SKB

TECHNICAL REPORT 92-08

Statistical inference and comparison of stochastic models for the hydraulic conductivity at the Finnsjön-site

Sven Norman

Starprog AB

April 1992

SVENSK KÄRNBRÄNSLEHANTERING AB
SWEDISH NUCLEAR FUEL AND WASTE MANAGEMENT CO

BOX 5864 S-102 48 STOCKHOLM

TEL 08-665 28 00 TELEX 13108 SKB S
TELEFAX 08-661 57 19

STATISTICAL INFERENCE AND COMPARISON OF STOCHASTIC MODELS FOR THE HYDRAULIC CONDUCTIVITY AT THE FINNSJÖN SITE

Sven Norman

Starprog AB

April 1992

STATISTICAL INFERENCE AND COMPARISON OF STOCHASTIC MODELS FOR THE HYDRAULIC CONDUCTIVITY AT THE FINNSJÖN SITE.

by Sven Norman, Starprog AB
March, 1992.

ABSTRACT

The origin of this study was to find a good, or even the best, stochastic model for the hydraulic conductivity field at the Finnsjö site. The conductivity fields in question are regularized, that is upscaled. The reason for performing regularization of measurement data is primarily the need for long correlation scales. This is needed in order to model reasonably large domains that can be used when describing regional groundwater flow accurately. A theory of regularization is discussed in this report.

In order to find the best model, jacknifing is employed to compare different stochastic models. The theory for this method is described. In the act of doing so we also take a look at linear predictor theory, so called kriging, and include a general discussion of stochastic functions and intrinsic random functions. The statistical inference methods for finding the models are also described, in particular regression, iterative generalized regression (IGLSE) and non-parametric variogram estimators. A large amount of results is presented for a regularization scale of 36 metre.

TABLE OF CONTENTS

1 BACKGROUND AND INTRODUCTION

The basic assumption of a continuum model of groundwater hydrology is :

> The volume averaged water velocity is proportional to the directed rate of change of the volume averaged hydraulic head. The proportionality constant, that may be different in each point, is called hydraulic conductivity.

The size of the volume over which the volume averages are taken will be referred to as the averaging scale. This law of proportionality is called Darcy´s law after the person who performed the first experiments to establish it [Darcy, 1856]. There are also theoretical derivations of this physical law, see for instance [Gray and O'Neill, 1976] or [Whitaker, 1985]. These derivations, as well as the original experiment, take place in a so-called porous medium and it is an open question whether fractured crystalline rock can be considered to behave as a porous medium. We will not go further with this subject here or elsewhere in this report but only stress the fact that, the larger the averaging scale, the more likely it is that the porous media model is valid, or to put it differently, that such a concept as conductivity in fact exists for fractured rock. The limiting case of a infinitely large averaging scale corresponds to a spatially constant value of the conductivity.

Note that the averaging volume need not be a REV and that the averages may be weighted averages in the sense of [Marle, 1967] see also [de Marsily, p 15-19]. The two crucial things about the size of the averaging volumes is first, the existence of the conductivity and second that, the results may be dependent on the chosen scale. This last point may be perceived as troublesome when considering, for instance, that water travel time may be dependent on the averaging scale. The explanation for such a scale effect is of course that the tortuosity of the stream lines decreases with increasing scale. Thus any subsequent model such as a transport model based on the results of a stochastic continuum model must explicitly consider the averaging scale.

It is not particularly bold to claim that there will always exist large uncertainty about the hydrological parameters of any model pertaining to subsurface hydrology. Therefore it is natural to view the parameters as stochastic functions. In the case of the continuum model the spatial conductivity function thus will be considered to be a stochastic function. The gain with this is that one can model the spatial variation of the parameters and also quantify the uncertainty in the model output instead of having to perform a large number of ad hoc variations of deterministic models. On the other hand the solution of problems involving stochastic functions requires much more computer resources and one is still not freed of the necessity to perform variations among models but this time models of stochastic functions.

Previous hydrologic analysis performed by SKB has often used these porous media models but without the consideration of averaging scale and with the spatial variation modelled solely by fracture zones as identified and quantified by geologists. For a recent example see [Lindbom et al, 1991].

In order to use a stochastic continuum model for the particular task of modelling the groundwater situation at a hypothetical site, one would like to translate all site-specific knowledge into site specific-conductivity. Ideally all knowledge, such as for instance inferred fracture zones, flow measurements, cross hole test, fracture statistics etc. should be collected into the distribution of the conductivity seen as a stochastic process. This is not simple. For instance one attempt, among many others, in this direction is that of [Clifton and Neuman, 1982] who used measurements of hydraulic head to, in part, determine the distribution of hydraulic conductivity.

Thus for simplicity this work relies entirely on stationary single hole packer test measurements. Clearly the rate at which water can be pumped into the rock at a given overpressure is related to the conductivity in the surroundings of the location of the test. The precise relation however is not at all easy to find. In order to simplify the analysis and the implementation of a stochastic model, the following assumption is added to the previous:

> A packer test can be evaluated by Moye´s formula to give the hydraulic conductivity at an averaging scale that is related to the length of the packed off section and at a point in space identical to the midpoint of the measurement section.

This assumption can certainly be criticized. One has only to look through the assumptions done in the derivation of Moye´s formula in Appendix C or in [Moye, 1966] or to consider the likelihood of the event that two borehole sections horizontally adjacent to each other give large differences in estimated conductivity due to local fracture closure. If this latter phenomenon is common in reality the above assumption may be invalid since one may easily construct examples where the conductivity drops from high values to zero over very short spatial lengths. Clearly this would be inadmissible with the notion of a large scale conductivity. On the other hand the likelihood for this phenomenon decreases with the length of the section.

To improve the situation other similar but more elaborate ideas based on cross-correlation between packer tests and large scale conductivities has been proposed in [Norman and Geier, 1991] . The advantage with such an approach is that one can accomodate the relation between the conductivity measured by means of a packer test and Moye´s formula and the actual effective conductivity in the vicinity of the measurement location. In order to simulate both the packer tests and the effective

conductivities the analysis starts from a discrete fracture model. However, this effort has been comleted within this study.

With these two basic assumptions the problem at hand is a so-called statistical inference problem, i.e. to deduce a model in terms of a stochastic function from a set of attained values, what is known as a sample. This model could for instance consist of an expectation value or trend function together with a second-order moment function describing the fluctuation of the residuals and an assumption of second-order stationarity. This type of model will be considered below together with so-called intrinsic models.

One standard method for statistical inference methods is regression which consists essentially of giving a model in terms of a finite number of parameters which are then determined by minimizing some error with respect to the observed values. In the ordinary case, regression assumes that the residuals are uncorrelated, which of course is totally illogical in our case. In this work we instead utilize iterative generalized least square regression (IGLSE) as for instance described by [Neuman and Jacobsson, 1984]. Other known methods are the maximum likelihood (ML) restricted maximum likelihood (RML) , see for instance [Loaiciga et al, 1988], or generalized covariances (GC), see for instance [Kitanidis, 1983] or [Delfiner 1976]. A difficulty which is not taken into account in this work, or in the other ones mentioned in this section, is the fact that the measurements are censored by the presence of a measurement limit. An attempt to account for this is given by [Lovius et al, 1990].

The drawback of a parametric method, such as those mentioned above, is of course that the choice of the parameterized model can be erroneous, thus nullifying the value of the subsequent analysis. One way to analyze the situation is what is known as residual analysis, see [Draper and Smith, 1966]. However in this work we employ another, somewhat less well-known, method known as cross-validation or jacknifing, [Russo and Jury, 1987]. The main idea in this approach is to test the assumed model by removing measured values, and then to predict them using the model and employing a linear predictor, so-called kriging. This results in a vector of kriging errors which can be analyzed statistically to give a measure of the "goodness" of the tested model.

The origin of this study was to find a good, or even the best, stochastic model for the conductivity field at the Finnsjö site. Therefore all measurements used are taken from this site as found in the SKB database GEOTAB. A description of the features specific to this site is contained in [Ahlbom and Tiren,1991], [Ahlbom et al, 1988] and [Andersson et al, 1991]. In order to perform the calculations the program INFERENS was written in FORTRAN77 as a part of the work. The studied averaging scale was mainly chosen to be 36 metre since it was found suitable for groundwater flow modelling on the scales needed to describe the regional groundwater situation.

2 SCALE AND REGULARIZATION

2.1 Regularization

The precise form of the fundamental assumption of stochastic analysis performed by HYDRASTAR is that :

> A packer test can be evaluated by Moye´s formula to give the conductivity at an averaging scale that is related to the length of the packed off section and at a point in space identical to the midpoint of the measurement section.

Thus what is needed in order to vary the scale is a methodology to add stationary conductivity measurements together to achieve new sets of measurements on packer intervals other than those originally used. The primary advantage with the increase of averaging scale is that the correlation scale of the studied parameter, in this case the conductivity, increases and thus makes it possible to study a larger domain. Hence if one want to study regional flow fields, which is also a necessity due to the need to find natural boundary conditions, regularization is imperative. Of course it is to be noted that the cost of increasing the averaging scale is the loss of resolution.

Moreover there are other interesting questions. In general the set of measurements performed at a site is on different scales, that is different packer interval lengths. For instance at the Finnsjön site the measurements are predominantly performed for two and three meter sections. The question arises whether these measurements can be considered as being from the same population, or if there is a systematic difference due to the difference in scale. Secondly we may ask how dependent our results are on the scale which we are using, that is, in the case of packer measurements, what packer interval lengths are used.

Let us first describe a mathematical model for a packer test. A similar model and subsequent numerical calculations has been developed by [Braester and Thunvik, 1982]. Figure 2.1.1 shows a rock block with a drilled hole in which a packer test is performed and a cylindrical coordinate system. The z- axis of the cylindrical coordinate system coincides with the borehole. The natural head field (after the drilling) is denoted by $h_0(\mathbf{x})$ and is assumed to satisfy the steady-state hydrology equation

$$\nabla K(\mathbf{x}) \nabla h_0(\mathbf{x}) = 0$$

along with the boundary conditions

$$h_0(\mathbf{x}) = x_3$$

for x at the groundwater surface, here x_3 denotes the vertical coordinate of x. Moreover it seems safe to assume that along the borehole the hydraulic head equals a given constant, namely the groundwater level in the borehole.

As the next step insert the packers, without disturbing the natural field and pressurize the section between the packers to an overpressure Δp_1. This overpressure corresponds to a head difference $\Delta h_1 = \Delta p_1/\rho g$. Let us now write the head field during the steady state part of the packer test as $h_0(\mathbf{x}) + \Delta h_1 \cdot h_1(\mathbf{x})$ where $h_1(\mathbf{x})$ satisfies

$$\nabla K(\mathbf{x}) \nabla h_1(\mathbf{x}) = 0$$

and

$$\begin{cases} h_1(\mathbf{x})\big|_{\rho=\rho_w} = 1 & z_1 \le z \le z_1 + L_1 \\ \left.\dfrac{\partial h_1(\mathbf{x})}{\partial \rho}\right|_{\rho=\rho_w} = -\dfrac{1}{\Delta h_1}\left.\dfrac{\partial h_0(\mathbf{x})}{\partial \rho}\right|_{\rho=\rho_w} & \begin{cases} z_1 + L_1 \le z \le z_1 + L_1 + d \\ z_1 - d \le z \le z_1 \end{cases} \\ \left.\dfrac{\partial h_1(\mathbf{x})}{\partial \rho}\right|_{\rho=\rho_w} = 0 & \begin{cases} z > z_1 + L_1 + d \\ z_1 - d > z \end{cases} \\ h_1(\mathbf{x}) = 0 & \textit{elsewhere on the boundary} \end{cases}$$

where we introduced z_1 as the position of the lowest point of interval between the packers, d as the length of the packers and L_1 as the length of the interval between the packers. The phrase "elsewhere on the boundary" is used to denote the boundary of the domain in question minus the borehole wall. For this analysis it should be interpreted as the groundwater surface. The third equality in the boundary conditions above expresses a basic approximation in this analysis, namely that the leakage back into the borehole induced by the test is negligible. The fourth equation implicitly states the assumption that the groundwater surface is assumed to be unchanged. Note that we supress the angle φ from the notation but that the analysis do not assume that any of the involved fields are independent of φ.

Next we will make the reasonable assumption that the natural head gradients are small in comparison to the applied difference head Δh_1 i.e.

$$-\frac{1}{\Delta h_1}\left.\frac{\partial h_0(\mathbf{x})}{\partial \rho}\right|_{\rho=\rho_w} \approx 0 \qquad z_1 - d \le z \le z_1 + L_1 + d \qquad 2.1.1$$

so that we may write the boundary conditions of $h_1(\mathbf{x})$ as

$$\begin{cases} h_1(\mathbf{x})\big|_{\rho=\rho_w} = 1 & z_1 \le z \le z_1 + L_1 \\ \left.\dfrac{\partial h_1(\mathbf{x})}{\partial \rho}\right|_{\rho=\rho_w} = 0 & \begin{cases} z > z_1 + L_1 \\ z_1 > z \end{cases} \\ h_1(\mathbf{x}) = 0 & \textit{elsewhere on the boundary} \end{cases}$$

Similar assumptions are made by [Braester and Thunvik, 1982]. However they do not perform the above division in h_0 and h_1 but assume that the natural conditions are given by a constant potential i.e. h_0=const and treats the actual head field divided into two cases

> No flow through the borehole wall at the packers. A constant potential along the rest of the borehole wall.
>
> No flow through the borehole wall i.e. the borehole is sealed.

Thus the equations resulting from our analysis above is much like the the second case of [Braester and Thunvik, 1982] but freed of the somewhat unrealistic assumtion that h_0=const . We note also that the assumptions made in 2.1.1 is slightly stronger than necessary for our immediate purposes, we will however make use of the remaining part of the assumption later on. Another test performed at the location z_2 is treated analogously i.e. it results in a field $h_0(\mathbf{x}) + \Delta h_2 \cdot h_2(\mathbf{x})$ where

$$\nabla K(\mathbf{x}) \nabla h_2(\mathbf{x}) = 0$$

and

$$\begin{cases} h_2(\mathbf{x})\big|_{\rho=\rho_w} = 1 & z_2 \le z \le z_2 + L_2 \\ \left.\dfrac{\partial h_2(\mathbf{x})}{\partial \rho}\right|_{\rho=\rho_w} = 0 & \begin{cases} z_2 + L_2 < z \\ z_2 > z \end{cases} \\ h_2(\mathbf{x}) = 0 & \textit{elsewhere on the boundary} \end{cases}.$$

Now the idea is to show that the function $h_0(\mathbf{x}) + h_1(\mathbf{x}) + h_2(\mathbf{x})$ is an approximation to the field $h_{1+2}(\mathbf{x})$ arising as a result of both packer tests being performed at the same time with unit difference head. Since the aim is to predict the result of a packer test of length $L_1 + L_2$ the case of the measurement sections being adjacent, i.e. $z_1 + L_1 = z_2$ is treated below. This case is also depicted in figure 2.1.2. The case of non adjacent sections will be treated in section 2.1.1. [1]

---

[1] In this connection we might ponder over the overall aim of the analysis. As the reader realizes this is to use the assumption in the beginning of the current chapter, the result of the current analysis together with Moyes formula to obtain an estimate of the effective conductivity on a large averaging scale at the center of the measurement section. With this ultimate goal in mind there does not seem to be much sense in requiring that the measurement sections should be adjacent. That is only needed to simulate a packer test performed with a larger packer interval length not to obtain the effective conductivity. Thus what one should wish for is a formula that connected a set of packer measurement with the effective conductivity of a rockblock containing them.

