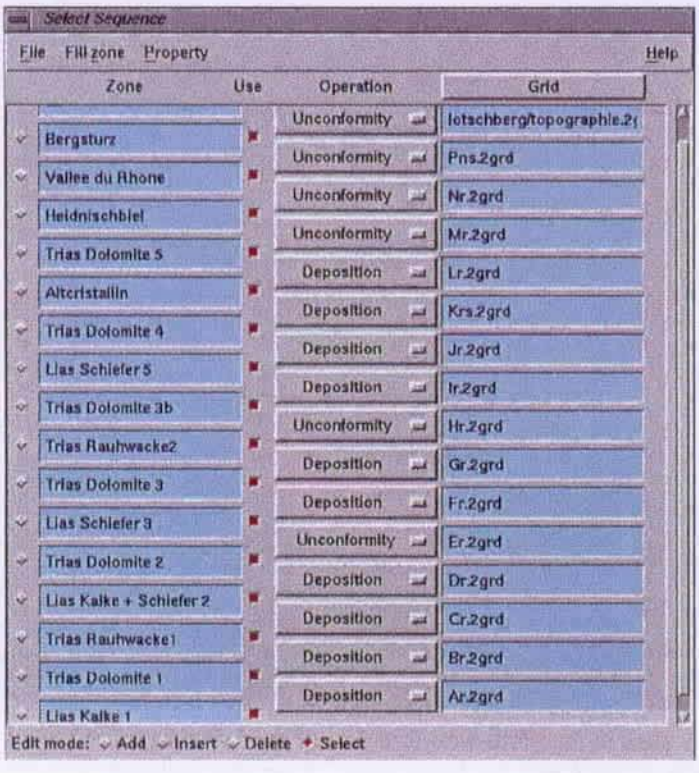


Figure 5-10 The stratigraphic sequence of the subsurface model as it is shown by the EarthVision® software. The surfaces (.2grd files) are the upper boundaries of the rock masses listed in the left column.

This sequence is closely related to the natural stratigraphic sequence. Most of the rock masses are concordant and few unconform interfaces are in fact either due to faults or to erosion surfaces: the fifth interface (Er.2grd in Figure 5-10) can represent the weathered, faulted zone observed by Dolivo (unpublished); the eighth interface (Hr.2grd in Figure 5-10) is also an unconformity and can represent the fault surface of the Maachi-Störung. The other unconformities are due to the later Heidnischbiel displacement (Mr.2grd) and the erosion of the Rhone glacier (Nr.2grd).

As can be noticed in this sequence, the 15 interfaces are bounding 16 rock masses. As some of these rock masses are in fact of the same type, a particular colour scale has been set up so that the similar rock types are appearing with the same colour in the geological model.

The resulting 3-D subsurface model in the area of the Raron portal is presented in both of the two following figures:

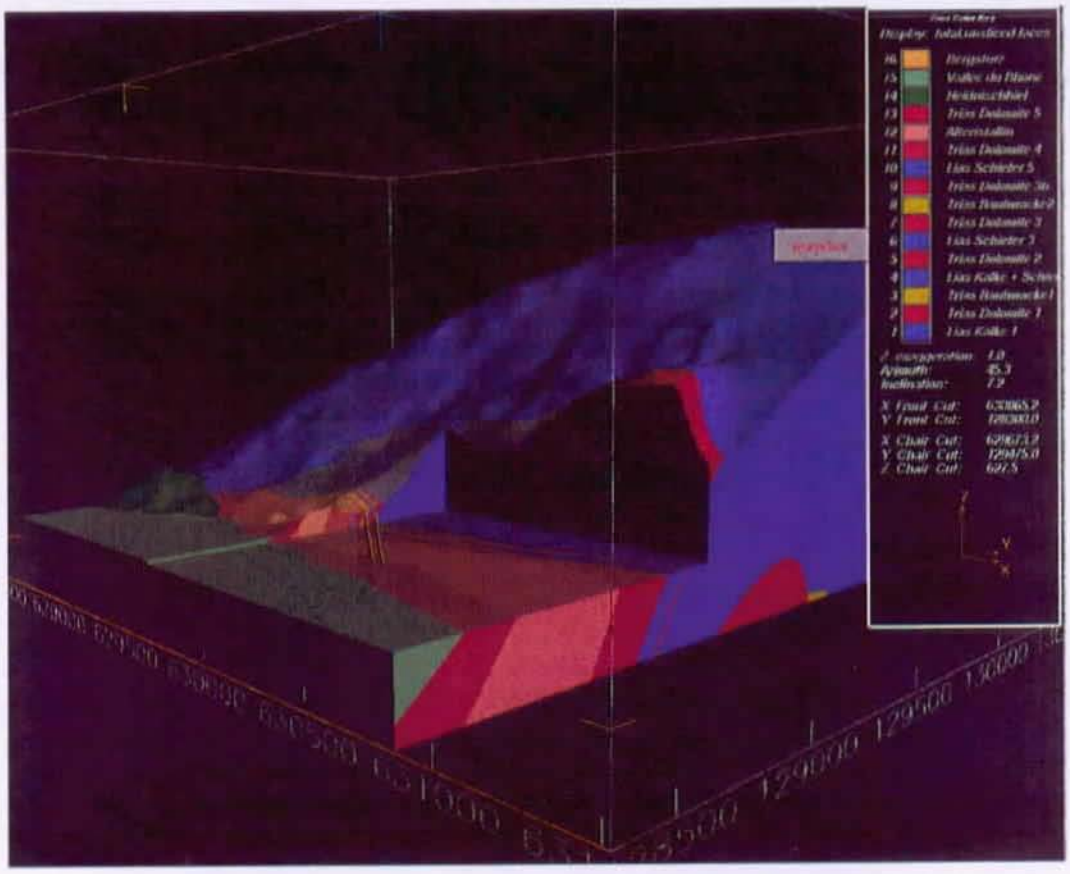


Figure 5-11 A view in the north-west direction of the 3-D subsurface model in the area of the Raron portal created with the EarthVision® modelling software. The two drives of the Lötschberg tunnel are represented with yellow tubes. The colours correspond to those of the simplified stratigraphy (Figure 5-7).

Figure 5-12 A closer view in the northwest direction of the 3-D subsurface model near St.-German. The drives of the Lötschberg tunnel are represented with yellow tubes. The St.-German landslide is shown with a brown, semi-transparent colour.

A more conventional - and easier to read – two-dimensional representation of subsurface structures along the tunnel drive is presented in the geological cross section of Figure 5-13. This cross section was extracted from the 3-D model.

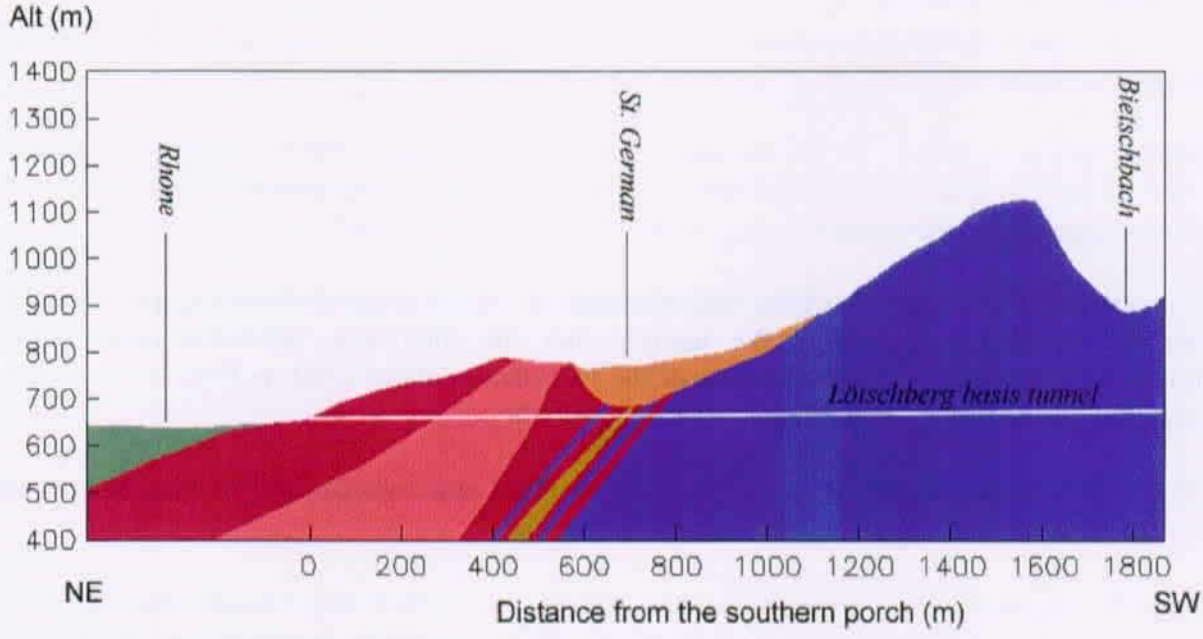


Figure 5-13 A geological cross section along the Lötschberg tunnel drive created with the EarthVision® modelling software. The colours are corresponding to those of the simplified stratigraphy (Figure 5-7).

The geological model represents the subsurface structures according to the small amount of information available and it is to a large extend interpretative. It represents the state of knowledge in the initial stage of the geological survey for the Lötschberg tunnel. Late available observations have not been taken into account when building this model.