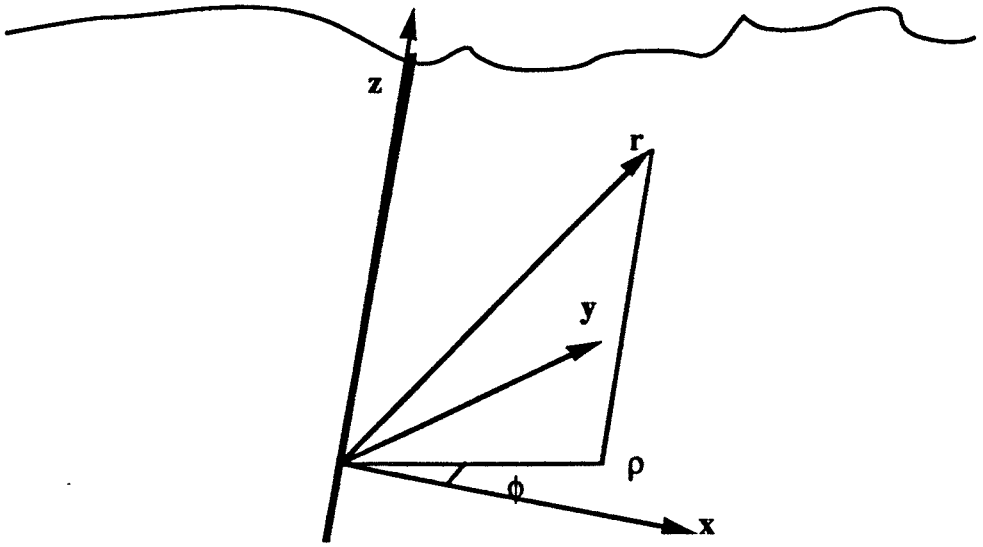


Figure 2.1.1 *Showing a borehole and the corresponding cylindrical coordinate system.*

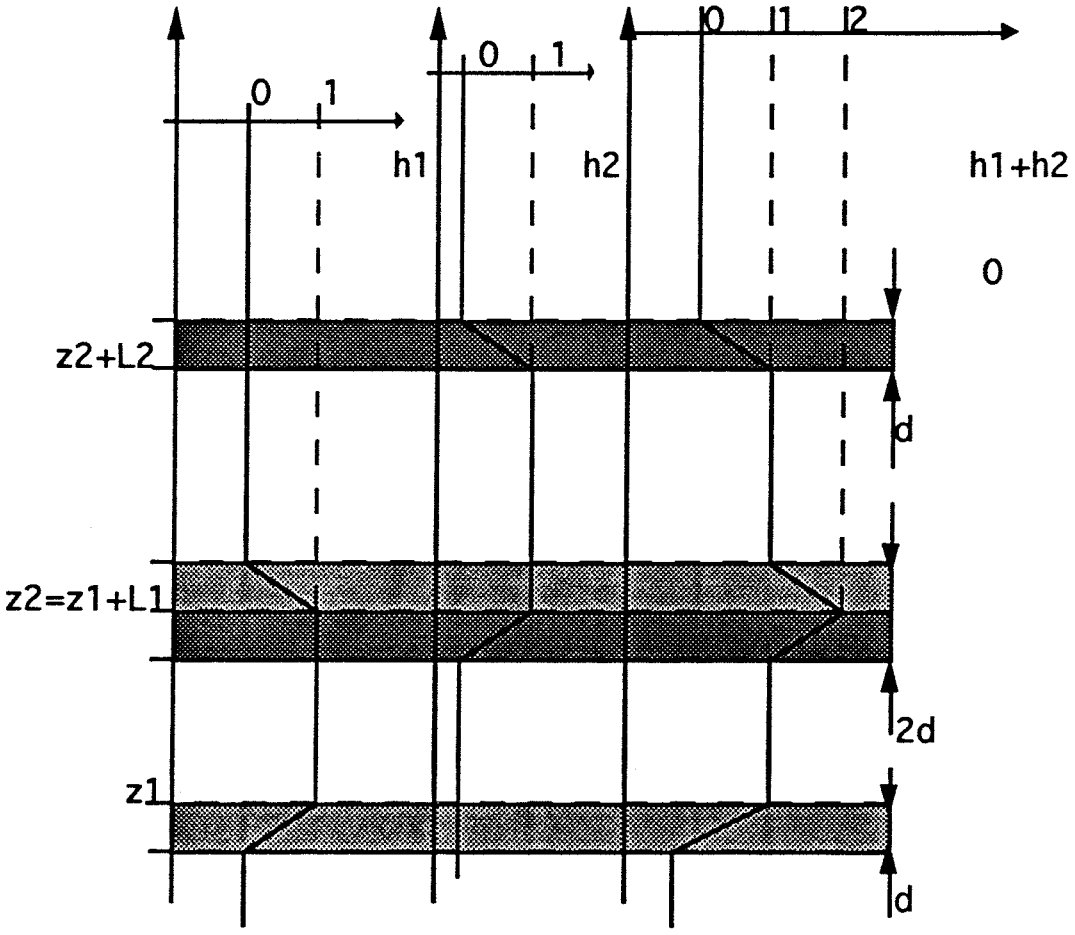


Figure 2.1.2. Showing the summing of two packer test responses.

To this end we first note that

$$0 \le h_1(\mathbf{x})\Big|_{\rho=\rho_w} \le 1 \qquad \text{for} \quad \begin{cases} z_1 + L_1 < z \\ z < z_1 \end{cases}$$

and similarly for $h_2(\mathbf{x})$. This is true since if there was a point along the borehole not satisfying the above inequality then there would exist a extremal value of $h_1(\mathbf{x})$ somewhere along the borehole outside the injection interval[2]. This point would then act like a source or a sink, contradicting the no-flow conditions.

Thus $h_1(\mathbf{x}) + h_2(\mathbf{x})$ satisfies

$$\nabla K(\mathbf{x})\nabla(h_1(\mathbf{x}) + h_2(\mathbf{x})) = 0$$

and

$$\begin{cases} 1 \le (h_1(\mathbf{x}) + h_2(\mathbf{x}))\big|_{\rho=\rho_w} \le 2 & z_1 \le z \le z_1 + L_1 + L_2 \\ \dfrac{\partial(h_1(\mathbf{x}) + h_2(\mathbf{x}))}{\partial\rho}\Big|_{\rho=\rho_w} = 0 & z < z_1,\ z > z_1 + L_1 + L_2 \\ h_1(\mathbf{x}) + h_2(\mathbf{x}) = 0 & \textit{elsewhere on the boundary} \end{cases}$$

whereas for the solution $h_{1+2}(\mathbf{x})$ of the problem posed by a packer test performed over the section length $L1 + L2$ we would require that

$$\begin{cases} h_{1+2}(\mathbf{x})\big|_{\rho=\rho_w} = 1 & z_1 \le z \le z_1 + L_1 + L_2 \\ \dfrac{\partial h_{1+2}(\mathbf{x})}{\partial\rho}\Big|_{\rho=\rho_w} = 0 & z < z_1,\ z > z_1 + L_1 + L_2 \\ h_{1+2}(\mathbf{x}) = 0 & \textit{elsewhere on the boundary} \end{cases}$$

Thus it is clear that

$$h_{1+2}(\mathbf{x}) \le h_1(\mathbf{x}) + h_2(\mathbf{x}) \le 2h_{1+2}(\mathbf{x}) \qquad z_1 \le z \le z_1 + L_1 + L_2$$

and that this double inequality implies that[3]

---

[2]Since it is clear that any function h(x) that satisfies $\nabla K(\mathbf{x})\nabla h(\mathbf{x})=0$ cannot have a local maximum in the interior of the domain of interest i.e it has to attain its extremal values on the boundary.

[3]This is a consequence of the statement: If the function h(x) satisfies $\nabla K(\mathbf{x})\nabla h(\mathbf{x}) = 0$ and

$$\begin{cases} h(\mathbf{x})\big|_{\rho=\rho_w} \ge 0 & z_1 \le z \le z_1 + L_1 + L_2 \\ \dfrac{\partial h(\mathbf{x})}{\partial\rho}\Big|_{\rho=\rho_w} = 0 & z < z_1,\ z > z_1 + L_1 + L_2 \\ h(\mathbf{x}) = 0 & \textit{elsewhere on the boundary} \end{cases}$$

then the flow from the section $[z_1, z_1+L_1+L_2]$ is nonnegative.i.e

$$\int_{z_1}^{z_1+L_1+L_2} K(\mathbf{x})\frac{\partial}{\partial\rho}h(\mathbf{x})\,\rho_w\,dz \ge 0$$

$$\int_{z_1}^{z_1+L_1+L_2} K(\mathbf{x})\frac{\partial}{\partial\rho}h_{1+2}(\mathbf{x})\,\rho_w dz \le \int_{z_1}^{z_1+L_1+L_2} K(\mathbf{x})\frac{\partial}{\partial\rho}(h_1(\mathbf{x}) + h_2(\mathbf{x}))\,\rho_w dz \le$$

$$2\int_{z_1}^{z_1+L_1+L_2} K(\mathbf{x})\frac{\partial}{\partial\rho}h_{1+2}(\mathbf{x})\,\rho_w dz$$

As the notation suggests these integrals are taken over the part of the surface of the borehole between the packers. Expanding the middle term, using again the assumption that the leakage flow is negligible, and turning the inequality inside out we have

$$\tfrac{1}{2}(q_1 + q_2) \le q_{1+2} \le q_1 + q_2 \qquad 2.1.2$$

where

$$q_1 = \int_{z_1}^{z_1+L_1} K(\mathbf{x})\frac{\partial}{\partial\rho}h_1(\mathbf{x})\,\rho_w dz,$$

$$q_2 = \int_{z_2}^{z_2+L_2} K(\mathbf{x})\frac{\partial}{\partial\rho}h_2(\mathbf{x})\,\rho_w dz$$

are known quantities proportional to the measured flows per applied difference head unit and

$$q_{1+2} = \int_{z_1}^{z_1+L_1+L_2} K(\mathbf{x})\frac{\partial}{\partial\rho}h_{1+2}(\mathbf{x})\,\rho_w dz$$

is the flow per difference head unit that would be measured if the test was performed on the section $L_1 + L_2$. It is to be stressed here that this is the crucial point where we made strong use of the no-leakage assumption since we used that

$$\int_{z_2}^{z_2+L_2} K(\mathbf{x})\frac{\partial}{\partial\rho}h_1(\mathbf{x})\,\rho_w dz = \int_{z_1}^{z_1+L_1} K(\mathbf{x})\frac{\partial}{\partial\rho}h_2(\mathbf{x})\,\rho_w dz = 0$$

It is moreover clear that an analysis such as this must make this kind of assumption since the leakage flows are not generally measured.

That the expressions for the q:s above are proportional to flow per difference head from a pressurized section follows from the assumption 2.1.1 since the flow from a pressurized section, the section $[z_1, z_1+L_1]$ say, can be written

$$q_1 := \frac{Q_1}{2\pi\Delta h_1} = \int_{z_1}^{z_1+L_1} K(\mathbf{x})\left(\frac{1}{\Delta h_1}\frac{\partial h_0(\mathbf{x})}{\partial\rho} + \frac{\partial h_1(\mathbf{x})}{\partial\rho}\right)\rho_w dz \approx$$

$$\int_{z_1}^{z_1+L_1} K(\mathbf{x})\frac{\partial h_1(\mathbf{x})}{\partial \rho}\rho_w dz \qquad 2.1.3$$

Now Moye's formula C.1, see appendix C or alternatively [Moye, 1967], can be written in terms of q = Q/2πΔh where Q is the measured flow as

$$LK = q\left(1-\ln\left(\frac{2\rho_w}{L}\right)\right)$$

and thus

$$(L_1+L_2)K_{1+2} \le (q_1+q_2)\left(1-\ln\left(\frac{2\rho_w}{(L_1+L_2)}\right)\right) =$$

$$\left(1-\ln\left(\frac{2\rho_w}{(L_1+L_2)}\right)\right)\left(\frac{L_1K_1}{\left(1-\ln\left(\frac{2\rho_w}{L_1}\right)\right)}+\frac{L_2K_2}{\left(1-\ln\left(\frac{2\rho_w}{L_2}\right)\right)}\right).$$

Introducing

$$K_{reg} = \frac{\left(1-\ln\left(\frac{2\rho_w}{(L_1+L_2)}\right)\right)}{(L_1+L_2)}\left(\frac{L_1K_1}{\left(1-\ln\left(\frac{2\rho_w}{L_1}\right)\right)}+\frac{L_2K_2}{\left(1-\ln\left(\frac{2\rho_w}{L_2}\right)\right)}\right)$$

we thus have the result that

$$\tfrac{1}{2}K_{reg} \le K_{1+2} \le K_{reg} \qquad 2.1.4$$

and hence

$$K_{1+2} \approx \frac{3}{4}K_{reg}$$

is a reasonable estimate but in this study we have used the more conservative $K_{1+2} = K_{reg}$.[4]

These results easily generalizes to n adjacent sections with the value of K_{reg} replaced by

---

[4]It may be argued that this is not conservative since choosing a certain value as a representative of a interval of possible values leads to an underestimation of the variance which in the end may lead to a more homogeneous conductivity field and thus less possibility of "fast flow paths".

$$K_{reg} = \frac{\left(1 - \ln\left(\frac{2\rho_w}{\sum_{i=1}^{n} L_i}\right)\right)}{\sum_{i=1}^{n} L_i} \sum_{i=1}^{n} \frac{L_i K_i}{\left(1 - \ln\left(\frac{2\rho_w}{L_i}\right)\right)}.$$

but due to presence of n added interval the estimate 2.1.4 instead becomes

$$\tfrac{1}{n} K_{reg} \le K_{1+2+\ldots+n} \le K_{reg}. \qquad 2.1.5$$

This is, however, a very pessimistic estimate since if we, in analogy with the above discussion, defines the scaled head response from the j:th packer test participating in the regularization by $h_j(\mathbf{x})$, the corresponding packer interval length by L_j and the packer length by d it is reasonable to assume that the $h_j(\mathbf{x})$:s are almost constant equal to zero outside the packers so that if $d<L_j$ the analog of 2.1.4

$$\tfrac{1}{2} K_{reg} \le K_{1+2+\ldots+n} \le K_{reg} \qquad 2.1.6$$

would still hold.

We note that in the common case of constant section length $L_i = L$ for all i the regularized conductivity value equals

$$K_{reg} = \frac{\left(1 - \ln\left(\frac{2\rho_w}{nL}\right)\right)}{\left(1 - \ln\left(\frac{2\rho_w}{L}\right)\right)} \frac{1}{n} \sum_{i=1}^{n} K_i$$

i.e. a corrected arithmetic mean value of the individual conductivity measurements. This behaves much different from the widely used geometric average. What is most striking is that the high measurement values dominate the sum and thus a plot of regularized measurements along the borehole tend to be constant in intervals, see the figures in section 2.2. This is rather natural if one consider how the conductivity of an averaging volume would vary when moved over a highly conductive fracture. We remark also that the size of the correction factor is of the order of one.

2.1.1 Imperfect match in regularization

We must consider two kinds of mismatch, positive (i.e. $z_1 + L_1 > z_2$) and negative i.e. ($z_1 + L_1 < z_2$), where the notation refers back to the two-measurement situation in the previous section. In the case of several sections added together with positive mismatch it is clear that 2.1.5 will always hold regardless of the magnitude of the overlap. On the other hand, following the reasoning above i.e. the assumption of constant zero head along the borehole outside the packers in each separate packer test, the estimate does not deteriorate provided that packed off sections, including the packers, does not overlap more than twice in which case 2.1.5 will still hold. The situation in the case of a

negative mismatch is much worse since there is no way to estimate the flow that would have resulted from the gap. Declining any further theoretical analysis, which could be done in the positive-overlap case, we accept the following definition.

Let us denote a measurement by the triple (z_i, L_i, K_i) where, as before, z_i denotes borehole coordinate of the lowest point of the packer test, L_i denotes the length of the packer interval used and K_i denotes the obtained conductivity value according to Moye´s formula. With a set of measurements $\{(z_i, L_i, K_i), i = 1, 2, \ldots, n\}$ we associate an interval along the borehole I_{reg} and a length L_{reg} by

$$I_{reg} = \left(\min_{1 \le i \le n}(z_i), \min_{1 \le i \le n}(z_i) + S \right)$$

and

$$L_{reg} = \max_{1 \le i \le n}(z_i + L_i) - \min_{1 \le i \le n}(z_i)$$

where S is the target regularization scale. The borehole interval I_{reg} is the packer interval of the packer test we are approximating whereas the interval $(\min(z_i), \min(z_i) + L_{reg})$ is the interval of the approximation itself with its positive and negative mismatches.

Now such a set of measurements is said to constitute a regularized measurement on the scale S at the positive tolerance level ε_p and the negative tolerance level ε_n precisely if

$$e^+_{reg} = \int \left(\sum_{i=1}^{n} \chi_{(z_i, z_i + L_i)} - 1 \right)^+ < \varepsilon_p S \quad , \qquad 2.1.1.1$$

$$e^-_{reg} = \int \left(\sum_{i=1}^{n} \chi_{(z_i, z_i + L_i)} - 1 \right)^- + |S - L_{reg}| < \varepsilon_n S \qquad 2.1.1.2$$

where χ_I denotes the characteristic function for the interval I for any interval I i.e. the function that is equal to one in the interval and identically zero outside and where $(\cdot)^+$ means $\max(\cdot,0)$ and $(\cdot)^-$ means $-\min(\cdot,0)$. The integrals extend over the interval $(\min(z_i), \min(z_i) + L_{reg})$.

In words this would amount to saying that the set of measurements $\{(z_i, L_i, K_i), i = 1, 2, \ldots, n\}$ is a regularized measurement on scale S if the sum of all positive mismatches is less than e_pS and the sum of the negative mismatches is less than e_nS. The reason for including the term $| S - L_{reg} |$ in the expression for the negative error is that it always represent a section from which the induced flow cannot be estimated in contrast with the positive type mismatch. Note that this in particular implies that

$$|S - L_{reg}| < \varepsilon_n S \quad .$$

The pair ε_n , ε_p are chosen tolerances. As is clear from the above discussion it is highly recommended to take $\varepsilon_n << \varepsilon_p$. Finally then the value of the regularized measurement is taken to be

$$K_{reg} = \frac{\left(1 - \ln\left(\frac{2\rho_w}{L_{reg}}\right)\right)}{L_{reg}} \sum_{i=1}^{n} \frac{L_i K_i}{\left(1 - \ln\left(\frac{2\rho_w}{L_i}\right)\right)}$$

where ρ_w is assumed to be a constant. Also in the case the conductivity K_i is below the measurement limit the value of the limit is substituted for the true conductivity value.

2.1.2 When are two measurements identical ?

The question in the heading arises because in certain situations it may be possible to patch together measurement sections satisfying the constraints of 2.1.1.1 and 2.1.1.2 in such a way that the resulting sections cover essentially the same part of the borehole. In spite of this the values obtained can be quite different. This is to be compared with the hypothetical situation that one performs the same measurement several times but obtains different answers due to uncontrollable factors i.e. what is usually referred to as measurement errors. This line of reasoning could be used to estimate the uncertainty of the regularized measurements but we will not pursue this any further in this report.

We will use the following definition. Two regularizations $\{(z_i, L_i, K_i), i = 1, 2, \ldots, n\}$ and $\{(z_i', L_i', K_i'), i = 1, 2, \ldots, n'\}$ represent the same measurement if

$$\left|\max_{1 \le i \le n}(z_i + L_i) - \max_{1 \le i \le n'}(z_i' + L_i')\right| + \left|\min_{1 \le i \le n}(z_i) - \min_{1 \le i \le n'}(z_i')\right| < \varepsilon_n S$$

and the value of this measurement is taken to be the arithmetic mean value of the two associated regularized conductivities i.e.

$$\frac{K_{reg} + K'_{reg}}{2} .$$

The extension to several regularizations is straightforward.

2.2 The effect of regularization

In this subsection we present some graphs on original and regularized measurement sections. The shown interpreted zone intersections are taken from [Ahlbom and Tiren, 1991]. The conclusions drawn from these graphs are discussed in the next section.

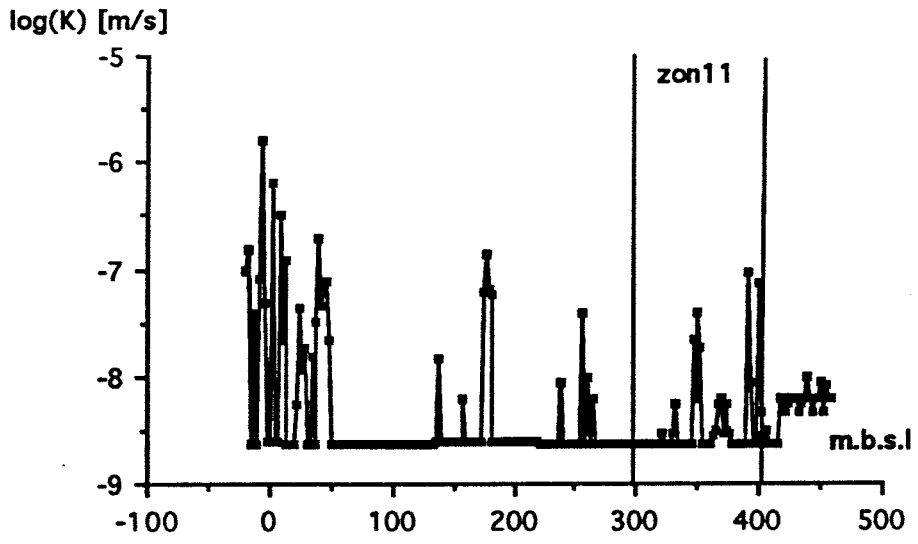


Figure 2.2.1 *The original measurements in borehole kfi01. The value of the measurement limit 2.4E-9 has been substituted for measurements below the measurement limit.*

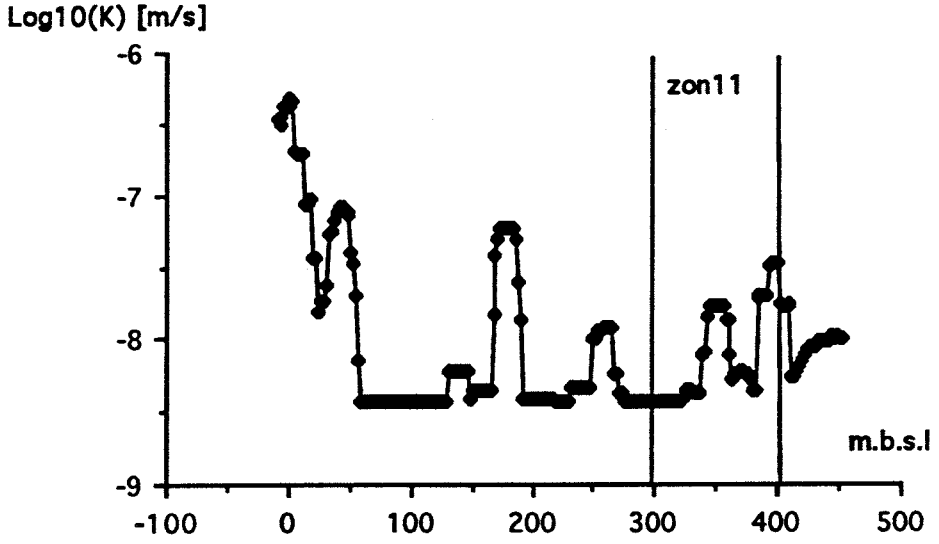


Figure 2.2.2 *The measurements in borehole kfi01 regularized to 18m.*

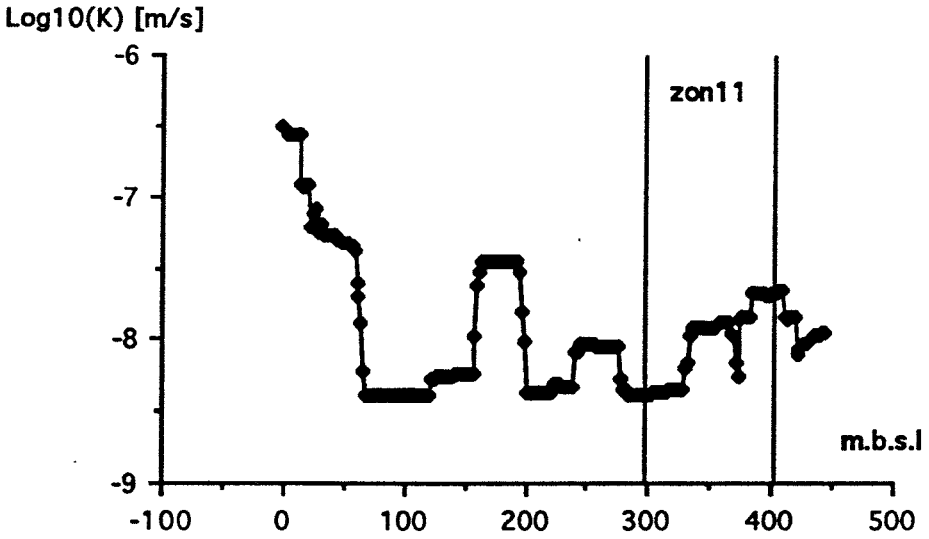


Figure 2.2.3 *The measurements in borehole kfi01 regularized to 36m.*

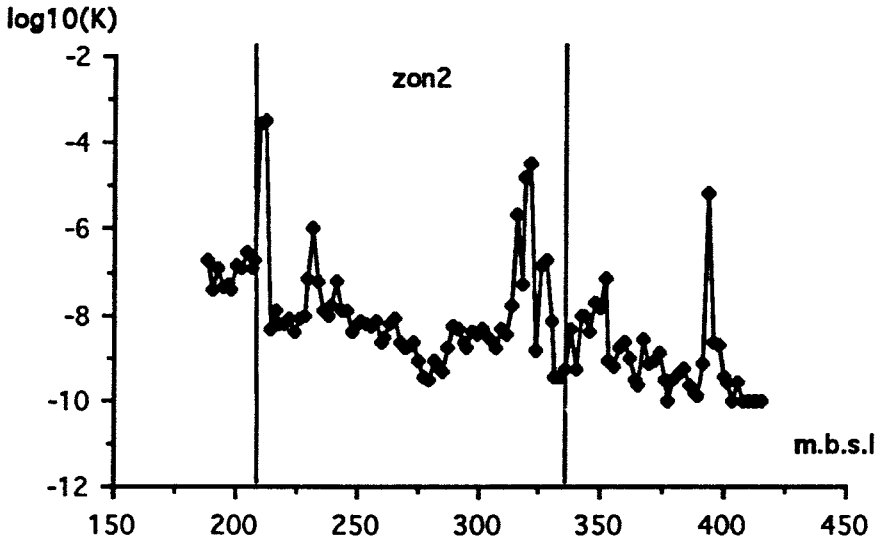


Figure 2.2.4 *The original measurements in borehole bfi01. The value of the measurement limit 1.0E-10 has been substituted for measurements below the measurement limit.*

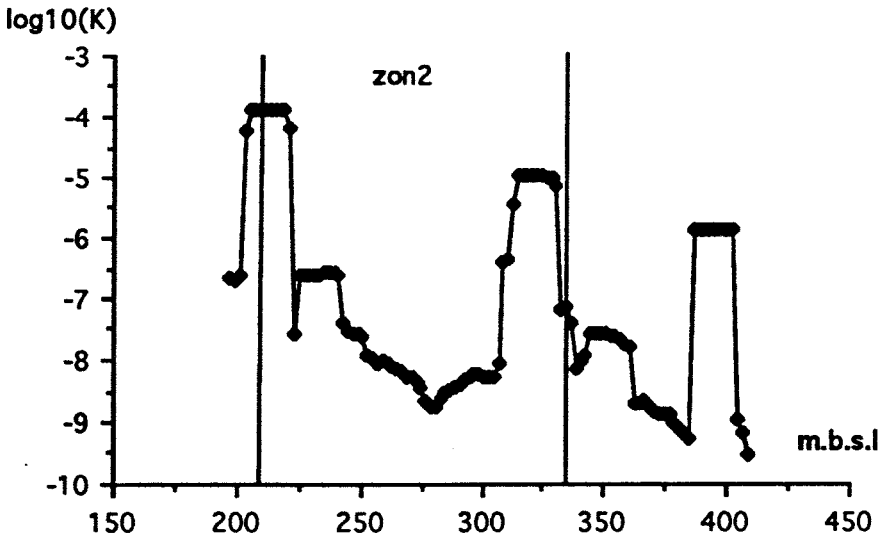


Figure 2.2.5 The measurements in borehole bfi01 regularized to 18 m.