Despite large uncertainties while building this model, it does not appear to be very different from the prognosis given by (Kellerhals and Isler 1998); when confronting the cross sections for example. This relative similarity may be due to the fact that the same information was used in both cases, but also - and mainly – to the fact that structures are strongly stretched in this area and thus quite invariant along the stretching direction (Dolivo 1982). In the Triassic and Liassic cover folds are nearly cylindrical.

There is, however, some evidences of small diversions in the position of some interfaces, which are due to the relatively unconstrained 2-D minimum curvature interpolation. These deviations could be corrected, but we kept this model as our best guess prognosis of subsurface structures.

5.2.5. Resulting Uncertainties

Resulting uncertainties related to the 3-D subsurface model of the area near the southern portal are described in this section. According to our methodology, the subsurface model presented in the previous section is considered as the best guess. We now have to propose a model of variability that can describe the fluctuations around the best guess.

Subsurface structures are anisotropic and a directional variogram is best suited in this case. This variogram should be oriented along the main tectonic directions, which are parallel and perpendicular to the fold axes.

In the study area, the size of the folds is on the kilometer scale (Dolivo 1982). So we assume that the minimum half-wavelength of folds is close to 500 m. Perpendicular to the fold axes, the variogram's range is thus set to $a_{\perp} = 500$ m.

As shown by the dispersion of the measurements on the stereogram of the Figure 5-8, Folds are not perfectly cylindrical. We assume that the minimum half-wavelength of the fluctuations around the mean direction of the fold axes is about 1500 m. Parallel to the fold axes, the variogram's range is thus set to $a_{//} = 1500$ m.

This large anisotropy ratio of $a_{//}/a_{\perp} = 3$ has been taken into consideration in order to account for the strong stretching of the rocks in the study area.

In both directions, the variogram's sill is set to c = 20'000 m^2, because the maximum variability around the expected position of the various interfaces is assumed to be of about $\pm\sqrt{20'000} \cong \pm 140$ m.

The model of variability is given by the following semi-variogram:

$$\gamma(\mathbf{h}) = \gamma(h_{//}, h_{\perp}) = 20'000 \cdot \left[1 - \exp\left(-\sqrt{\left(\frac{h_{//}}{1'500}\right)^2 + \left(\frac{h_{\perp}}{500}\right)^2}\right)\right]$$

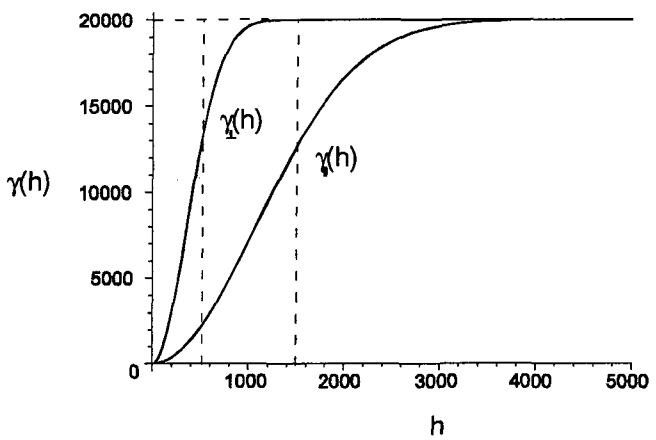


Figure 5-14 Resulting semi-variograms $\gamma(h_{//})$ and $\gamma(h_{\perp})$ parallel and perpendicular to the fold axis, respectively.

Observations available on outcrops (339 digitized points) and along boreholes (22 digitized points) are taken into account. Initial measurement errors are set to 0 for all these observations; all these observations are thus supposed to be hard data.

Here, the same stratigraphic sequence and stratigraphic relations as those used for the construction of the 3-D subsurface model (see Figure 5-10) have been considered.

These uncertainties are expressed in terms of probability of encountering the various rock types. They are calculated according to our methodology and are hence based on the subsurface model described in the preceding section and thus depend on available information as well as on various structural and stratigraphical constraints.

According to these given parameters, the probabilities could be computed with our CREER program in a sub-area of the 3-D subsurface model.

The resulting probabilities are shown as 4-D data (three coordinates and one probability). These results can be represented either as a 3-D probability field of a given rock type, or as a 3-D joint probability field of a given set of rocks, such as shown in Figure 5-15 where probabilities of the Liassic rocks are represented, and such as shown in Figure 5-16 where probabilities of the Triassic dolomites are represented.

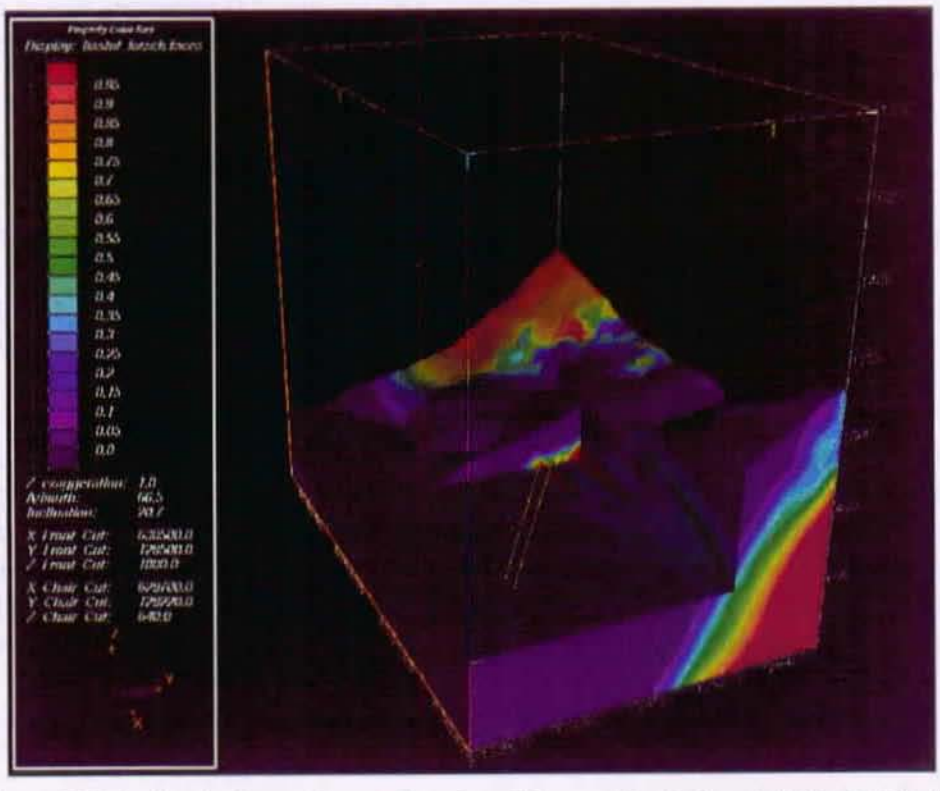


Figure 5-15 A view in the north-west direction of the resulting 3-D probability field in the area near St. German calculated with the CREER program and represented by the software EarthVision®. The colours are corresponding to the probability of encountering the Liassic rocks according to the colour scale on the left. The two yellow tubes show the two drives of the Lötschberg tunnel.

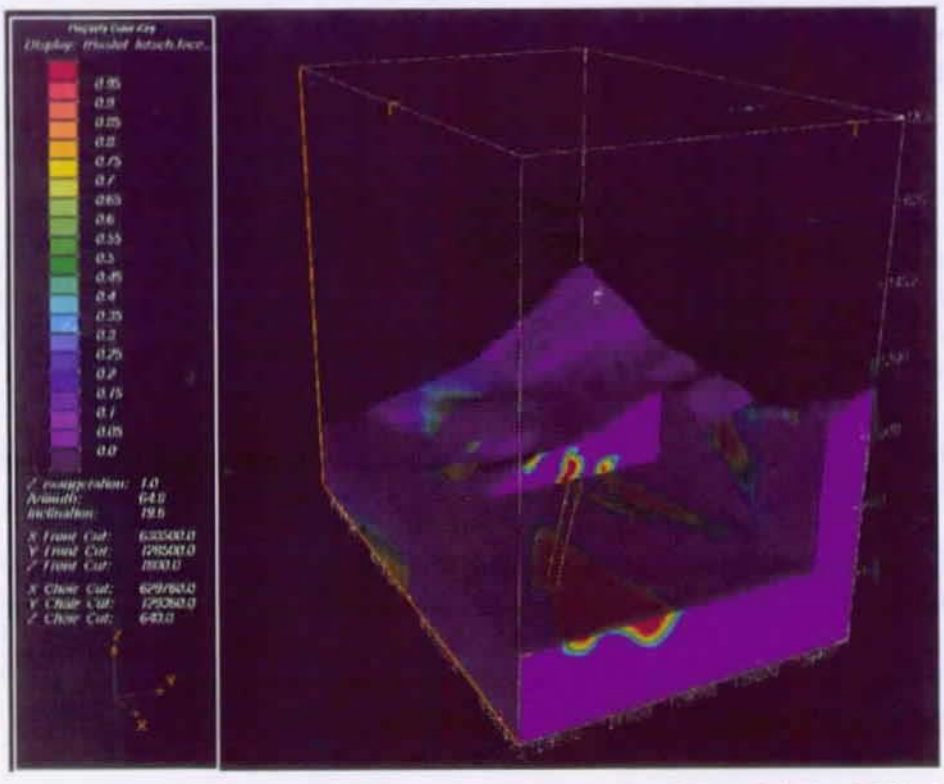


Figure 5-16 A view in the north-west direction of the resulting 3-D probability field in the area near St.-German calculated with the CREER program and represented by the software EarthVision®.