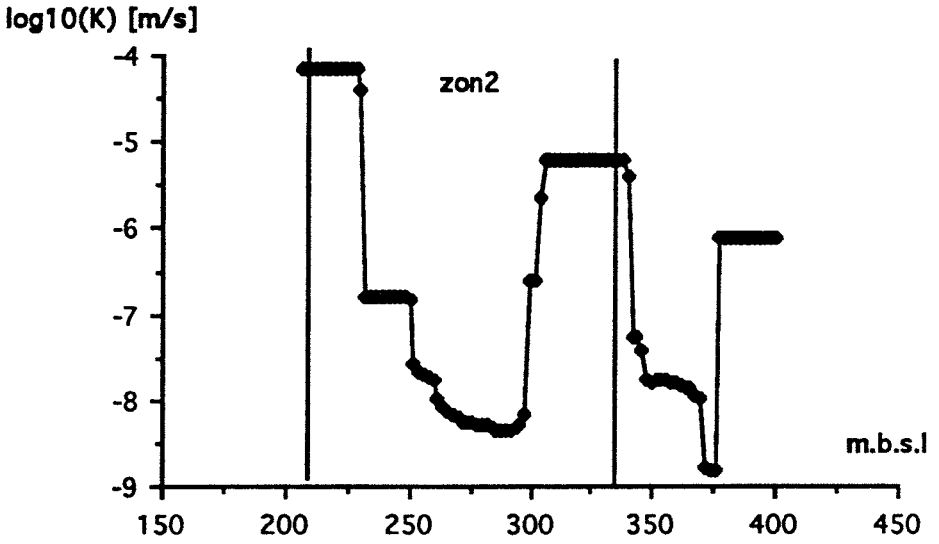


Figure 2.2.6 The measurements in borehole bfi01 regularized to 36m.

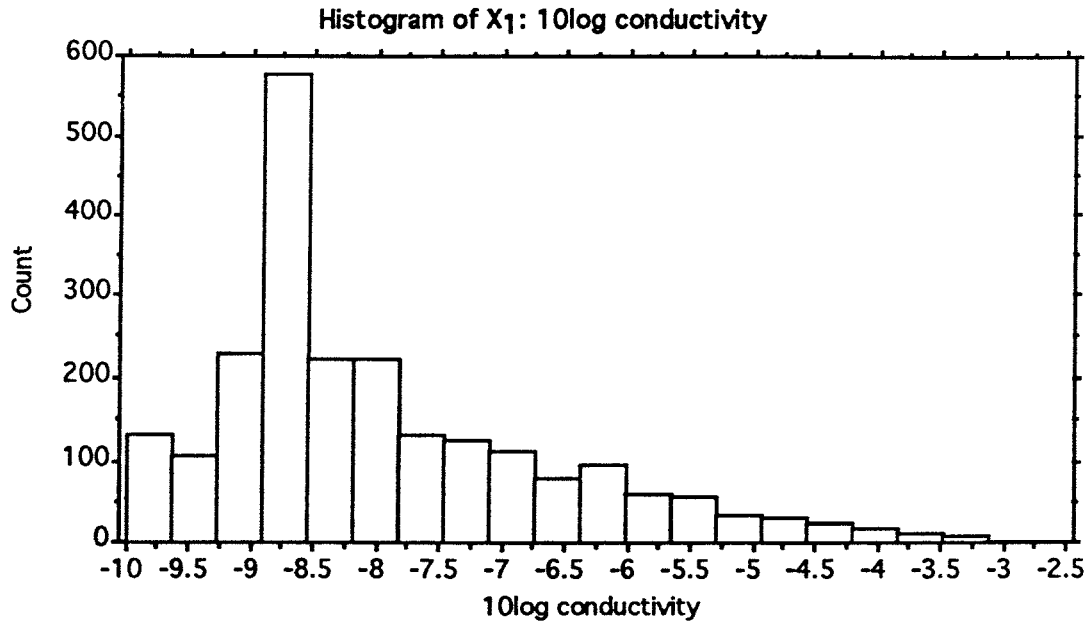


Figure 2.2.7. *Histogram of all (2258) original measurements of log conductivity. The value of the measurement limits have been substituted for measurements below the measurement limit..*

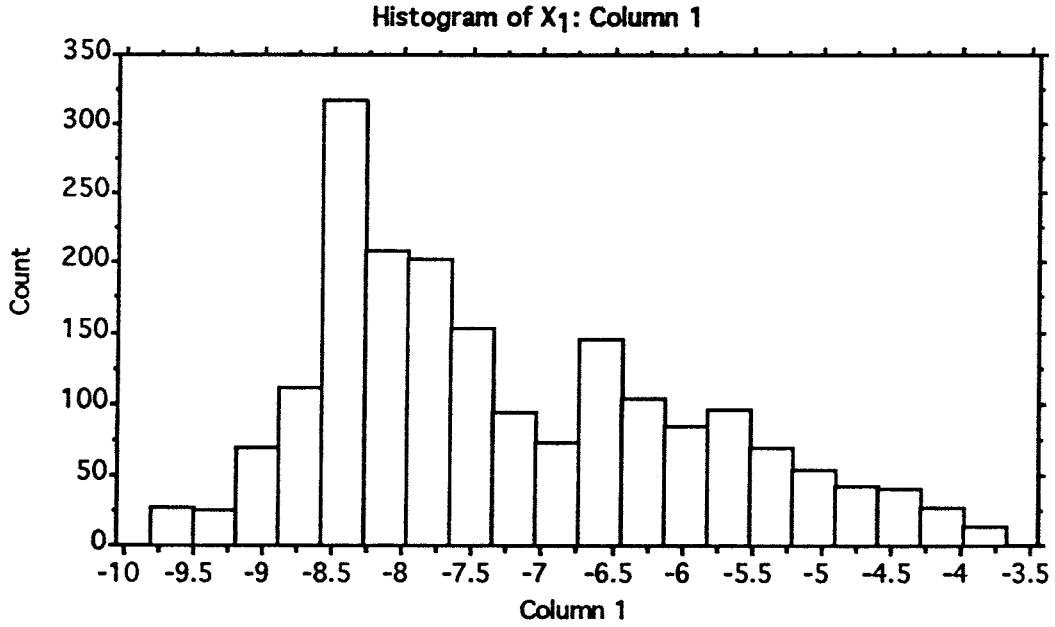


Figure 2.2.8. *Histogram of all (1952) 18 m measurements of log conductivity.*

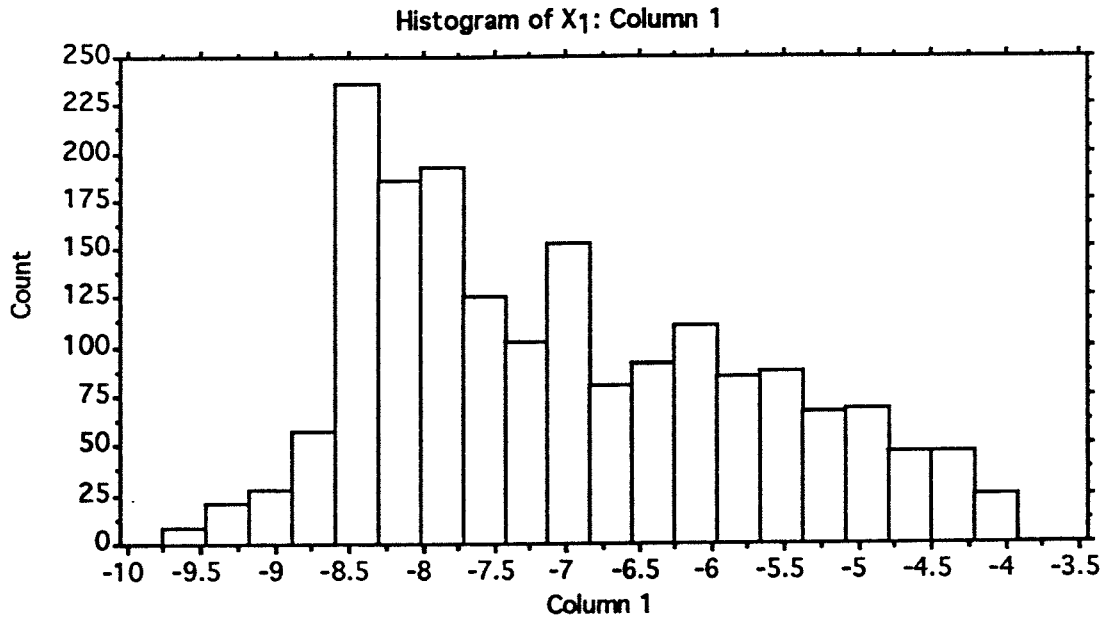


Figure 2.2.9. Histogram of all (1815) 36 m measurements of log conductivity.

2.3 Conclusions

As pointed out by geologists the high conductivities in zone 2 are located at the upper and lower bounding surface of the zone. It is shown by the above graphs that this quality holds for the regularized measurements, at least up to 36m, as well. The graphs also show that the high conductivity values are (almost) confined within the interpreted zone boundaries. Thus a trend function for zone 2 should be bimodal, contrary to the one used below, which corresponds solely to the geometric characteristics of the zone. However another important feature shown in the above graphs is that there exists bands of high conductivities outside the interpreted zones as well. This shows the difficulty with the use of trend functions to describe bands of high conductivities. That is, one may find them in the interpreted zones but, on the other hand, they may be outside the interpreted zones completely. This, together with the simplicity, suggests that trendless or intrinsic models are most suitable.

As for the histogram the effect of the regularization is to slightly diminish the range of the conductivity values and to even out the shape. Also there is an increase in average log conductivity with increasing scale. However the most striking feature of these histograms are that they do not look as if taken from a normal distribution, which is a usual assumption in the literature, and also in this work. Thus it would be a worthwhile effort to find new Gaussian transforms. In this connection we remark that the regularization formula does not preserve log-normality thus the possible finding of a Gaussian transform should be done separately on each scale.

3 STOCHASTIC FUNCTIONS

3.1 General

In this report it will mostly suffice to think of a stochastic function Y as a function of two arguments $Y(\mathbf{x}, \omega)$, the space argument $\mathbf{x}$, and the event space argument ω. For fixed $\mathbf{x}$ the function is an ordinary stochastic variable and for fixed ω it becomes a spatial function, a realization. Let us take the opportunity to introduce some notation below that will be used throughout the report.

Define

$$m_Y(\mathbf{x}) = E[Y(\mathbf{x}, \omega)]$$

where $E[\cdot]$ denotes the expectation value operator. A primed quantity will always denote the mean removed form so for instance

$$Y'(\mathbf{x}, \omega) = Y(\mathbf{x}, \omega) - E[Y(\mathbf{x}, \omega)]$$

and $C_Y(\mathbf{x}, \xi)$ is the centered covariance function of Y the definition of which can be written as

$$C_Y(\mathbf{x}, \xi) = E[Y'(\mathbf{x} + \xi, \omega) Y'(\mathbf{x}, \omega)]$$

where ξ denotes the vector separating two points of consideration, the so-called lag vector. A stochastic function is said to be weakly (or second order) stationary if the expectation value function $m_Y(\mathbf{x})$ and the centered covariance function $C_Y(\mathbf{x}, \xi)$ is independent of $\mathbf{x}$ i.e. if

$$\begin{cases} m_Y(\mathbf{x}) = m_Y \\ C_Y(\mathbf{x}, \xi) = C_Y(\xi) \end{cases}.$$

Finally $\sigma(\mathbf{x})$ is the standard deviation at $\mathbf{x}$ i.e. if Y is weakly stationary

$$\sigma_Y(\mathbf{x})^2 = C_Y(\mathbf{0}).$$

In the above notational definitions $Y(\mathbf{x})$ may be replaced by any other stochastic function without altering the meaning of the notation however the index will be dropped when it is apparent by context which stochastic function is referred to. Furthermore the dependence on the event space variable ω will be suppressed from now on. For a more thorough description of random function see appendix A.

3.2 Intrinsic random functions

A stochastic function $Y(\mathbf{x})$ is said to be intrinsic if the following requirements hold:

(i) The second order moment of an increment

$$\gamma_Y(\xi) = \tfrac{1}{2}E\left[\left(Y(\mathbf{x}+\xi) - Y(\mathbf{x})\right)^2\right] \qquad 3.2.1$$

exists and is independent of **x**. The function $\gamma(\mathbf{x})$ is known as the semivariogram.

(ii) The first order moment of the increment is zero i.e.

$$E[Y(\mathbf{x}+\xi) - Y(\mathbf{x})] = 0 \qquad 3.2.2$$

Note that the condition above is a much weaker condition than to require weak second order stationarity. In particular an intrinsic random function need not to have a finite variance i.e. need not to be of second order. In case the stochastic function has a finite variance the semivariogram is closely related to the covariance function since

$$\gamma_Y(\xi) = \tfrac{1}{2}E\left[\left(Y(\mathbf{x}+\xi) - Y(\mathbf{x})\right)^2\right] =$$

$$\tfrac{1}{2}E\left[\left(Y'(\mathbf{x}+\xi) - Y'(\mathbf{x})\right)^2\right] =$$

$$\sigma(\mathbf{x})^2 - C_Y(\xi) \qquad 3.2.3$$

4 KRIGING

Assume that we have a random function Y(x) which is known at a number of points, the data support, x_i, i = 1, 2, . . . , N. What is the best linear unbiased estimator Y*(x)

$$Y(\mathbf{x})^* = \sum_{i=1}^{N} \lambda_i(\mathbf{x}) Y(\mathbf{x}_i) \qquad 4.1$$

of Y(x), given this information, in the sense that the centered variance of the interpolation error i.e.

$$V\left[(Y(\mathbf{x})^* - Y(\mathbf{x}))^2\right] = E\left[(Y(\mathbf{x})^* - Y(\mathbf{x}))^2\right] - E\left[(Y(\mathbf{x})^* - Y(\mathbf{x}))\right]^2$$

should be minimal ? Here V[·] is the centered variance operator. The multiplicators $\lambda_i(\mathbf{x})$ appearing in 4.1 will be referred to as the kriging weights at x. These are functions of the point of estimation, x. The reader is urged to make clear the difference between the estimator Y(x)*, which is a stochastic variable, and the estimate y*(x) which is realization of Y(x)*. The estimator in 4.1 is the so-called kriging estimator and the process of employing it for estimating for instance level curves of a stochastic function is known as kriging.

There are a number of different approaches to this simple problem and we shall consider them one by one.

4.1 Residual kriging

Here we assume that the trend is known i.e. in some way estimated. We may then work with the residual process Y´(x) = Y(x) - E[Y(x)] only. Since the expectation of both the residual and any linear estimator of it then becomes zero the centered variance of interpolation error above is written as

$$V\left[(Y'(\mathbf{x})^* - Y'(\mathbf{x}))^2\right] = E\left[(Y'(\mathbf{x})^* - Y'(\mathbf{x}))^2\right] =$$
$$\sum_{i,j=1}^{N} \lambda_i(\mathbf{x})\lambda_j(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x}_j) - 2\sum_{i=1}^{N} \lambda_i(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x}) + C(\mathbf{0})$$

By differentiating with respect to $\lambda_i(\mathbf{x})$ we see that the centered variance is minimized[5] when

---

[5]That this stationary point is in fact a minimum follows, since a covariance function is always positive definite.

$$\sum_{i=1}^{N}\lambda_j(\mathbf{x})C(\mathbf{x}_i-\mathbf{x}_j)=C(\mathbf{x}_i-\mathbf{x}) \qquad i=1,\ldots N$$

which are the so-called kriging equations. Inserting this equality into the above expression for the interpolation error, we obtain the kriging variance as

$$E\left[\left(Y'(\mathbf{x})^*-Y'(\mathbf{x})\right)^2\right]=C(\mathbf{0})-\sum_{i=1}^{n}\lambda_i(\mathbf{x})C(\mathbf{x}_i-\mathbf{x})$$

4.2 Kriging in the locally stationary case

If one does not know the trend or wants to avoid estimating it, one can instead make the assumption that the trend is locally stationary . This means that for the task of determining the best linear estimator of Y at **x**, $Y^*(\mathbf{x})$, one makes the assumption that in a neighbourhood of **x**, the trend E[Y(**x**)] is unknown but approximately equal to a constant. To this neighborhood there corresponds a set D(**x**) of measurement locations defined by the requirement that the points in D(**x**), i.e. $\{\mathbf{x}_i\}_{i\in D(\mathbf{x})}$, belongs to the neighbourhood. The question of finding the best linear unbiased estimator is then restricted to using only points in D(**x**) i.e. $Y(\mathbf{x})^*$ is given by

$$Y(\mathbf{x})^*=\sum_{i\in D(\mathbf{x})}\lambda_i(\mathbf{x})Y(\mathbf{x}_i) \qquad 4.2.1$$

The set D(**x**) is usually called a kriging neighbourhood and can also be used, in general, to restrict the size of the kriging equation system by taking only the closest points into account.

With these assumptions we easily see that in order to ensure that the linear estimator should be unbiased , that is have the correct expectation value , we have only to require that

$$\sum_{j\in D(\mathbf{x})}\lambda_j(\mathbf{x})=1 \qquad 4.2.2$$

Next, with this in mind, we can write the centered variance of the interpolation error as

$$V\left[\left(Y(\mathbf{x})^*-Y(\mathbf{x})\right)^2\right]=E\left[\left(Y(\mathbf{x})^*-Y(\mathbf{x})\right)^2\right]=$$

$$E\left[\left(\sum_{i\in D(\mathbf{x})}\lambda_i(\mathbf{x})Y(\mathbf{x}_i)-Y(\mathbf{x})\right)^2\right]=$$

$$E\left[\left(\sum_{j\in D(\mathbf{x})}\lambda_i(\mathbf{x})(Y(\mathbf{x}_i)-m(\mathbf{x}_i))-(Y(\mathbf{x})-m(\mathbf{x}))\right)^2\right]$$

$$\sum_{i,j\in D(\mathbf{x})}\lambda_i(\mathbf{x})\lambda_j(\mathbf{x})C(\mathbf{x}_i-\mathbf{x}_j)-2\sum_{i\in D(\mathbf{x})}\lambda_i(\mathbf{x})C(\mathbf{x}_i-\mathbf{x})+C(\mathbf{0}) \qquad 4.2.3$$

in principle as before. Note the use of the assumption of local stationarity in the third equality above i.e. we used

$$m(\mathbf{x}) - \sum_{j \in D(\mathbf{x})} \lambda_i(\mathbf{x}) m(\mathbf{x}_i) = 0 \qquad 4.2.4$$

which is equivalent with the assumption a if Y is locally stationary.

The minimization of this interpolation error 4.2.3 under the constraint of unbiasedness 4.2.2 is obtained by differentiating the Langrangian function

$$\sum_{i,j \in D(\mathbf{x})} \lambda_i(\mathbf{x})\lambda_j(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x}_j) - 2 \sum_{i \in D(\mathbf{x})} \lambda_i(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x}) + C(\mathbf{0}) +$$

$$\mu(\mathbf{x}) \left[\sum_{j \in D(\mathbf{x})} \lambda_j(\mathbf{x}) - 1 \right]$$

which leads to the linear system for determining a stationary point[6]

$$\begin{cases} \sum_{j \in D(\mathbf{x})} \lambda_j(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x}_j) + \mu(\mathbf{x}) = C(\mathbf{x}_i - \mathbf{x}) & i \in D(\mathbf{x}) \\ \sum_{i \in D(\mathbf{x})} \lambda_j(\mathbf{x}) = 1 \end{cases} \qquad 4.2.5$$

where $\mu(\mathbf{x})$ is a Lagrange multiplier for the kriging system at $\mathbf{x}$. The interpolation error associated with the above approach is given by inserting the kriging system into the expression for centered variance of the interpolation error 4.2.3 as

$$E\left[(Y(\mathbf{x})^* - Y(\mathbf{x}))^2 \right] = C(\mathbf{0}) - \mu(\mathbf{x}) - \sum_{i \in D(\mathbf{x})} \lambda_i(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x})$$

The above approach can be generalized to more general forms of local trend. That is, instead of assuming that the trend is locally constant we may assume another, more complicated, form. The most common form of the local trend is a second order polynomial, where we assume that locally the expectation of Y(x) is given by

$$E[Y(\mathbf{x})] = m_0 + \mathbf{m}_1 \bullet \mathbf{x} + \mathbf{m}_2 \bullet \mathbf{x}^T \mathbf{x}$$

where m_0 is a constant scalar, $\mathbf{m}_1$ is a constant vector, $\mathbf{m}_2$ a constant symmetric matrix and $\bullet$ signifies the inner (tensor) product [7]. The nonbias equation 4.2.4 is thus replaced by

---

[6]Which is also a minimum since if C is positive semi definite it is positive semi definite on any subspace as well.

[7]The outer matrix (tensor) product is defined as: Let S_1 and S_2 be two arbitrary NxM matrices then

$$m_0\left(1-\sum_{j\in D(\mathbf{x})}\lambda_i(\mathbf{x})\right)+\mathbf{m}_1\bullet\left(\mathbf{x}-\sum_{j\in D(\mathbf{x})}\lambda_i(\mathbf{x})\mathbf{x}_i\right)+\mathbf{m}_2\left(\mathbf{x}^T\mathbf{x}-\sum_{j\in D(\mathbf{x})}\lambda_i(\mathbf{x})\mathbf{x}_i{}^T\mathbf{x}_i\right)=0$$

or since we want this to hold independent of the choice of m_0, $\mathbf{m}_1$ and $\mathbf{m}_2$ with the ten conditions

$$\begin{cases}1-\sum_{j\in D(\mathbf{x})}\lambda_i(\mathbf{x})=0\\ \mathbf{x}-\sum_{j\in D(\mathbf{x})}\lambda_i(\mathbf{x})\mathbf{x}_i=0\\ \mathbf{x}^T\mathbf{x}-\sum_{j\in D(\mathbf{x})}\lambda_i(\mathbf{x})\mathbf{x}_i{}^T\mathbf{x}_i=0\end{cases}$$

and thus the number of Lagrange multipliers is likewise increased to ten. This is known as universal kriging. The disadvantage of this approach, independent of the the number of constants to describe the trend (also in the main case of one constant above), is that one must estimate the form of the covariance function of the residuals. First of all this is a difficult task and secondly, if one completes the task, one obtains the trend as a part of the result. See section 6.3 below and [de Marsily, 1986, pp 310-312].

4.3 Kriging in the intrinsic case

When using intrinsic random functions in relation to kriging, one usually assumes that 3.2.2

$$E[Y(\mathbf{x}+\xi)-Y(\mathbf{x})]=0$$

holds only locally. This equation is equivalent to the corresponding assumption made on the expectations in the section above.

To derive the kriging system in this case one first notes that the condition for unbiasedness again becomes

$$\sum_{j\in D(\mathbf{x})}\lambda_j(\mathbf{x})=1 \qquad .\qquad 4.3.1$$

And because of this, one may express the (centered) variance of the kriging error as

---

$$\mathbf{S}=\sum_{i=1}^{N}\sum_{j=1}^{M}S_{1_{i,j}}S_{2_{i,j}}$$

$$V\left[(Y(\mathbf{x})^* - Y(\mathbf{x}))^2\right] = E\left[(Y(\mathbf{x})^* - Y(\mathbf{x}))^2\right] =$$

$$E\left[\left(\sum_{i \in D(\mathbf{x})} \lambda_i(\mathbf{x}) Y(\mathbf{x}_i) - Y(\mathbf{x})\right)^2\right] =$$

$$E\left[\left(\sum_{i \in D(\mathbf{x})} \lambda_i(\mathbf{x})[Y(\mathbf{x}_i) - Y(\mathbf{x})]\right)^2\right] =$$

$$\sum_{i,j \in D(\mathbf{x})} \lambda_i(\mathbf{x})\lambda_j(\mathbf{x}) E\left[(Y(\mathbf{x}_i) - Y(\mathbf{x}))(Y(\mathbf{x}_j) - Y(\mathbf{x}))\right]$$

Now what is needed is the following observation

$$2(Y(\mathbf{x}_i) - Y(\mathbf{x}))(Y(\mathbf{x}_j) - Y(\mathbf{x})) =$$

$$(Y(x_i) - Y(x))^2 + (Y(\mathbf{x}_j) - Y(\mathbf{x}))^2 - (Y(\mathbf{x}_i) - Y(\mathbf{x}_j))^2$$

Making use of this identity we can express the variance of the interpolation error in terms of the semivariogram as

$$2\sum_{i \in D(x)} \lambda_i(x)\gamma(\mathbf{x}_i - \mathbf{x}) - \sum_{i,j \in D(\mathbf{x})} \lambda_i(\mathbf{x})\lambda_j(\mathbf{x})\gamma(\mathbf{x}_i - \mathbf{x}_j)$$

and differentiation of the corresponding Lagrangian function, incorporating the side condition 4.3.1, with respect to the kriging weights gives the kriging system

$$\begin{cases} \sum_{j \in D(\mathbf{x})} \lambda_j(\mathbf{x})\gamma(\mathbf{x}_i - \mathbf{x}_j) - \mu(\mathbf{x}) = \gamma(\mathbf{x}_i - \mathbf{x}) & i \in D(\mathbf{x}) \\ \sum_{i \in D(\mathbf{x})}^{n} \lambda_j(\mathbf{x}) = 1 \end{cases} \qquad 4.3.2$$

where $\mu(\mathbf{x})$ again is a Lagrange multiplier for the kriging system at $\mathbf{x}$.

The interest in this derivation is mainly in the differences between it and the derivation for the locally constant case. The kriging variance is obtained as previously by inserting the kriging system into the expression for the variance of the interpolation error to give

$$\sum_{i \in D(x)} \lambda_i(x)\gamma(\mathbf{x}_i - \mathbf{x}) - \mu(\mathbf{x})$$

For an alternative and more elegant derivation of the results of this section see appendix A.

4.4 Solution of the kriging equations

First of all let us stress the fact that any covariance matrix $\mathbf{C} = \{C(\mathbf{x}_i - \mathbf{x}_j)\}_{i,j}$ is positive semidefinite, a fact that which we have already used extensively. This is easily seen since

$$\lambda^T C\lambda = \sum_{1\le i,j\le n} \lambda_i C_Y(\mathbf{x}_i - \mathbf{x}_j)\lambda_j = V\left[\sum_{1\le i\le n} \lambda_i Y(\mathbf{x}_i)\right] \ge 0 \quad [8]$$

This is an important quality to keep when solving the kriging system. Now the kriging matrix i.e. the matrix in the solution in the kriging system 4.2.5 or its corresponding generalization to universal kriging is

$$\mathbf{D} = \begin{bmatrix} \mathbf{C} & \mathbf{X} \\ \mathbf{X}^T & 0 \end{bmatrix} \qquad 4.4.1$$

or in the case 4.3.2 [9]

$$\mathbf{D} = \begin{bmatrix} \Gamma & \mathbf{X} \\ \mathbf{X}^T & 0 \end{bmatrix} \qquad 4.4.2$$

are both indefinite and thus it is good strategy to write the solution of the kriging system

$$\mathbf{D}\begin{bmatrix} \lambda(\mathbf{x}) \\ \mu(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{c}(\mathbf{x}) \\ 1 \end{bmatrix} \qquad 4.4.3$$

or

$$\mathbf{D}\begin{bmatrix} \lambda(\mathbf{x}) \\ \mu(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{x}) \\ 1 \end{bmatrix} \qquad 4.4.4$$

in terms of the inverse of the covariance matrix, $\mathbf{C}^{-1}$. In the first case 4.4.1 we note that

$$\mathbf{D} = \begin{bmatrix} \mathbf{C} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{X}^T & -\mathbf{X}^T\mathbf{C}^{-1}\mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{C}^{-1}\mathbf{X} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

and thus the solution to the kriging system 4.4.3 can be written

$$\mu = \frac{1 - \mathbf{X}^T\mathbf{C}^{-1}\mathbf{c}(\mathbf{x})}{s}$$

$$\lambda(\mathbf{x}) = \mathbf{C}^{-1}\mathbf{c}(\mathbf{x}) - \mu\mathbf{w}$$

---

[8] An analogous property holds for the semivariogram namely it is conditionally negative semidefinite, see Appendix «airf».

[9] Or its generalization to intrinsic random functions of order higher than zero,see Appendix «irf».

where we introduced

$$\mathbf{w} = \mathbf{C}^{-1}\mathbf{X}$$

$$s = -\mathbf{X}^T \mathbf{C}^{-1}\mathbf{X}.$$

We call the vector **w** the "Lagrange correction vector" since it arises from the Lagrange multiplier formulation of the minimum variance formulation above.