Such a representation provides the overall spatial distribution of possible occurrences of a certain rock type in the study area. In Figure 5-16 for example, Triassic rocks have high probabilities to occur in areas coloured with red and in contrary, there are few chances of finding Triassic rocks in areas coloured with purple. Areas with large uncertainties are those where colour changes are smooth.

This 3-D representation has the advantage that it can provide a quick overall view of resulting uncertainties in the whole study area, but the interpretation of such a representation is rather uncomfortable, because the results cannot be directly compared with the occurrence of others rock masses. For this reason, the more conventional 2-D representation is often more useful when studying a particular problem. For example, cross sections along the tunnel drive can show the estimated probabilities that are directly concerning the tunnel construction.

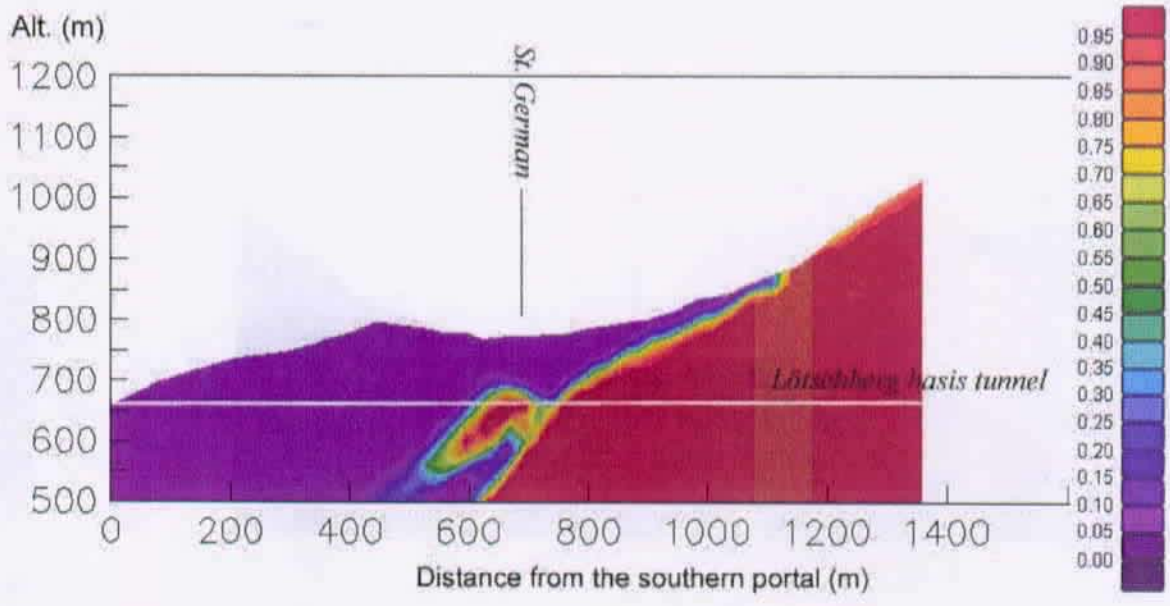


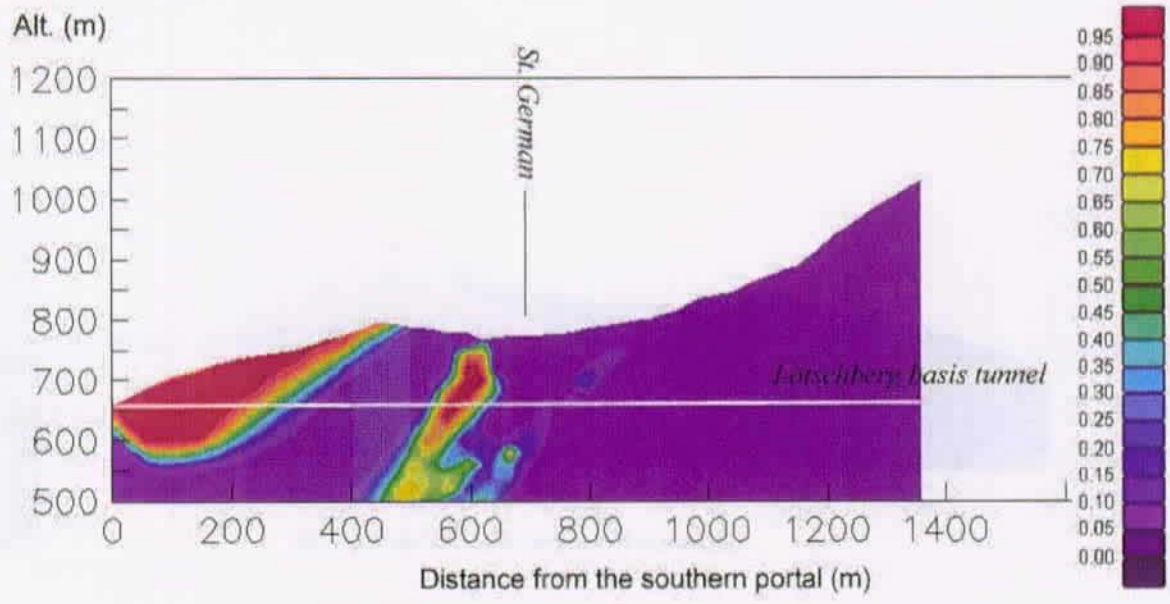
Figure 5-17 A cross section along the Lötschberg tunnel drive (white line) into the 3-D probability field, which represents the probability of encountering the Liassic rocks calculated by the CREER program.

In this last figure, we can see that the distribution of probabilities fairly represents the occurrence of Liassic rocks of the subsurface model of the area near the southern portal (see Figure 5-13). However, the Liassic lenses under the St.-German landslide are too thin to be accurately outlined by the probability field. This is either due to the low resolution of our study area or to a variability range that is larger than the size of these thin structures.

Probabilities can also be calculated in one dimension along the tunnel track itself. This can out an estimation of uncertainties at any location along the tunnel drive. Such result can then be directly integrated into some decision aid system in order to evaluate geological risks for the tunnel construction.

The next figure (Figure 5-18 i) shows a cross section extracted from the 3-D probability field related to the occurrence of Triassic rocks. This cross section can easily be compared to the corresponding probability profile calculated along the Lötschberg tunnel drive (Figure 5-18ii).

i)



ii)

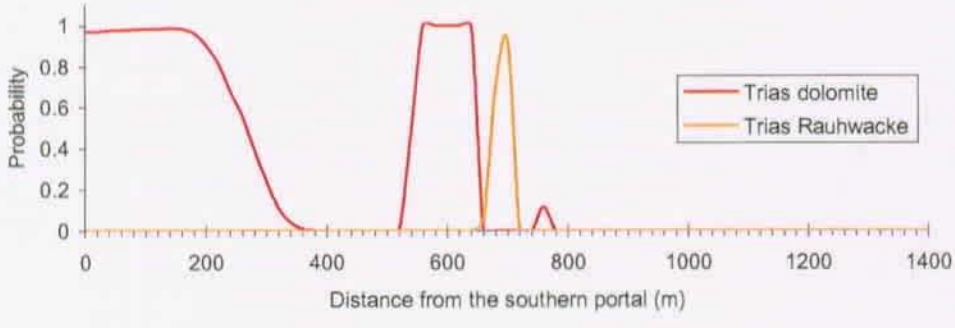


Figure 5-18 *i)* A cross section along the Lötschberg tunnel drive (white line) into the 3-D probability field, which represents the probabilities of encountering the Triassic rocks calculated by the CREER program.

ii) The corresponding probability profile along the Lötschberg tunnel drive. This profile shows the occurrence of the Triassic dolomite and of the Triassic Rauhwacke.

A one-dimensional profile can allow representing the whole probabilities along the tunnel drive on the same figure. This way, probabilities can be confronted in order to compare the relative probabilities of encountering the various rock masses. The next figure shows such a profile along the Lötschberg tunnel drive.

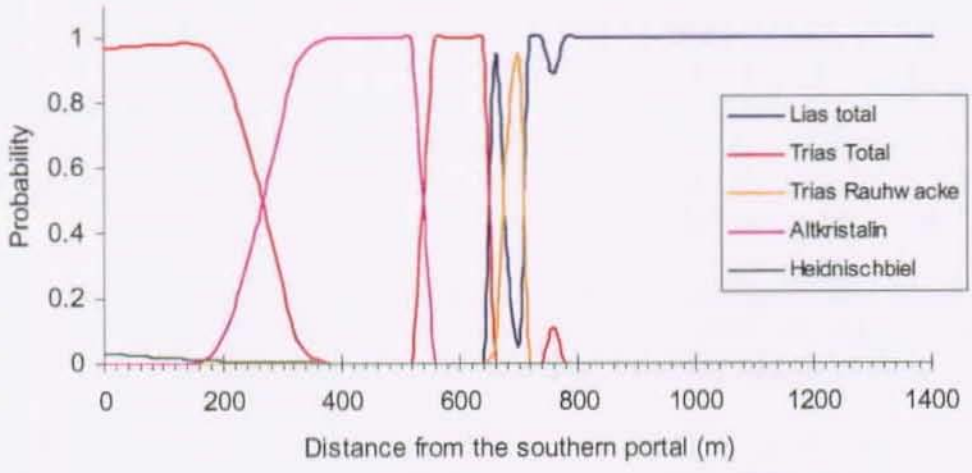


Figure 5-19 The probability profile along the Lötschberg tunnel drive calculated by the CREER program. All rocks of the subsurface model are accounted into this profile.