In the case 4.4.2, where the matrix Γ results from a semivariogram without a sill, a pseudo covariance [Journel and Huijbregts, 1978, p 306] $C_p(\mathbf{x})$ is defined by

$$\gamma(\xi) = C_0 - C_p(\xi)$$

and thus

$$\mathbf{C}_p = C_0\mathbf{1}\mathbf{1}^T - \Gamma \qquad 4.4.5$$

where C_0 is a constant greater than the maximal element of Γ. Note however that it is not certain that the resulting matrix is positive definite.

5 MODELS OF STOCHASTIC FUNCTIONS

In this chapter some characteristics of models of stochastic functions will be discussed. The main difference between the models employed in this work is whether a trend is explicitly modelled, in which case we will talk of a residual type model, or if the trend is modelled implicitly by a local stationarity assumption as described in section 4.2. This latter case is referred to as, somewhat misleading, the intrinsic case.

5.1 Covariance models.

As discussed in section 3 and appendix A many properties of a stochastic function can be explained if one knows the semivariogram. However, as we have already seen, not any function can be chosen as a semivariogram or for that matter a covariance function. Two models that guarantee the definiteness properties are presented in this section and will be used in our subsequent work. These models are both what is referred to as **transition models** [Journel and Huijbregts, 1978] which means that they possess a finite variance and thus have equivalent formulations in terms of covariance functions.

The models are basically isotropic. The first one is the so-called **spherical model** and is written in the form of a covariance function as

$$C(\xi) = \begin{cases} V\left(1 - \frac{3}{2}\frac{\|\xi\|}{a} + \frac{1}{2}\frac{\|\xi\|^3}{a^3}\right) & 0 \le \|\xi\| \le a \\ 0 & \|\xi\| > a \end{cases} \qquad 5.1.1$$

and the the other is an **exponential model**

$$C(\xi) = V \exp(-\lambda\|\xi\|) \qquad \forall\, \xi \qquad 5.1.2$$

where in both these expressions V signifies the variance, ξ is the vector separating two measurement points, i.e. the lag vector and a and λ are parameters determining the range of the stochastic function with the corresponding covariance function.[10] The **range**, or correlation length, of a stochastic function Y(x) is the maximal distance separating two points x_1 and x_2 over which the the stochastic variables $Y(x_1)$ and $Y(x_2)$ are correlated. We note that as a rule of thumb the practical range of a stochastic function with an exponential covariance function is $3/\lambda$ whereas for a stochastic function with an spherical covariance function the range is equal to a.

A simple way to model stochasticly anisotropic fields is to use so-called geometrical anisotropy i.e. to write

---

[10]These models could of course have been stated in the semivariogram form, see «sfung.c».

$$C(\xi) = C_{iso}(G\xi)$$

where C_{iso} is an isotropic model and G is the matrix of geometrical anisotropy. This will transform the level surfaces of the covariance functions from concentric spheres to confocal ellipsoids. The mapping

$$\eta = G\xi$$

will be said to transform the (lag) space into the isotropic (lag) space. In this work we will restrict G to the form

$$G = \begin{bmatrix} \lambda_h & 0 & 0 \\ 0 & \lambda_h & 0 \\ 0 & 0 & \lambda_v \end{bmatrix}.$$

This means that the resulting covariance ellipsoids have two main axes in the horizontal plane and one orthogonal to it.

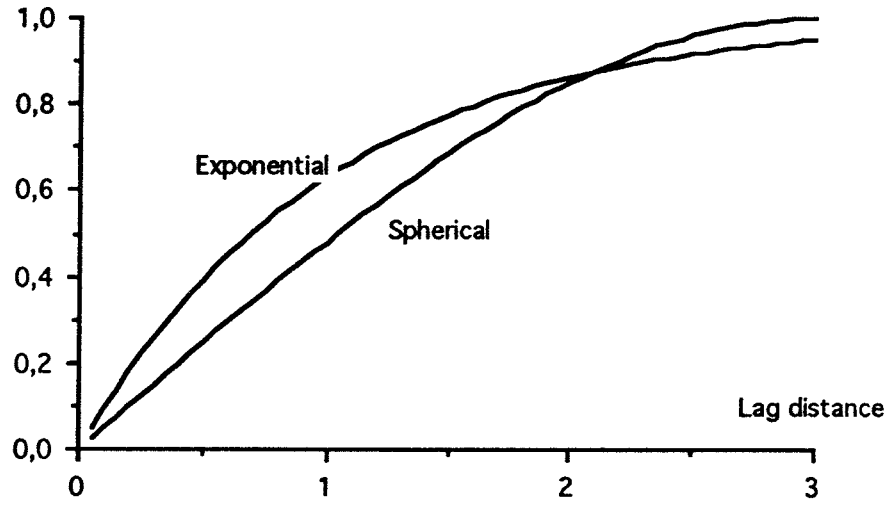


Figure 5.1.1 *Showing the exponential and spherical variogram models. They both have finite variance equal to 1.0 and practical range equal to 3.0.*

Further models of covariance functions can be obtained by adding covariance functions of the above classes to obtain what is known as nested models [Journel and Huijbregts, 1978]. This however will not be employed in this work.

5.2 Kriging neighbourhoods

Kriging neighbourhoods are discussed in chapter 4. The particular form of these employed in this study are inclined slices, see figure 5.2.1, for instance the kriging neighbourhood for the point x_1 is given by

$$D(\mathbf{x}_1) = \left\{ \mathbf{x} \in \mathbf{R}^3 : -o \leq (\mathbf{x} - \mathbf{x}_p(\mathbf{x}_1)) \bullet \hat{\mathbf{n}} \leq w + o \right\}$$

where the width w and the overlap o are positive real numbers, $\mathbf{n}\hat{}$ is the normal of the slices and $\mathbf{x}_p(\mathbf{x}_1)$ is a point defining a kriging set[11] . The kriging set containing the point $\mathbf{x}_1$, $E(\mathbf{x}_1)$ is defined as

$$E(\mathbf{x}_1) = \left\{ \mathbf{x} \in \mathbf{R}^3 : 0 \leq (\mathbf{x} - \mathbf{x}_p(\mathbf{x}_1)) \bullet \hat{\mathbf{n}} < w \right\}$$

and thus consists of slices of width w inclined in the same dirstion as the kriging neighbourhoods. This implies that the kriging neighbourhood is unchanged for all points in a kriging set

$$D\left(\mathbf{x}_p(\mathbf{x}_1) + s\hat{\mathbf{n}} + \mathbf{b}\right) = D(\mathbf{x}_1)$$

if $s \leq w$ is a real number and $\mathbf{b}$ is any vector orthogonal to the kriging neighbourhood normal $\mathbf{n}\hat{}$. The function $\mathbf{x}_p(\mathbf{x}_1)$ is chosen so that the set of all kriging sets is mutually disjoint, i.e. there is no point that lies in two kriging sets, and covers the domain of interest. [12]

Since, as explained in chapter 4, the primary requirement on a kriging neighbourhood is the constancy of the expected value function this choice of kriging neighbourhoods is motivated by the well known decrease of hydraulic conductivity with increasing depth. This decrease will thus be achieved without the use of an explicit trend function. Also inclination of the slices can be motivated by an assumption that the level surfaces of the hydraulic conductivity is oriented parallel with the subhorizontal zone 2.

---

[11]This is a nonstandard notation invented by the author.

[12]It may be mentioned here that HYDRASTAR.1.1 allows the definition of primary kriging neighbourhoods and sets prior to the slicing defined here. This can be used to treat the zones as different from the rest of the rock. This possibility has not been used here however.

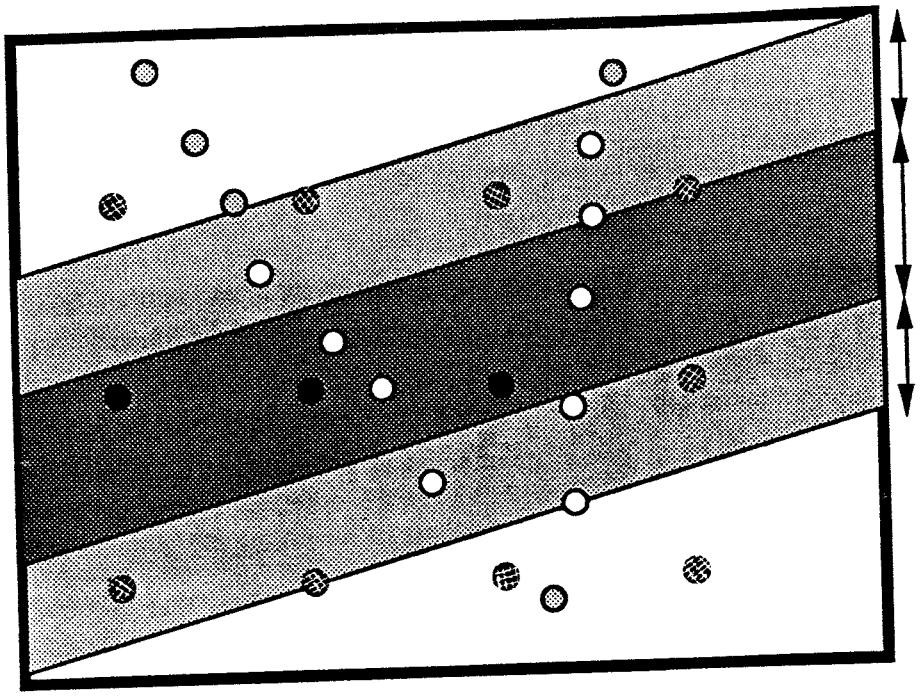


Figure 5.2.1 *Showing a kriging set and a corresponding kriging neighbourhood.*

- ● *Nodes inside the kriging set i.e.* $\in E(x_1)$.
- ◍ *Nodes outside the kriging set.*
- ◎ *Measurements inside the kriging neighbourhood i.e.* $\in D(x_1)$.
- ○ *Measurements outside the kriging neighbourhood.*

As for the parameter w it should not be too small in comparison to the correlation scale of the stochastic function considered. On the other hand it should not be too large in view of the constancy requirement.

5.3 Trends

If one does not believe that a zero:th order intrinsic model, even using kriging neighbourhoods, is sufficient for describing the phenomenon at hand one may want to include the expectation value functions explicitly in the analysis. In this work we will, for simplicity, confine ourselves to the use of linear models. That is we assume that we may write

$$m_Y(\mathbf{x}) = \phi(x)^T \beta \qquad 5.3.1$$

where β is a Mx1 vector of multiplicators and $\phi(\mathbf{x})$ is a Mx1 vector of basis functions. The choices of the basis functions is done based on examination of the general appearance of measurement data. The following basis functions will be utilized in the different trend models:

A constant trend $\phi_i(\mathbf{x}) = 1$.

Exponential type trend function $\phi_i(\mathbf{x}) = \mathbf{x}\cdot\mathbf{z}$ where $\mathbf{z}$ is unit vector pointing downward.

Potential type trend function $\phi_i(\mathbf{x}) = \log(c + \mathbf{x}\cdot\mathbf{z})$ where c is a given constant.

Characteristic trend function $\phi_i(\mathbf{x}) = I(\mathbf{x} \in \text{zone } 2)$ where I is an indicator function that equals one if the condition in parenthesis is fulfilled, zero if not. In Appendix E we discuss the geometrical characteristics of zone 2.

The reason for the names exponential type and potential type is obvious since they are models for the log conductivity. See sections 9.4 and 9.5.

Denoting the data support by

$$\{\mathbf{x}_i\}_{i=1}^{N}$$

and replacing the functions defined above by vectors containing the values at the data points as components we write 5.3.1 as

$$\mathbf{m}_Y = \mathbf{X}\beta \qquad 5.3.2$$

where $\mathbf{X}$ is a NxM matrix which we name the trend matrix. The columns of $\mathbf{X}$ are the basis functions above evaluated at the data points. It is essential for the regression analysis which follows that the columns are linearly independent.

6 STATISTICAL INFERENCE

6.1 Non-parametric variogram estimator

First we define a spherical coordinate system for the lagspace in terms of a given basis system {x^, y^, z^} where z^ is directed upward by referring to figure 6.1.1.

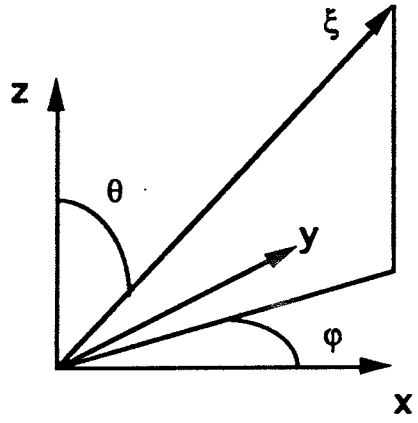


Figure 6.1.1 Defining a spherical coordinate system for the lag space.

Secondly, in order to estimate a semivariogram, we divide the lagspace into lag classes $L_{i,j}$ defined by

$$L_{i,j} = \left\{ \xi : r_{i-1}^2 \le |\xi|^2 \le r_i^2, \tan\left(\tfrac{\pi}{2} - \theta_j\right) \le \frac{(\hat{\xi})^2}{|\xi|^2 - (\hat{\xi})^2} \le \tan\left(\tfrac{\pi}{2} - \theta_{j-1}\right) \right\}$$

$$1 \le i \le R, 1 \le j \le \Theta \qquad 6.1.1$$

where r_i are the radial lag class delimiters, θ_i are the angular lag class delimiters ordered in an increasing sequence, R is the number of radial lag classes and Θ is the number of angular lag classes. Since we define θ_i to be an increasing sequence the ordering of the angular lag classes as defined by 6.1.1 is from the most horizontal to the most vertical.

Since the semivariogram is symmetric[13] it is only necessary to let θ vary from 0 to $\pi/2$ thus the lag class delimiters always satisfy $\theta_1 = 0$ and $\theta_\Theta = \pi/2$, also trivially r_0 is always chosen equal to zero. Note in particular that there is no radial lag class defined by the sole condition of being greater than r_R. The lag space could also be divided in the φ -direction as defined by figure 6.1.1, i.e. the angle in the horizontal plane, as well. However on the interest of simplicity this has not been done, see also the next section.