In this last profile all rocks are accounted for. It is thus interesting to compare the occurrence of the various rock masses. Uncertain predictions can be revealed when the probabilities are smoothly varying from one place to another and, on the other hand, when the probability profile is steep, little uncertainties are expected.

Summing all calculated probabilities can easily check results. In this case, the sum should be equal to 1 according to the second axiom of probabilities (see Chapter 4).

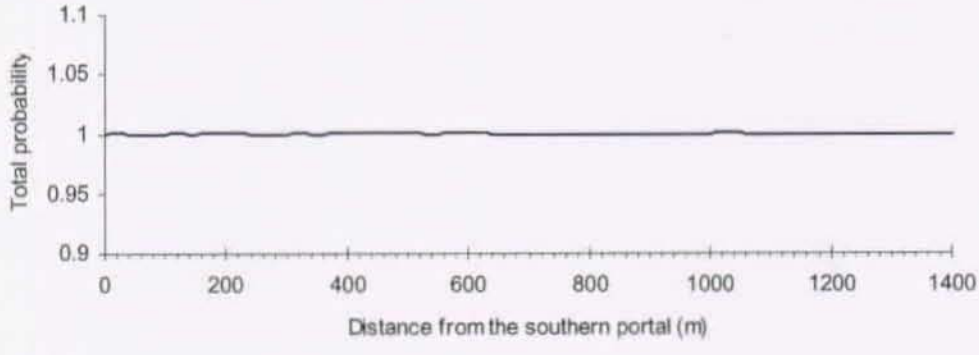


Figure 5-20 The sum of all probabilities calculated by the CREER program along the Lötschberg tunnel drive.

A more detailed description of the resulting uncertainties related to this subsurface model could of course be done, but the calculations have only been computed in order to test our methodology and a complete interpretation of these results is unfortunately beyond the scope of this research.

5.2.6. Remarks

At this point we should bring up some precisions about the calculation process with our CREER program. The first remark concerns the fact that the calculation process is time consuming. The time required for calculating a probability at one given location is about one second, so the bigger is the study area the longer will be the process. Due to the updating process, the duration is also proportional to the number of rock types and to the number of erosion surfaces that are present in the model.

We have added many enhancements in the calculation process in order to gain as much time as possible, but this was made to the detriment of the accuracy. We have therefore made the best compromise between duration and accuracy.

In the above case of the Lötschberg tunnel, the initial model size was too big to allow a calculation in a reasonable time. The size of this model is 3'660 x 2'360 x 1'600m and, in spite of a rather low resolution of 20m, there are already 1'723'120 nodes in the study area.

It was thus necessary to reduce the study area to a reasonable size. The range of the sub-area was set to 1'000 x 1'500 x 1'400m with the same 20m resolution. With this reduced grid of about 275'190 nodes and with 16 different rock types and 5 erosion surfaces, about two days were necessary to calculate the resulting probabilities. The program was executed by a personal computer using an 866MHz Pentium®III processor, using 64Mo RAM and running WindowsNT.

Due to calculation inaccuracies, the resulting probabilities can sometimes exceed the value of 1.0, but the maximum cumulated error measured when summing the whole probabilities was about 0.2, this means that a maximal error of 1.25% can be expected on the resulting probabilities. However, we must be aware that additional errors are induced by the interpolation process, particularly when using the 3-D minimum curvature interpolation method available in the EarthVision® software.

Chapter 6.
Summary and Conclusions

6.1. Results

This research presents a methodology for the estimation of uncertainties about subsurface structures in 3-D geological models. The overall goals in developing this methodology were to better define these uncertainties. The main challenge of this study was to account for various geological constraints and subjective guesses in order to bring a detailed evaluation of uncertainties in a formal probabilistic framework.

The main achievements of our work can be summarized as follows:

- Accounting for several tectonic constraints such as fold shape or constant thickness as well as stratigraphic constraints such as depositional or erosion relationship,
- estimating the probabilities of encountering any rock type in a full three-dimensional context in order to describe the occurrence of this rock,
- accounting for various types of hard and soft data in the estimation of structural uncertainties such as borehole or outcrop observations.

New methods to account for several geological constraints in the estimation of uncertainties were achieved in this research. These constraints, such as fold shapes and stratigraphic relations, can bring to a detailed description of uncertainties, which is closer to the natural phenomenon under investigation than the usual geostatistical methods.

Different types of information can be taken into account in the evaluation of uncertainties. Soft data with initial inaccuracy can be integrated together with hard data into the estimation of uncertainties. Prior guess is also accounted for, because uncertainties are evaluated according to the best-guessed geological model.

Starting from a 3-D model of subsurface structures, a probabilistic 3-D model of these structures is built following our methodology. At each location of the study area our methodology is able to evaluate the chances of finding a certain rock type. This evaluation is expressed in terms of probabilities, which are a direct quantification of uncertainties. As these probabilities account for several geological constraints, they thus provide an accurate description of uncertainties about subsurface structures.

The CREER program, which has been developed during this research, incorporates all the methodology described above. This program allows to estimate uncertainties according to several structural and geological constraints thanks to the calculation of the probabilities throughout the whole 3-D study area. Resulting probabilities are calculated either in the entire 3-D study area or just along a line such as a tunnel drive, for example.

When rendering the results as a three-dimensional probability field with a visualization software such as EarthVision®, a quick overview of the distribution of uncertainties is available and reliable areas can then be easily distinguished from those where large uncertainties are present.

When a particular sub-domain is targeted, it is also possible to focus on the evaluation of uncertainties in this area in order to get a more detailed description. This option can be particularly helpful when uncertainties must, for example, be assessed along a tunnel drive.

This methodology can successfully help engineers and geologists in the evaluation of global uncertainties over a whole 3-D study area as well as detailed uncertainties over a targeted area. It can be useful for engineers and geologists to identify areas of greater uncertainties, which can be targeted for further data collection, thus optimizing the site characterization. Our methodology can also serve to an accurate evaluation of geological risks that often affect the construction performance in tunneling. The resulting probabilities could easily be integrated into an existing decision aid system.

The results that have been obtained when applying our methodology on two different geological models (Chapter 5) show that this methodology is able to correctly describe uncertainties for many different subsurface structures.

6.2. Limits of the Methodology

6.2.1. Intrinsic Limits

Several assumptions are required in order to build the model of uncertainties.

The uncertainties are described by possible fluctuations around the expected position of the various interfaces in the best-guessed geological model. This estimation is thus based on the subjective model provided by the geologist. The resulting uncertainties are an expression of the spatial variability of the structures under study, but they do not describe uncertainties about the best-guessed model itself (see section 2.2). One should thus be aware that this methodology does not attempt to provide an estimation of global uncertainties which account for wrong choices or omissions the geologist could have made when building up his model.

The choice of limiting this methodology to the description of the spatial variability around a best-guessed model amounts to say that the best guess is an approximative but reliable description of the subsurface structures as long as the contrary cannot be proven. However, in areas where structures are known to be complex, it is right to think that the best guess can be far from the reality. Clearly, in faulted or strongly folded domains, the geological model can

be very uncertain. In such cases, a large sill value can be considered in the model of variability in order to account for large fluctuations. A nugget can also be added to the variogram model in order to account for an initial uncertainty around the best guess. At the extreme, a pure nugget variogram (white noise) can account for totally unknown structures.

Another limiting issue is due to an unfortunate consequence of the fact that a Gaussian distribution is assumed for the description the possible dispersions around the predicted position of an interface (see section 4.1). As the Gaussian distribution law is not bounded, the possible outcomes of such a distribution law are defined all over the interval from $-\infty$ to $+\infty$. This implies that an outcome can be found, not only near the expectation, but also very far from the prediction. Therefore, when representing the uncertainties with uncertainty bounds, nothing warrants that the reality will in fact fall within these boundaries.

The fact that the Gaussian distribution law is not bounded induces another annoyance. It is not possible to honour any threshold value within the range of possible outcomes. So, in our case, we cannot exclude the fact that layers can overlap and thus violate the stratigraphic rules. This issue is encountered in other cases. For example, when one studies the sizes in a human population, one assumes that its distribution is Gaussian. Although, many random phenomena are closely following a normal distribution law in practice, some inconsistent outcomes are not excluded from such a description, because this distribution can in fact not take a limit value into account; and a surprising size for a human such as –0.1m can thus not be excluded when considering a Gaussian distribution law.

Besides these limits, one must be aware that, just as there cannot be a prediction without a prior model, there cannot be an assessment of uncertainty of that prediction without a prior model. And again, models of uncertainty are not unique and cannot be "objective" (Journel 1996). Our model is like all models arbitrary in essence, and its ability to adequately represent uncertainties is of course always debatable.

6.2.2. Technical Limits

The technical limits are less serious than the intrinsic ones, because they are not inherent to the approach we have chosen. They can therefore be suppressed by enhancing the methodology we have described here.

One of the major technical limit of our methodology is due to the fact that the estimation of uncertainties is computed on a bi-dimensional surface. Our method only considers information that concerns the interface. Locations over and under this interface are not "visible". For example, even if the tail of a borehole is located near an interface, the observations made along this borehole are not taken into consideration as long as it does not reach the interface.

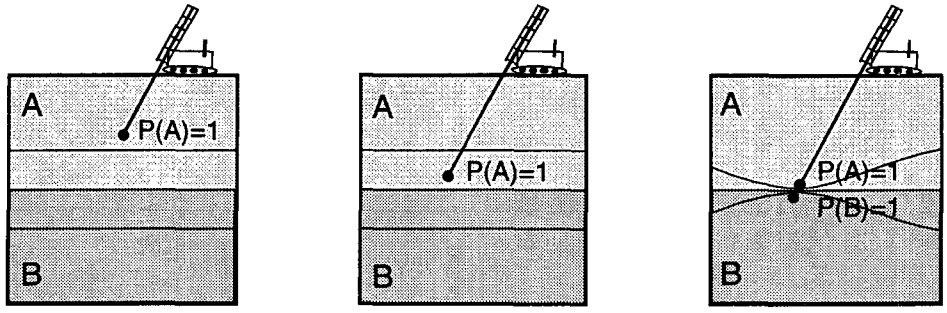


Figure 6-1 Three stages of the progression of a borehole. On the left picture, the borehole is still distant from the interface and its uncertainty bounds. On the middle figure, the borehole has reached the area contained between the uncertainty bounds and observations are there in contradiction with the uncertainty model. Finally, when the borehole has reached the interface (right figure), uncertainty bounds are tightened by the presence of an observation on the interface and no contradiction is present.

The first problem is that a distribution around one interface is independent of the distributions around the other interfaces. This independence allows outcomes for a given interface that can be inconsistent with the information concerning other interfaces. For example, if we except the constrain of isopach layers (see Chapter 3), when we have observed the true position of a given interface we can be sure about the presence or the absence of the various rock types in the neighborhood of this observation (Figure 6-2).

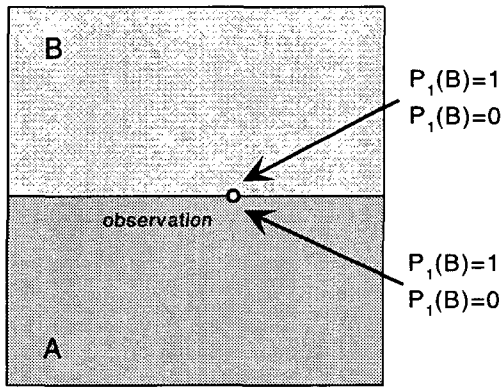


Figure 6-2 Around an observation of the real position of the bounding surface between *A* and *B* probabilities of occurrence are close to 0 or 1.

When a new interface is added to the model by the application of the iterative process previously described (see Chapter 4), initial probabilities are updated according to the presence of this additional interface. At this point, the probabilities of occurrence near the observation are modified (Figure 6-3).

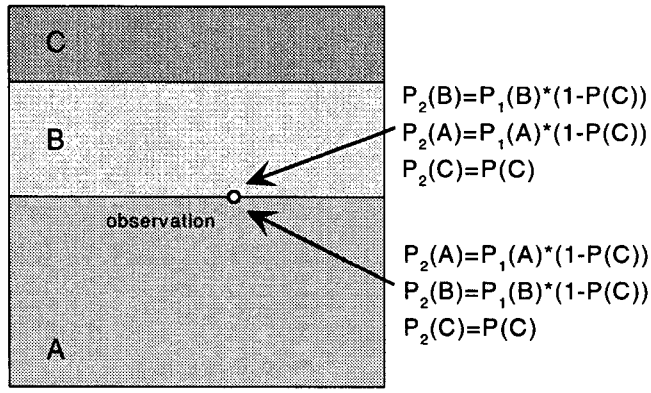


Figure 6-3 When adding a new interface that is an erosion surface, initial probabilities $P_1(A)$ and $P_1(B)$ are modified so that the new layer C can take place in the model.

As we can see in the last figure, the resulting probabilities are diminished by a ratio of [1-$P(C)$] because of the possible presence of the new layer C. Observations are in a certain sense altered by the updating process.

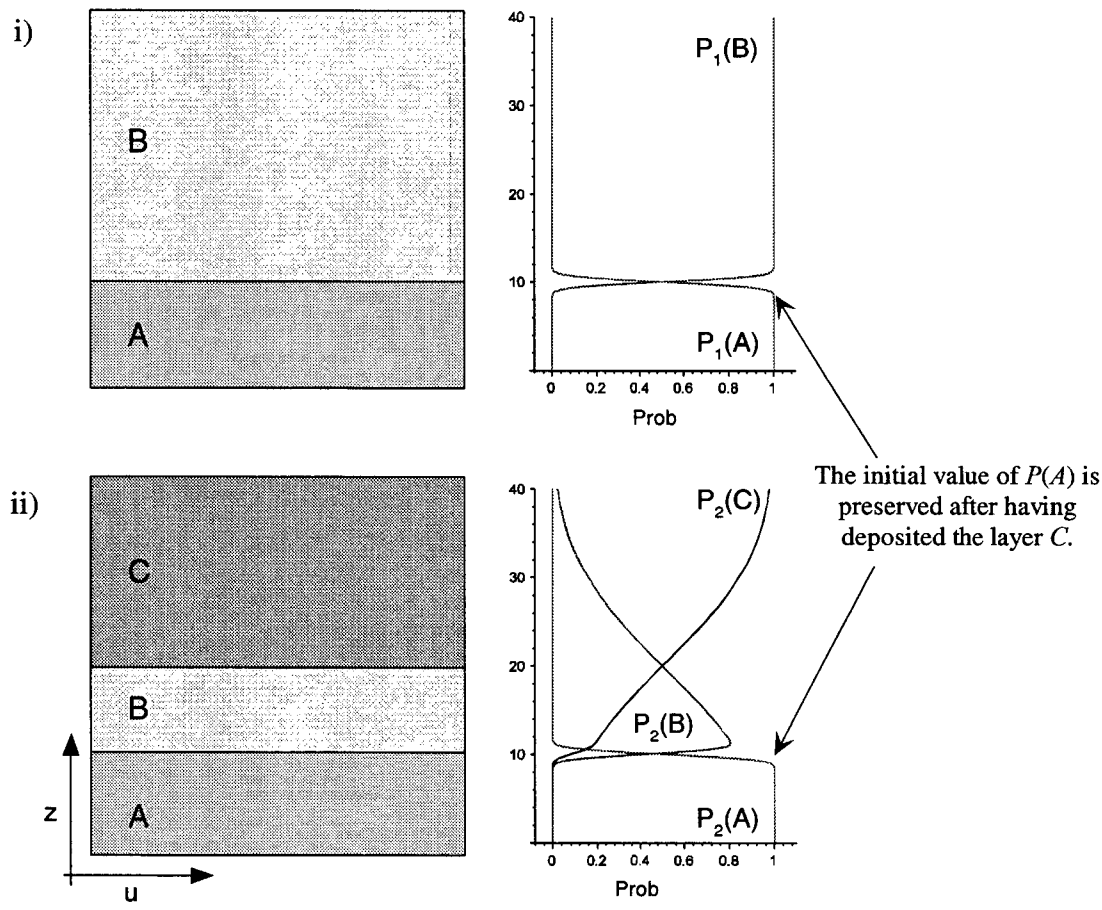


Figure 6-4 A stratigraphic sequence with three deposited layers. In figure i) the first iteration is represented with its initial corresponding probabilities $P_1(A)$ and $P_1(B)$ on the right. In figure ii) the second iteration with its resulting probabilities $P_2(A)$, $P_2(B)$ and $P_2(C)$.

This issue can be observed in Figure 5-4, where the probabilities around observations are diminished after the erosion of the sequence. This is particularly visible around the observation at x = 2.1, where $P(B)$ is lower than it should obviously be. However, this problem is only encountered when adding an erosion surface to the model. In other cases, overlapping is avoided. As it can be seen in Figure 6-4, The layer C cannot be found under the layer B, thus $P(C \cap A) = 0$.

To solve this issue, we must prevent the overlapping of the various random fields related to the interfaces of the model. This can be done by adding an additional constraint to the calculation of probabilities presented in section 4.2.

Such an alteration is due to the fact that the distribution of possible positions of the various interfaces is assumed to be Gaussian (section 3.2) and such a distribution allows possible outcomes for Z ranging between $-\infty$ and $+\infty$. As two interfaces are two independent random fields, there is no statistical or probability rule that can make them interact. All the above issues are caused by the fact that these random fields are independent one to the others and the overlapping of two interfaces in the sequence cannot be excluded.

Another limit that can affect the results is related to the measurement of separation distances. With the CREER program, the separation distances are measured along a same surface only, so when a geological body is delimited by several interfaces, as it is the case with complex structures, the total separation distance is not accessible and only the segmented distances are measured (see Figure 6-5).

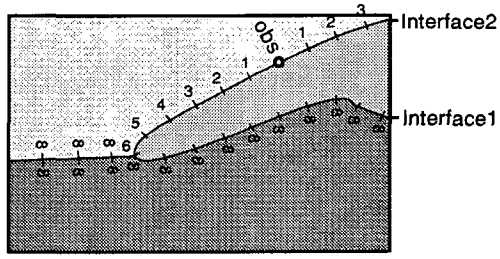


Figure 6-5 Two interfaces compose an overturned fold. An observation is available on interface2. The distances to the observation are measured along interface2, but on the interface1 no observation is available. Separation distances are thus infinite along this interface.