As a nonparametric semivariogram estimator we use

---

[13]That is $\gamma(\xi) = \gamma(-\xi)$ which implies $\gamma(r,\theta) = \gamma(r,-\theta)$ with θ as defined by figure «nve.A».

$$\Gamma_{i,j} = \frac{1}{2|L_{i,j}|} \sum_{\xi \in L_{i,j}} [Y(x+\xi) - Y(x)]^2$$

where the sum for each i, j is taken over all pairs of measurement points such that the vector connecting them belongs to $L_{i,j}$ and $|\cdot|$ denotes as always when applied to a set the number of points in (or more general, size of) the set. Note that lags between different holes participate as well as lags between points in a single hole. Each estimate $\Gamma_{i,j}$ is interpreted as the value of the semivariogram at the centroid of the lag class. This centroid is defined by

$$r_{i,j} = \int_{L_{i,j}} r dV \Big/ \int_{L_{i,j}} dV \qquad 6.1.2$$

and

$$\theta_{i,j} = \int_{L_{i,j}} \theta dV \Big/ \int_{L_{i,j}} dV \quad . \qquad 6.1.3$$

In the case that the lag class is given in spherical coordinates by 6.1.1 this is expressed as

$$r_{i,j} = \frac{3(r_i^4 - r_{i-1}^4)}{4(r_i^3 - r_{i-1}^3)}$$

and

$$\theta_{i,j} = \frac{-\theta_j \cos\theta_j + \theta_{j-1}\cos\theta_{j-1} + \sin\theta_j - \sin\theta_{j-1}}{\cos\theta_{j-1} - \cos\theta_j} \quad .$$

Other nonparametric estimators, mentioned in [Russo and Jury, 1987], are developed in [Cressie and Hawkins, 1980] and [Omre, 1984].

6.1.1 Fitting of variogram models

From the discussion in section 5.1 it follows in general that our semivariogram models can be written $\gamma(V, G, \xi)$ where V is the variance, G is the matrix of geometric anisotropy and ξ , as always, denotes the lag vector. In particular the anisotropic exponential variogram model is

$$\gamma(\xi) = V\left(1 - \exp\left(-\sqrt{\lambda_p^2 \cdot [\xi_1^2 + \xi_2^2] + \lambda_z^2 \cdot \xi_3^2}\right)\right) \quad . \qquad 6.1.1.1$$

The fitting of such a model to the nonparametric semivariogram estimates $\Gamma_{i,j}$ is done by minimizing the weighted square norm

$$\min_{V,G} \sum_{i,j} W_{i,j} (\gamma(V, G, \xi_{i,j}) - \Gamma_{i,j})^2 \quad . \qquad 6.1.1.2$$

Here $\xi_{i,j}$ are the representative point of the lag class i, j as defined by the averages 6.1.2 and 6.1.3, $W_{i,j}$ are the weights of each estimated semivariogram value $\Gamma_{i,j}$ and is set equal to the number of lags participating in the lag class i, j.

The motivation for this choice is that roughly this situation is equivalent to weighted regression where one minimize

$$(\Gamma - \gamma(V, G))^T C^{-1} (\Gamma - \gamma(V, G)).$$

Here Γ is the vector $\{\Gamma_{i,j}\}_{i,j}$, $\gamma(V, G)$ is the vector $\{\gamma(V, G, \xi_{i,j})\}_{i,j}$ and C is the covariance matrix of Γ i.e. $C = E[\Gamma\Gamma^T]$. Now the variance of $\Gamma_{i,j}$, that is a diagonal term of C, should be proportional to $1/N_{i,j}$ if the $\Gamma_{i,j}$ were independent.

The solution to the problem posed by 6.1.1.2 is obtained by the use of a conjugate gradient method as described in [Press et al, 1989, chap. 10].

It is to be pointed out here that the method described here has a number of disadvantages which unfortunately was not understood at the time of implementation. The use of lag classes certainly speeds up the calculation of the target function and its gradient but it makes the analysis somewhat arbitrary since the optimum depends on the choice of the lag classes and also on the choice of the weights and representative point for each lag class. A better approach would be to minimize

$$\min_{V,G} \sum_{(i,j)} (\gamma(V, G, x_i - x_j) - Y(x_i)Y(x_j))^2$$

as suggested in [Kitanidis, 1983], where x_i are the measurement locations and the sum is over all unique pairs.

6.2 Regression

This section contains a summary of our needs from the theory of regression analysis. The material is certainly contained in [Draper and Smith, 1966] or [Placket, 1960]. It is inserted for easy access of the reader and a reader familiar with the material could skip the section.

6.2.1 Normal equations

We will assume throughout that Y = log(K) is a Gaussian process. This assumption paves the way for deriving the normal equations by maximum likelihood. In fact let **Y** denote the stochastic vector obtained by restricting the stochastic function Y to all measurement points $\{x_i\}_{i=1,\ldots N}$ and let **y** denote our realization of **Y** then the "probability for the outcome" **y** for a given parameter vector β, the so-called likelihood function $L(\mathbf{y}, \beta)$, is

$$L(\mathbf{y},\beta) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} \exp\left(-\frac{(\mathbf{y}-\mathbf{X}\beta)^T \mathbf{C}^{-1}(\mathbf{y}-\mathbf{X}\beta)}{2}\right)$$

differentiating with respect to β and setting the deriviative equal zero we obtain directly

$$\mathbf{X}^T \mathbf{C}^{-1}(\mathbf{y}-\mathbf{X}\beta) = 0.$$

or

$$\mathbf{X}^T \mathbf{C}^{-1}\mathbf{X}\beta = \mathbf{X}^T \mathbf{C}^{-1}\mathbf{y} \qquad 6.2.1.1$$

with solution

$$\hat{\beta} = (\mathbf{X}^T \mathbf{C}^{-1}\mathbf{X})^{-1} \mathbf{X}^T \mathbf{C}^{-1}\mathbf{y} \qquad 6.2.1.2$$

As an alternative, distribution-free derivation we show that the estimator 6.2.1.2 is the unbiased estimator of β with the least variance. This is known as Gauss theorem and may be shown as follows.

Assume that

$$\hat{\beta} = \mathbf{Z}\mathbf{Y}$$

is a linear unbiased estimator of β where β, the process Y and the data support is connected by the equation

$$E[\mathbf{Y}] = \mathbf{X}\beta. \qquad 6.2.1.3$$

The requirement for nonbiasedness is written as

$$E[\hat{\beta}] = \mathbf{Z}E[\mathbf{Y}] = \beta$$

i.e. using 6.2.1.3

$$\mathbf{Z}\mathbf{X}\beta = \beta$$

or better, since we want the above equality to hold for all β,

$$\mathbf{Z}\mathbf{X} = \mathbf{I}_M$$

where $\mathbf{I}_M$ denotes the M by M unit matrix.

Moreover the centered variance of the estimator is

$$tr\, E\left[(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])^T\right] = tr\,(\mathbf{Z}\mathbf{C}_Y\mathbf{Z}^T)$$

where tr(·) denotes the trace operator. [14]

Now minimizing this variance taking the requirement of the nonbiasedness into account amounts to finding a stationary point to the Langrangian

$$L(\mathbf{Z},\mu) = tr\left(\mathbf{Z}\mathbf{C}_Y\mathbf{Z}^T\right) + (\mathbf{Z}\mathbf{X} - \mathbf{I}_M) \bullet \mu$$

where now dot stands for outer matrix (tensor) product. [15]

Differentiating this we obtain[16]

$$\begin{cases} \dfrac{\partial L(\mathbf{Z},\mu)}{\partial \mathbf{Z}}(\mathbf{Z},\mu,\delta\mathbf{Z}) = 2tr\left(\delta\mathbf{Z}\mathbf{C}_Y\mathbf{Z}^T\right) + \delta\mathbf{Z}\mathbf{X} \bullet \mu = 0 \\ \dfrac{\partial L(\mathbf{Z},\mu)}{\partial \mu}(\mathbf{Z},\mu,\delta\mu) = (\mathbf{Z}\mathbf{X} - \mathbf{I}_M) \bullet \delta\mu = 0 \end{cases}$$

Now since

$$tr\left(\delta\mathbf{Z}\mathbf{C}_Y\mathbf{Z}^T\right) = \delta\mathbf{Z}\mathbf{C}_Y\mathbf{Z}^T \bullet \mathbf{I}_M$$

and, as is easily established, it is true for general matrices $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$ that

$$\mathbf{S}_1\mathbf{S}_2 \bullet \mathbf{S}_3 = \mathbf{S}_3\mathbf{S}_2^T \bullet \mathbf{S}_1$$

provided that both sides of the equation is defined, the stationarity condition translates into

$$\begin{cases} \dfrac{\partial L(\mathbf{Z},\mu)}{\partial \mathbf{Z}}(\mathbf{Z},\mu,\delta\mathbf{Z}) = \left(2\mathbf{Z}\mathbf{C}_Y + \mu\mathbf{X}^T\right) \bullet \delta\mathbf{Z} = 0 \\ \dfrac{\partial L(\mathbf{Z},\mu)}{\partial \mu}(\mathbf{Z},\mu,\delta\mu) = (\mathbf{Z}\mathbf{X} - \mathbf{I}_M) \bullet \delta\mu = 0 \end{cases}$$

The fact that $\delta\mathbf{Z}$ and $\delta\mu$ are arbitrary is now used to derive the system of equations

---

[14]The trace operator is defined as: Let S be a general NxN matrix then

$$tr(\mathbf{S}) = \sum_{i=1}^{N} S_{i,i}$$

[15]The outer matrix (tensor) product is defined as: Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two arbitrary NxM matrices then

$$\mathbf{S} = \sum_{i=1}^{N}\sum_{j=1}^{M} S_{1}{}_{i,j}\, S_{2}{}_{i,j}$$

[16]This is in the Frechet sense i.e the derivatives is interpreted as linear operators on $\delta\mathbf{Z}$ and $\delta\mu$ parameterized by $\mathbf{Z}$ and μ. For a nice treatment of Frechet derivatives see [Ladas and Lakshmikantham, 1972, chapter 1]

$$\begin{cases} 2ZC_Y + \mu X^T = 0 \\ ZX - I_M = 0 \end{cases}$$

for points of stationary for the Lagrangian.

If we multiply the first equation from the right with one half times $C_Y^{-1}X$ and subtracts the result from the second equation it follows that $\mu = -2(XC_Y^{-1}X)^{-1}$ and thus it follows that $Z = (XC_Y^{-1}X)^{-1}X^TC_Y^{-1}$ which is precisely 6.2.1.2 that is obtained in the derivation using the normal distributional assumption.

6.2.2 Variance of the parameters

Introduce the variance operator in terms of the expectation value operator as

$$V[\hat{\beta}'] = E\left[(\hat{\beta}' - E[\hat{\beta}'])(\hat{\beta}' - E[\hat{\beta}'])^T\right] = E\left[\hat{\beta}\hat{\beta}'^T\right]$$

where primed quantities always denotes the mean removed form of a stochastic entity. Since it follows from 6.2.1.2 that

$$\hat{\beta}' = (X^T V^{-1} X)^{-1} X^T V^{-1} y' \qquad 6.2.2.1$$

the variance matrix of the parameters is easily calculated as

$$E\left[\hat{\beta}\hat{\beta}'^T\right] = (X^T V^{-1} X)^{-1} \qquad 6.2.2.2$$

6.2.3 Variance of the estimate.

Since the estimator $\beta\hat{}$ is unbiased

$$E[\hat{Y}] = XE[\hat{\beta}]$$

and thus

$$V[\hat{Y}] = E\left[(\hat{Y} - E[\hat{Y}])(\hat{Y} - E[\hat{Y}])^T\right] = XE\left[\hat{\beta}\hat{\beta}^T\right]X^T$$

so by 6.2.2.2 we may write

$$V[\hat{Y}] = X(X^T V^{-1} X)^{-1} X^T .$$

6.2.4 Confidence ellipsoid

First we note that the parameter vector $\beta\hat{}$ is Gaussian since it is linearly related to Y by 6.2.2.1. Hence in particular any component of $\beta\hat{}$, $\beta\hat{}_i$, is normally distributed with expectation value $E[\beta\hat{}]_i$ and variance $V[\beta\hat{}]_{i,i}$. Invoking standard mathematical statistics we derive a $100(1 - \alpha)$ confidence ellipsoid for the vector $\beta\hat{}$ as

$$(\hat{\beta}-\beta)^T \mathbf{X}\mathbf{C}^{-1}\mathbf{X}^T(\hat{\beta}-\beta) \leq$$

$$\left[\frac{N}{N-M}\right]\left(\mathbf{y}^T\mathbf{C}^{-1}\mathbf{y} - \hat{\beta}\mathbf{X}\mathbf{C}^{-1}\mathbf{y}\right)F(M, n-M, 1-\alpha)$$

where as usual N is the number of measurements, M is the number of parameters used to describe the trend function and F(M, N-M, ·) is an F distribution with (M, N-M) degrees of freedom. This result is stated, but not proved, in [Draper and Smith, p 108]. A more comprehensive derivation of these results are found in [Plackett, 1960].

6.3 Iterative Generalized Least Squares Estimation

Let us write the logarithm of the hydraulic conductivity Y as

$$Y(\mathbf{x}) = m(\mathbf{x}) + Y'(\mathbf{x})$$

where m(**x**) again denotes the deterministic drift component i.e.

$$m_Y(\mathbf{x}) = E[Y(\mathbf{x})]$$

and primed quantities as always signify mean removed form. Assume in the following that the residual Y′ is a second-order stationary stochastic function, see 3.1.

If the drift component m(**x**) were known, we would be able to retrieve structural information and estimate Y at unmeasured locations using residual kriging and the residual variogram 3.2.1

$$\gamma'_Y(\xi) = \tfrac{1}{2}\mathrm{E}\left[Y'(\mathbf{x}+\xi) - Y'(\mathbf{x})\right]^2$$

or analogously the covariance function. Since this is not readily available one could use the variogram of Y $\gamma_Y(\mathbf{x}, \xi)$ as an approximation of $\gamma_{Y'}(\xi)$. This leads to the error

$$\gamma_Y(\mathbf{x}, \xi) - \gamma'_Y(\xi) = \tfrac{1}{2}[m(\mathbf{x}+\xi) - m(\mathbf{x})]^2$$

as easily verified by expanding $\gamma_Y(\mathbf{x}, \xi)$ and substituting $\gamma_Y(\xi)$. If this error is small in comparison with the variance of Y′ for points in the lagspace **x** such that $|\mathbf{x}|$ is less than the correlation scale of Y′ we could use the local stationarity assumption of section 4.2. If this is not the case one should try to estimate the trend.

Thus we write the trend using a linear model as

$$\mathbf{m}_Y(\mathbf{x}) = \mathbf{X}\beta$$

where now the vector $\mathbf{m}_Y(\mathbf{x})$ signifies the Nx1 vector obtained by evaluating the function m(**x**) at all measurement locations the remaining notation has previously been explained in section 6.1.

As explained in section 6.2.1 the maximal likelihood or the minimum variance equations for β are

$$\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X}\beta = \mathbf{X}^T \mathbf{C}^{-1} \mathbf{y} \quad 6.3.1$$

In the most common case of regression the assumption is that the residuals are independent. This would in the above equation correspond to the case that

$$\mathbf{C} = \sigma^2 I \quad 6.3.2$$

where **I** is the NxN unity matrix. In our case however there is certainly no ground for such an assumption, since on the contrary we plan to use the covariance function of Y' for simulating the process. To use the assumption expressed in 6.3.2 and then estimate the covariance function of the residual for subsequent use for simulation purposes is self contradictory.