In this case, the measured distances are independent from one interface to another and large discontinuities in measured distances can occur at the junction of the various interfaces. This issue can cause a wrong estimation of uncertainties when several interfaces are in fact segments of the same boundary.

The process of measuring distances should therefore be implemented in order to allow a connection between various interfaces during the measurement process.

Such an implementation could also be interesting in the case of faulted domains. In such a case, the various fault-blocks could be related so that uncertainties estimated in one block influence the estimation on the other blocks.

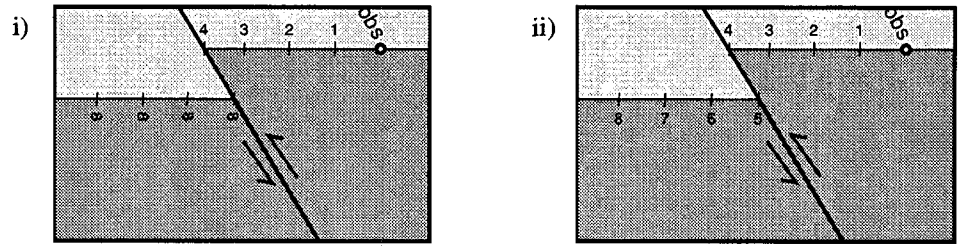


Figure 6-6 i) The two fault-blocks are not related and the observation available in the right block is thus not taken into account in the left block. ii) If the two blocks are related, the separation distances measured in the left block are the continuation of those measured in the right block.

6.3. Recommendations for Future Work

Several aspects of this research could be further developed.

We have currently tested our CREER program on many simple cases (one of them is presented in section 5.1), but few tests have so far been computed with real 3-D study cases. The overall goal was to ensure the potentiality of the application of the methodology and this was done on one large and complex area, which is the Lötschberg case. Additional tests could thus be computed in order to check the reliability of this methodology when handling various kinds of complex three-dimensional structures.

Since we have made many assumptions for the building up of an effective stochastic model, additional researches should thus be engaged in order to confirm - or possibly infirm - the description we have brought. Some further investigations are needed to bring a reliable assessment of the variogram function. Further experiments are also required in order to bring out the most representative statistical description possible. A parametrical study would be particularly interesting for ascertain the various parameters of the model of uncertainties, but it was beyond the scope of this research. Such a study could thus be done in a future work.

Some limits of our methodology, which were enumerated in the previous section, should be further investigated. The fact that random fields are not constrained in order to avoid any overlapping in the stratigraphic sequence, affect the accuracy of the methodology. In this case, additional constraints should be brought into our calculation process. This constraint can be expressed as a conditioning rule in the calculation of probabilities (i.e. the probability of encountering a certain rock type at a given location, knowing that it cannot be found below or over a certain point). The fact that faulted blocks are not related one to the other Figure 6-6, is less problematic. This limit is not due to the methodology, but only require a further development of the calculation program.

Additional deterministic information can be taken into account in a further uncertainty model. Information like dip measurements is an invaluable information and the geological prognosis

is often improved if such measurements are available. Uncertainty about subsurface structures are therefore certainly sensitive to such information. Dip measurement can be integrated in the evaluation of uncertainty by introducing an additional constraints in the correlation function.

Appendix A.
Common Notation

a: amplitude

A: random amplitude

$\overline{A}$: mean amplitude

c: sill parameter

$C(\mathbf{h})$: stationary covariance between any two random variables $Z(\mathbf{u})$, $Z(\mathbf{u}+\mathbf{h})$ separated by vector $\mathbf{h}$

$Cov\left(Z(\mathbf{u}_i), Z(\mathbf{u}_j)\right)$: non-stationary covariance of two variables $Z(\mathbf{u}_i)$, $Z(\mathbf{u}_j)$

$C(0)$: covariance value at separation vector $\mathbf{h} = 0$. It is also the stationary variance of random function $Z(\mathbf{u})$

$F(\mathbf{u};z)$: non-stationary cumulative distribution function of random function $Z(\mathbf{u})$

$F(z)$: cumulative distribution function of a random variable Z

$E\{Z(\mathbf{u})\}$: expectation of the random function $Z(\mathbf{u})$

$\gamma(\mathbf{h})$: semi-variogram between any two random variables $Z(\mathbf{u})$, $Z(\mathbf{u}+\mathbf{h})$ separated by lag vector $\mathbf{h}$

$2\gamma(\mathbf{h})$: variogram

$2\hat{\gamma}(\mathbf{h})$: experimental variogram

$\gamma^*(\mathbf{h})$: empirical semi-variogram

$\mathbf{h}$: separation vector (vectors are assigned with bold letters)

$\|\mathbf{h}\| = < \mathbf{u}_i; \mathbf{u}_j >$: Euclidian distance that separates two points $\mathbf{u}_i$ and $\mathbf{u}_j$

$i(\mathbf{u})$: binary indicator value at location $\mathbf{u}$

λ_α: kriging weights associated to datum α

$m(\mathbf{u})$: mean at location $\mathbf{u}$, expected value of random variable $Z(\mathbf{u})$

$\hat{m}(\mathbf{u}) = \frac{1}{N}\sum_{i=1}^{N} z(\mathbf{u}_i)$: estimated mean of $Z(\mathbf{u})$ at location $\mathbf{u}$

$N(1,0)$: normal distribution law

n: number of samples

N: number of modes or iterations

$n(\mathbf{h})$: number of pairs of data available at a lag vector $\mathbf{h}$

OK : acronym for ordinary kriging

ω: wavelength

Ω: random wavelength

$\text{Prob}\{Z(\mathbf{u})\in A|(n)\}$: probability that $z(\mathbf{u})$ belongs to rock type A at the location $\mathbf{u}$ conditioned by the n available observations $z_1,..,z_n$

$P(Z\in A) = P(A)$: probability of finding rock type A (simplified notation)

$P_i(L_j)$: probability of finding the j^{th} layer of a stratigraphic sequence at the i^{th} iteration

r: range parameter

$\rho(\mathbf{h}) = C(\mathbf{h})/\sigma_{h+}\cdot\sigma_{h-}$: stationary correlogram (or autocorrelation) function $\in$ [-1,+1]. It is the normalized covariance

s: nugget parameter

SK: acronym for simple kriging

$\sigma_Z^2(\mathbf{u})$: variance of $Z(\mathbf{u})$ at location $\mathbf{u}$

σ_{OK}^2: ordinary kriging variance of $Z(\mathbf{u})$

$\hat{\sigma}_Z^2 = \hat{\sigma}^2(\mathbf{u}) = E\left\{\left[Z(\mathbf{u})-\hat{m}(\mathbf{u})\right]^2\right\}$: estimated variance of $Z(\mathbf{u})$ at location $\mathbf{u}$

$\mathbf{u}$: generic coordinate vector (vectors are assigned with bold letters)

$\mathbf{x} = (x,y,z)$: usual Cartesian coordinate vector

$\{z(\mathbf{x}_1), z(\mathbf{x}_2), .., z(\mathbf{x}_n)\}$: sample set of n observations

$z_i(\mathbf{u})$: z-datum at location $\mathbf{u}$

$Z(\mathbf{u})$: generic random function of location $\mathbf{u}$ (random variables are assigned with capital letters)

$z(\mathbf{u})$: study variable. It is function of location $\mathbf{u}$

$Z^*(\mathbf{u})$: estimator of $Z(\mathbf{u})$

$Var\{Z(\mathbf{u})\}$: variance of $Z(\mathbf{u})$ at location $\mathbf{u}$

Appendix B.
CREER: Uncertainties Calculation Program

B.1. Introduction

CREER is a program developed to evaluate the uncertainties related to 3-D geological models. The source code of the CREER program was written in Fortran90, because this language is well-suited to an intensive calculation process. This code was compiled with the Compaq Fortran platform (formerly DIGITAL Fortran) on an IBM compatible personal computer. However, as this code was written in conformity with the Fortran language standards, it can actually be compiled by any other compiler. No user interface has been created for this program, first because it is mainly a calculation program and second, because it would have taken too much time to the detriment of the development of the methodology.

The CREER program is mainly designed to estimate geological uncertainties related to 3-D geological models. It calculates the probabilities of encountering the various rock types in a geological model. This program was originally designed to aid geologists and engineers through the risk evaluation process and the project design for tunneling, but it could also be applied to many other fields like mining exploration, oil exploration or groundwater modelling.

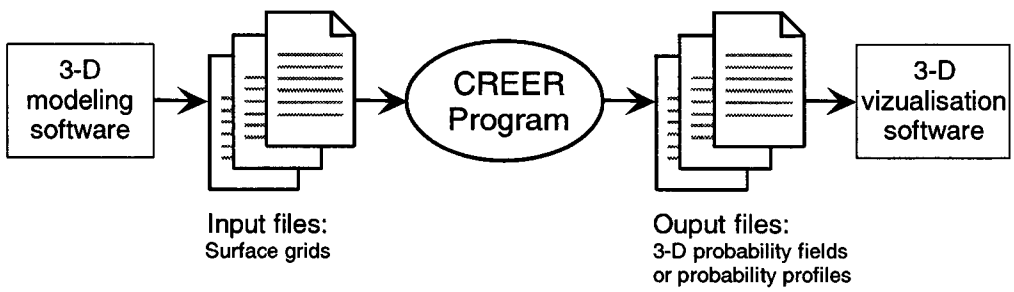


Figure 6-7 A simple scheme illustrating the use of the CREER program. Data must first be prepared by a modeller program in order to contain regularly spaced data points. The CREER program calculates the various probabilities and writes the results in output files. These files can then be read and plotted by a 3-D visualization program.

Besides the calculation of uncertainties, the program also allows to:

- manipulate gridded data files of the various interfaces and to add or suppress nodes according to their stratigraphic relations,

- build a 3-D grid of properties $p(x,y,z)$ representing the various rock types according to the stratigraphic relation of each unit,

- calculate path distances on any gridded surface,

- calculate local thickness between two gridded surfaces,

- calculate a probabilistic model according to tectonic and stratigraphic criteria, which describes the occurrence for the various rock types along a given track or in the entire 3-D domain,

- calculate the variogram surface with an enhanced algorithm for 2-D regularly spaced data.

The program was designed to handle simple raw data files as easily as possible. Main input files are regularly spaced data in a grid files coming from the 3-D geological modelling software EarthVision™. Fortunately, these files have a simple format, so, according to the required format, it is still possible to avoid the use of the EarthVision™ software and to build one's own data files (see section B.2.1 for details on this file format). However, the program is able neither to read any kind of irregularly distributed raw data files, nor to represent the results graphically, so it cannot be used alone, but should rather be used as a modulus working together with other softwares.

Much of the programming in CREER is original work. Some of the algorithms used in this program, as the Dijkstra's algorithm (Kreyszig 1993) for the minimal path calculation, are inspired from literature.

The program was intensively tested and used on several real cases, but there will undoubtedly remain some "bugs".

B.2. User's Guide

B.2.1. Input Files

There are three kinds of input files:
- data files
- observations files
- track files
- parameter file

The **data files** contain the data points. The first data files needed by the program are the gridded surfaces. They consist of ASCII space, coma or tab delimited files that contain the

coordinates of the grid nodes belonging to a given interface. Each node constitutes a row in the data file. For each node, there are three columns containing the x, y and z coordinates, plus two columns containing the column and row numbers of the grid node.

```
# Example of a input data file
# Header lines are preceded with the # character
# There are five columns: x, y, z, #col, #row
0      0      -1E+30      1      1
100    0      59.8223     2      1
200    0      31.9671     3      1
300    0      4.85748     4      1
0      100    -1E+30      1      2
100    100    59.8223     2      2
200    100    31.9671     3      2
300    100    4.85748     4      2
```

Figure 6-8 Example of an input data file. It is a small 4x2 grid containing eight nodes. When the value of the z coordinate equals -1E+30, it means a missing value for this node.

There are two important constraints about the input data files:

- Each input file must be a grid with regularly spaced nodes. If such is not the case, the program will not be able to run correctly and will display an error message (error 003) when reading the file.

- All data files must also have the same grid size. This means that they must have the same number of columns and rows. If there are not enough data to fulfill this condition, the grid can be completed by adding the nodes with the z value equal to -1E30 which corresponds to a missing value; the program will ignore these data.

The **observation files** are ASCII space, tab or coma delimited files containing raw data of observed points. The required format was established so as to make the use of information easier. These files are made up of one or more header lines followed by the lines containing information about the observed points. The observed points are made up of three columns containing the coordinates of the point, one column giving the value of the measurement error, and a fifth column containing the name of the data file of the concerned interface.

```
# Example of an observed data file
# Five columns: x, y, z, error, filename
12     3      12.100      0.0    interface1.dat
120    0.5    59.8223     0.2    interface2.dat
```

Figure 6-9 Example of an observed data file. The fifth column makes reference to the filename of the interface to which the observed point belongs.

For this type of file, there is no need to have regularly spaced data, but the main constraint is bound to the name of the file containing the concerned interface. If this file does not exist the program will display an error message (error 077) and ask the user to enter the name of the data file manually.

The **track file** is only needed when calculating the probabilities along a track. The track can be made of either a set of points located on the track, or the first and the last point of the track. Each line is simply composed of the x, y and z coordinates.

```
# Example of a track file
# first line: start
# second line: end
2.0 0.1 -0.5
7.0 0.1 1.0
```

Figure 6-10 Example of a track file.

The **parameter file** is necessary to run the program properly and must be named "vario.var". This file contains the parameters *a* (range) and *c* (sill) for the various variograms, as shown in following figure (Figure 6-11). These parameters can (and should) be customized by the user in order to use proper variogram parameters.

```
# The vario.var file
1 0.5 default
1 0.5 default
1 0.5 interface1.dat
1 0.3 interface1.dat
1 0.9 interface2.dat
1.5 0.5 interface2.dat
```

Figure 6-11 The "vario.var" file. This file contains the parameters of the various variograms.

Each line concerns a particular variogram. The first column gives the variogram sill, the second column indicates the variogram range and the last column shows the filename of the interface to which the variogram is applied. Note that there always are two lines of parameters for each interface, in order to allow directional variograms in the main tectonic directions (see Chapter 3).

As for observation files, the third column contains a filename, which refers to the file that should be concerned by the parameters. So, on each interface of the model, one can set up a custom variogram. If a row contains a filename that does not exist, the parameters given in this row will not be taken into consideration.

In this file, the name "`default`" is reserved. It indicates which parameters the program should use if some parameters are missing. So one should be aware that, when some parameters are missing, the default parameters would be used.

If the "vario.var" file does not exist, the program will display an error message (error 041) and the user will be asked to enter the parameters of the default variogram manually.

B.2.2. Output Files

Output files are files that contain the results of the calculation process. These results are simply written in ASCII space separated files and the program does not allow any graphical display. The results should then be visualized in 3-D modelling software like EarthVision™ or 2-D mapping software, like Surfer™ for example (most of the graphics of this work were made with this software).

Main results of this program is a file that contains calculated probabilities for a given rock type.

B.2.3. Output Filenames

Due to the fact that our calculation is processed through a series of stages, we chose to set up a particular file-naming rule. This rule may seem confusing, but it allows the identification of each resulting file at any stage of the process.

At the beginning, the user will be asked to enter a generic filename (like a project name). If needed, the file extension must be added. This name will be used as a suffix for all resulting output files.

`"name.dat"` example of a filename as it can be entered by the user.

The program will fit a first prefix to the name in order to identify the type of resulting file. This prefix will be `"track_"` when probabilities are calculated along a track or `"prob_"` when probabilities are calculated for the whole 3-D area.

`"prob_name.dat"` example of a filename with its prefix.

A second prefix is added to identify the number of the formation. If there are N interfaces in the geological model, there will be N+1 formations. This number corresponds to the fifth column of the output file (see Figure 6-12).

`"iprob_name.dat"` example of filename with its second prefix (i = 1,..,N+1).

Finally, a last prefix is added which makes reference to the updating process of probabilities. If there are n erosion surfaces in the geological model, there will be n successive updates. So original files have the prefix "0" (zero) while each new update files numbered from 1 to n

"j-iprob_name.dat" example of filename with its third prefix (j = 1,..,*n*).

When calculations are done, original files and files resulting from previous updates can be deleted and it is now possible to keep only the last n^{th} resulting updated files.

```
x y z Prob #formation co li dist
1.600    0.100    -0.900   0.11644001
1.600    0.100    -0.800   0.13909250
1.600    0.100    -0.700   0.16905718
1.600    0.100    -0.600   0.19893467
1.600    0.100    -0.500   0.22522965
1.600    0.100    -0.400   0.24807036
1.600    0.100    -0.300   0.27369380
...
```

Figure 6-12 Example of the "12prob_sample.dat" output file. This file concerns the formation #2 as it is written in the fifth column. Each line represents a point located in the 3-D domain. These points are located by the three columns x, y and z. The fourth column contains the calculated probabilities of encountering the formation #2.

B.2.4. Parameters Required by the CREER Program

The program was conceived so as to reduce as much as possible the annoying procedure of entering parameters. However, several parameters are required by the program in order to be able to calculate the probabilities.

- First, the number *N* of input files must be entered. This number corresponds to the number of interfaces that the program should account for.
- The *N* filenames will then be asked for starting from the bottom interface of the geological model.
- For each of these interfaces, the user must indicate whether it is a depositional or an erosion surface.
- For each of these interfaces again, the user must indicate if folds are concentric or parallel
- If needed, new boundaries for the study area can be entered in order to reduce its size.
- The program will search for the fold axis direction. However, the user still may enter some particular direction if needed.
- Finally, the user will be asked to enter the number and the name of the various input files that contain the observations.

At this point, the program is fed with all required parameters and the calculation process can thus begin. If no error is encountered during the calculation, the program will calculate all the probabilities and will write them into the various resulting files while covering the whole 3-D study area.

References

Abrahamsen, P. and H. Omre (1994). Random Functions and Geological Subsurfaces. 4th European Conference on the Mathematics of Oil Recovery, Røros, Norway.

Alabert, F. G. (1994). Reservoir Engineering and Uncertainty How much do we know about what we don't know? 4th European Conference on the Mathematics of Oil Recovery, Røros, Norway.

Blanchin, R. and J.-P. Chilès (1993). "The Channel Tunnel: Geostatistical Prediction of the Geological Conditions and its Validation by the Reality." Mathematical Geology **25**(7): 963-974.

Blanchin, R., P. Margron, et al. (1990). Tunnel sous la Manche - Les Calculs à l'Epreuve de la Réalité. Franchissements Souterrains pour l'Europe. L. (ed.). Rotterdam, A. A. Balkema: 139-143.

Budd, C. J. and M. A. Peletier (2000). "Approximate Self-Similarity in Models of Geological Folding." SIAM Journal on Applied Mathematics **60**(3): 990-1016.

Chilès, J.-P. (1999). Geostatistics - Modelling Spatial Uncertainty. New York, Wiley-Interscience.

Dagbert, M., M. David, et al. (1984). Computing Variograms in Folded Strata-controlled Deposits. Geostatistics for Natural Resources Characterisation. G. Verly, M. David, A. G. Journel and A. Marechal, D. Riedel Publishing. **122:** 71-89.

Davis, J. (1973). Statistics and Data Analysis in Geology. New York, John Wiley & Sons Inc.

de Sitter, L. U. (1957). "Boudins and Parasitic Folds in Relation to Cleavage and Folding." Geol. en Mijinb. **20**: 277-286.

Deutsch, C. and A. Journel (1998). GSLIB Geostatistical Software Library and User's Guide. New York, Oxford University Press Inc.: 369.

Dolivo, E., Ed. (1982). Nouvelles observations structurales au SW du massif de l'Aar entre Visp et Gampel. Matériaux pour la Carte Géologique de la Suisse. Basel, Commision Géologique Suisse.

Dolivo, E. (unpublished). Nouvelles observations structurales au SW du massif de l'Aar entre Visp et Gampel. Lausanne, unpublished.

Dudt, J.-P. and F. Descoeudres (1999). Quantifizierung der Prognose(un)sicherheit im Tunnelbau am Beispiel der AlpTransit Basistunnel. Symposiums Geologie AlpTransit, Zürich, A. A. Balkema.

Dynamic Graphics, I. (1997). Earth Vision 5.1 User manual.

Einstein, H. H., V. B. Halabe, et al. (1996). Geological Uncertainty in Tunneling. Uncertainty '96 (ASCE), Madison, Wisconsin, ASCE.

Einstein, H. H. and G. S. Baecher (1982). "Probabilistic and Statistical Methods in Engineering Geology. 1. Problem Statement and Introduction to Solution." Rock Mechanics **Supp. 12**: 47-61.

Einstein, H. H. and G. B. Baecher (1992). Tunneling and Mining Engineering. Techniques for Determining Probabilities of Geological Events and Process. R. L. H. a. C. J. Mann. New York, Oxford University Press: 46-74.

Haldorsen, H. and E. Damsleth (1990). "Stochastic Modelling." J. of Petroleum Technology **44**(4): 404-412.

Holden, L. and P. F. Mostad (1997). Handling of Uncertainty in Petroleum Applications. Oslo, Norvegian Computing Center.

Isaaks, E. and R. Srivastava (1989). An Introduction to Applied Geostatistics. New York, Oxford University Press Inc.

Isaaks, E. H. and M. R. Srivastava (1988). "Spatial Continuity Measures for Probabilistic and Deterministic Geostatistics." Mathematical Geology **20**(4): 313-341.

Jaquet, O. (1989). "Factorial Kriging Analysis Applied to Geological Data from Petroleum Exploration." Mathematical Geology **21**(7): 683-691.

Jones, T. A. (1990). "Why 3-D modelling?" Geobyte **5**(1): 25-26.

Journel, A. G. (1996). The Abuse of Principles in Model Building and the Quest for Objectivity. Geostatistics Wollongong'96, Wollongong, Australia, Kluwer Academic Publishers.

Journel, A. G. and J. J. Gomez-Hernandez (1993). "Stochastic Imaging of the Wilmington Clastic Sequence." SPE Formation Evaluation(March 1993): 33-40.

Journel, A. G. and C. J. Huijbregts (1978). Mining geostatistics. London a.o., Academic Press.

Kanevski, M., V. Demyanov, et al. (1997). Spatial Estimation and Simulations of Environmental Data by Using Geostatistics and Artificial Neural Networks. IAMG97 Annual conference.

Kellerhals, P. and A. Isler (1998). Lötschberg-Basistunnel: Geologische Voruntersuchungen und Prognose. Bern, Landeshydrologie und -geologie.

Kinnikutt, P. and H. Einstein (1996). Incorporating Uncertainty Objective and Subjective Data in Geologic Site Characterization. Uncertainty '96 (ASCE), Wisconsin, University of Wisconsin, Madison, Wisconsin.

Kreyszig, E. (1993). Advanced Engineering Mathematics, Wayne Anderson.

Krige, D. G. (1951). "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand." Journal of the Chemical, Metalurgical and Mining Society of South Africa **52**: 119-139.

Loew, S. (1997). "Wie Sicher Sind Geologische Prognosen?" Bulletin für Angewandte Geologie **2**(2): 83-97.

Luo, J. (1996). Transition Probability Approach to Statistical Analysis of Spatial Qualitative Variables in Geology. Geological Modelling and Mapping. A. F. a. D. F. Merriam. New York, Plenum Press: 281-299.

Matheron, G., Ed. (1963). Traité de Géostatistiques Appliquée, Tome II: Le Krigeage. Memoires du Bureau de Recherches Géologiques et Minières. Paris.

Mayoraz, R. (1993). Modélisation et Visualisation Infographiques Tridimensionnelles de Structures et Propriétés Géologiques. GEOLEP - Laboratoire de Géologie. Lausanne, Ecole Polytechnique Fédérale de Lausanne: 215.

McArthur, G. J. (1988). "Using Geology to Control Geostatistics in the Hellyer Deposit." International Association for Mathematical Geology **20**(4): 343-366.

Najjar, Y. M. and I. A. Basheer (1996). A Neural Network Approach for Site Characterization and Uncertainty Prediction. Uncertainty '96 (ASCE), Wisconsin, University of Wisconsin, Madison, Wisconsin.

Omre, H. (1987). "Bayesian Kriging - Merging Observations and Qualified Guesses in Kriging." Mathematical Geology **19**(1): 25-39.

Pilz, J., G. Spoeck, et al. (1996). Taking Account of Uncertainty in Spatial Covariance Estimation. Geostatistics Wollongong'96, Wollongong, Australia, Kluwer Academic Publishers.

Ramsay, J. G. (1967). Folding and Fracturing of Rocks. New York and London.

Ramsay, J. G. (1987). The Techniques of Modern Structural Geology. London, Academic Press Inc.

Ross, S. M. (1987). Initiation aux Probabilités. Lausanne, Presses polytechniques romandes.

Royer, J.-J., A. Shtuka, et al. (1997). Apport du 3D dans l'Estimation des Ressources du Sous-sol et des Gisements Miniers. Colloque Modélisation du Sous-sol, Paris.

Schenker, M. (1946). Geologische Untersuchung der Mesozoischen Sedimentkeile am Südrand des Aarmassivs zwischen Lonza und Baltschiedertal. Bern, Geologischen Kommission der Schweizerischen Naturforschenden Gesellschaft.

Tacher, L. and A. Parriaux (1997). Calcul et représentation de l'incertitude associée aux modèles géologiques. Colloque Modélisation du sous-sol, Ecole des Mines, Paris.

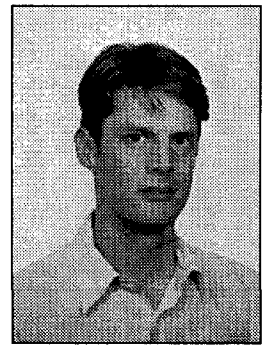
Vick, S. G. and H. H. Einstein (1974). A Probabilistic Approach to Geological Investigation for Hard-Rock Tunnels. Third International Congress of the International Society for Rock Mechanics, Denver.

Wingle, W. (1997). Evaluating Subsurface Uncertainty Using Modified Geostatistical Techniques. Department of Geology and Geological Engineering. Golden, CO, Colorado School of Mines: 180.

Ziegler, H. J. (1997). Lötschberg-Basistunnel: Neue geologische Erkenntnisse aus den Sondierarbeiten. AlpTransit: Das Bauprojekt-Schlüsselfragen und erste Erfarungen, Bern, SIA.

Ian POMIAN-SRZEDNICKI

Laboratoire de Géologie (GEOLEP)
Département de Génie Civil
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne



Né le 30 décembre 1968,
originaire de Genève (GE),
célibataire.

Formation

1998 à ce jour	Candidat au doctorat, Geolep - EPFL.
1998	Certificat de maîtrise de spécialisation en géologie appliquée à l'ingénierie et à l'environnement, Geolep - EPFL.
1997	Diplôme de Géologue de l'Université de Lausanne.
1995	Licence en Sciences de la Terre de l'Université de Lausanne.
1989	Maturité de type C (scientifique) cantonale et fédérale au collège de Candolle (GE).

Formation permanente

1998	Cours bloc de 3^{e} cycle de Cartographie géostatistique, réseaux neuronaux, SIG, Institut de Minéralogie de l'Université de Lausanne.
1993	Stage d'Hydrogéologie, Centre d'Hydrogéologie de Neuchâtel CHYN, Université de Neuchâtel
1992	Cours de "Conception d'interfaces utilisateurs C/C++", Faculté de HEC de l'UNI Lausanne.

Langues

Français	Langue maternelle.
Italien	Lu, parlé
Anglais	Lu, parlé, écrit
Allemand	Lu, parlé, écrit