Since neither the drift nor the true covariance function is known this calls for an iterative procedure. Starting with an initial estimation and fitting of the semivariogram we solve the system 6.3.1. We then calculate the residuals and estimate the corresponding residual semivariogram which is then used to recalculate the drift components β etc. The iterative procedure is halted when the difference in the regression parameters β in two successive iterates is tolerably small. This method has previously been employed in for instance [Neuman and Jacobsson , 1984] and [Lovius et al. , 1990]

7 JACKNIFING

As discussed above in section 5 there are many possible models for a stochastic function meant to explain the pattern of the observed conductivities. After the choice of model has been made, methods for estimating the model parameters, as for instance those described in chapter 6, are employed. These methods are constructed in order to produce, in some sense, a best estimate of the model parameters. In simple cases, always using an a priori assumption on the distribution on the sample itself, one may derive the distribution of the estimator and thus produce confidence intervals etc. see section 6.2.4 for example. If the situation is to complicated for the actual derivation of the distribution of the estimator, simulation or distribution free methods such as the Chebyshev inequality may be used.

Such a procedure however does not tell you if the a priori assumptions are any good, see for instance [Draper and Smith, p. 41]. If one is interested testing these a priori assumptions or in finding the best model it is imperative to have a way to say " Model A is no good" or similarly a way to order models by statements of the kind "Model A is better than Model B". In the sequel we let model mean the a priori model together with some parameter estimates, for instance those obtained by the methods described in chapter 6. Note however that in the case of intrinsic functions the form of the kriging neighbourhoods constitutes an important model characteristic that cannot be estimated but only chosen and evaluated by the methods to be described.

Here a method known as cross-validation or jacknifing will be described. The objective is to measure the predictive capabilities or the "goodness" of a given stochastic model with given parameters . The technique is based on removing one sample point at a time and then to use the suggested model and the remaining sample points to predict the removed value with kriging. In this fashion one obtains a vector of kriging errors which can be analyzed statistically. The method is similar, at least as the objectives are concerned, to the examination of residuals in regression analysis see [Draper and Smith, p. 141–183].

Let us start by refreshing the notation used in section 4. Removing the I:th data point where $1 \le I \le N$ the kriging equations for predicting the value at the I:th data location becomes

$$\begin{cases} \sum_{j \in D(\mathbf{x}_I), j \neq I} \lambda_j(\mathbf{x}_I) C(\mathbf{x}_i - \mathbf{x}_j) + \mu(\mathbf{x}_I) = C(\mathbf{x}_i - \mathbf{x}_I) \qquad i \in D(\mathbf{x}_I), i \neq I \\ \sum_{i \in D(\mathbf{x}_I), j \neq I} \lambda_j(\mathbf{x}_I) = 1 \end{cases}$$

which leads to the kriging estimate

$$Y^*(\mathbf{x}_I) = \sum_{j \in D(\mathbf{x}_I), j \neq I} \lambda_j(\mathbf{x}_I) Y(\mathbf{x}_j)$$

and the kriging error

$$\varepsilon_I = Y(\mathbf{x}_I) - Y^*(\mathbf{x}_I) = Y(\mathbf{x}_I) - \sum_{j \in D(\mathbf{x}_I), j \neq I} \lambda_j(\mathbf{x}_I) Y(\mathbf{x}_j)$$

We note that including the nonbias condition above are strictly speaking unnecessary when the model includes a trend, since we then could use residual kriging. It is clear however that the inclusion of it does not affect the results of the jacknifing procedure. This is true since the assumptions done in deriving the kriging system above is certainly valid but of course the estimator is not the best since it would be improved by removing the nonbias condition.

From the section 4 it follows that the two first moments of the kriging error component ε_i are

$$E[\varepsilon_i] = 0 \qquad 7.1$$

and

$$E[\varepsilon_i^2] = C(0) - \mu(\mathbf{x}_I) - \sum_{i \in D(\mathbf{x}), j \neq I} \lambda_i(\mathbf{x}_I) C(\mathbf{x}_i - \mathbf{x}_I) \qquad 7.2$$

respectively. We stress again the fact that these results follow only if the a priori assumptions are correct. That is:

(i) The log conductivity process or residual process has the covariance function C(·).

(ii) The expectation value E[Y(**x**)] or E[Y´(**x**)] is constant in D(**x**).

Defining the reduced kriging error vector by

$$\hat{\varepsilon}_i = \frac{\varepsilon_i}{\sqrt{E[\varepsilon_i^2]}} \qquad i = 1, 2, .., N \qquad 7.3$$

we introduce the three jacknifing measures **mean reduced error, mean square reduced error** and **mean square error**

$$MRE = \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_i$$

$$MSRE = \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_i^2$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2$$

Assuming that the model is correct it follows from 7.1 and 7.2 that MRE ≈ 0 and MSRE ≈ 0 In fact we know that if the model is correct then

$$E[\hat{\varepsilon}_i] = 0 \qquad 7.4$$

and

$$E[\hat{\varepsilon}_i^2] = 1 \qquad 7.5$$

Now it would be tempting to assume that the kriging errors are normally distributed and perform a t-test, [Fisz, 1963, p. 348] to check the validity of the hypothesis that 7.4 and 7.5 holds. However since the kriging errors are not independent[17] this does not seem correct. This will be further discussed in chapter 8.

On the following pages some typical results from a jacknifing calculation are presented. Shown is the original measurements and the successive prediction errors when removing one point at a time. One notes that

(i) The errors are negligible on the flat portions of the conductivity plot.

(ii) There are large errors occurring in negative positive pairs at points where the conductivity plot shows rapid changes. This is very clear when studying figure 7.2.

This explains the fact, observed in chapter 9, that the jacknifing measure MSE is independent of the model choice since the sum of non reduced error is dominated by the very large errors in top graph of figure 7.2.

[17]That the kriging error cannot be uncorrelated in general is easily shown by a counter example.

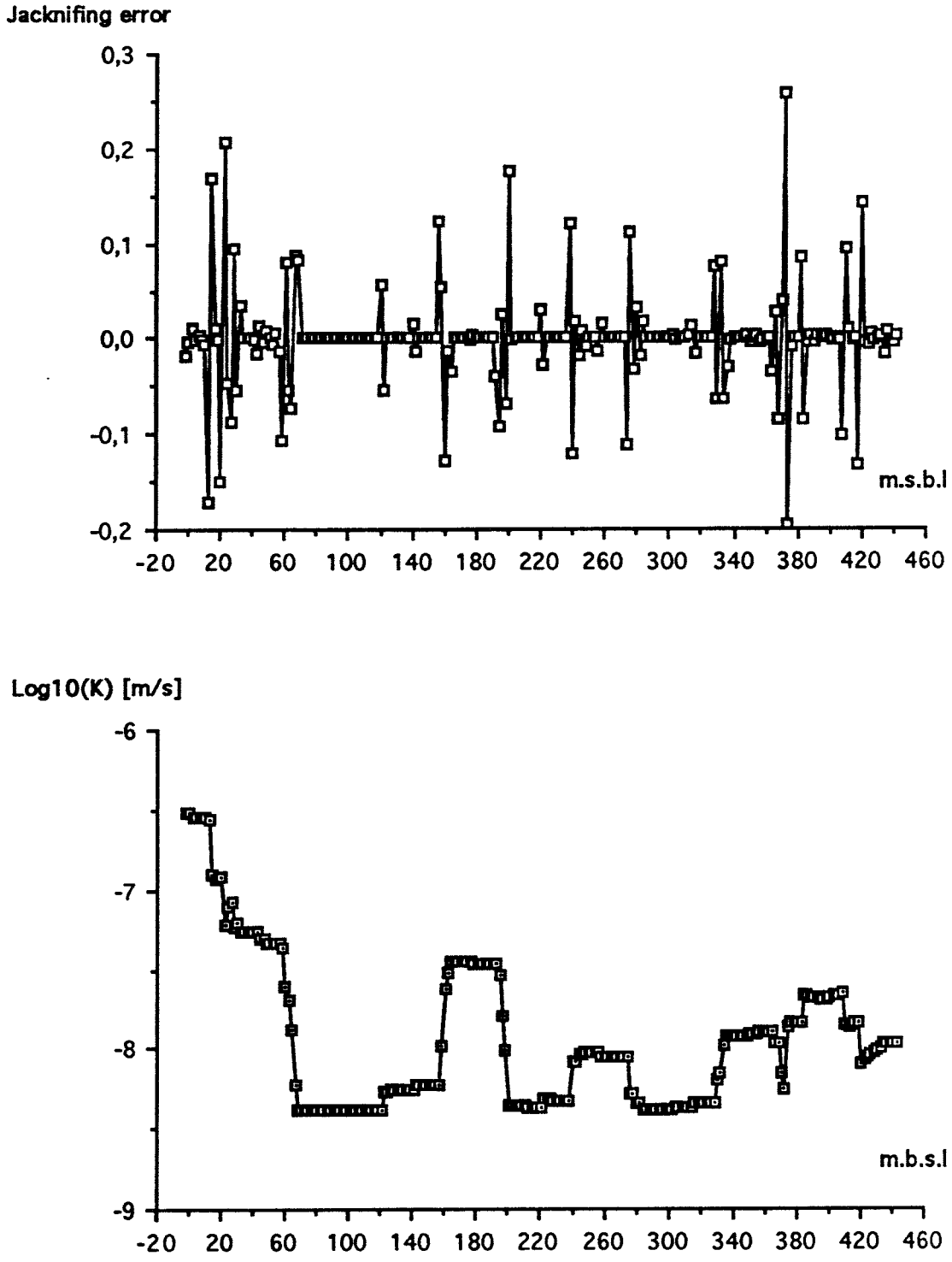


Figure 7.1 *Jacknifing errors and original regularized measurements in kfi01. The estimates are calculated using the exponential isotropic model in section 9.1.2 and measurements regularized to 36m from all boreholes.*

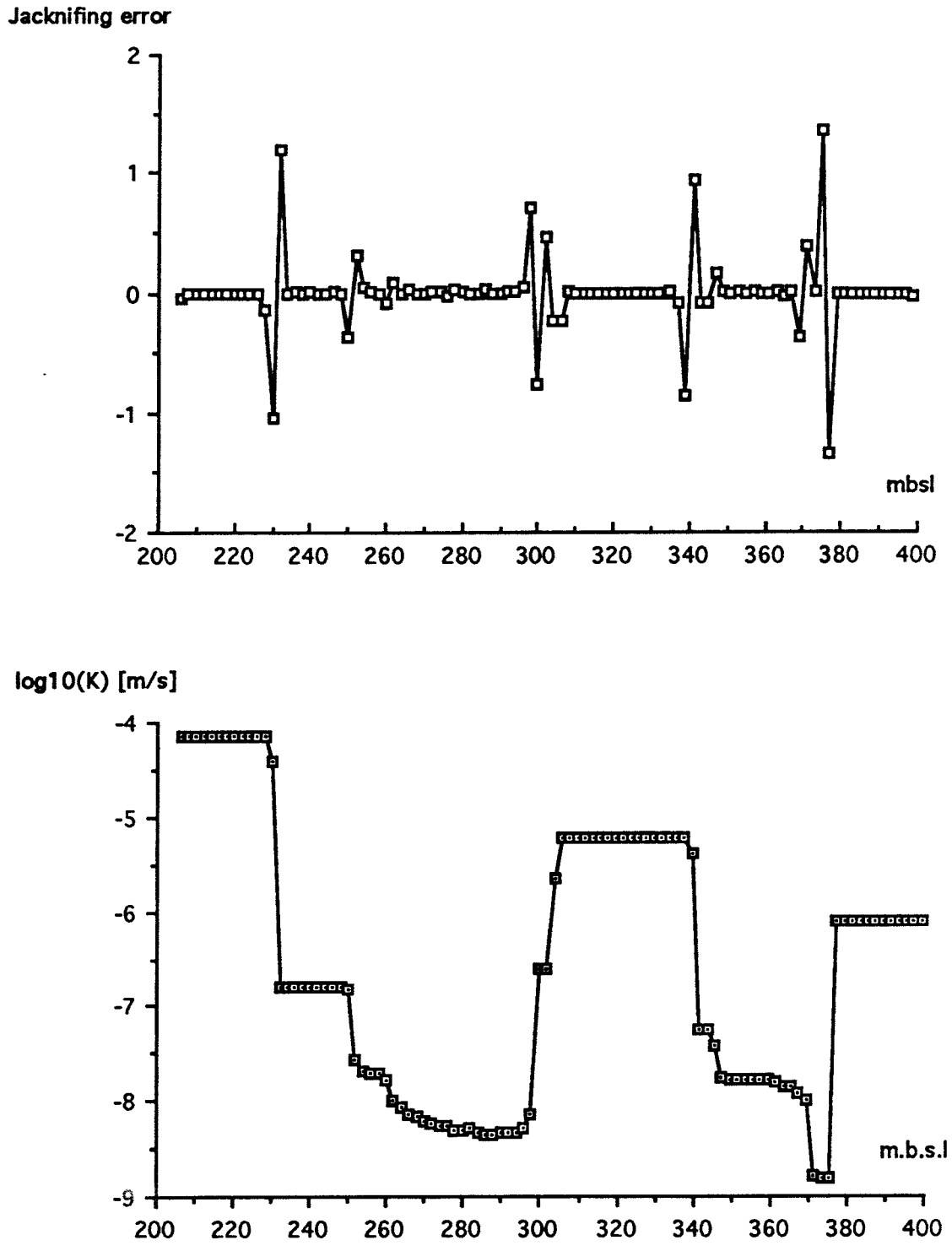


Figure 7.2 *Jacknifing errors and original regularized measurements in bfi01. The estimates are calculated using the exponential isotropic model in section 9.1.2 and measurements regularized to 36m from all boreholes.*

7.1 Calculation of the jacknifing measures using Woodbury's formula

The ordinary kriging system 4.2.5 can be written as

$$\mathbf{D}\begin{bmatrix} \lambda(\mathbf{x}) \\ \mu(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{c}(\mathbf{x}) \\ 1 \end{bmatrix} \qquad 7.1.1$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}$$

and where in turn **C** as usual denotes the NxN covariance matrix $\{C(\mathbf{x}_i - \mathbf{x}_j)\}_{i,j=1,N}$, $\mathbf{c}(\mathbf{x})$ is the vector $\{C(\mathbf{x}_i - \mathbf{x})\}_{i=1,N}$ and **1** is the Nx1 vector with all elements equal to one.

As previously stated what we want to do is to remove one data point , the I:th say where $1 \le I \le N$, and thus solve 7.1.1 with

$$\mathbf{C} = \{C(\mathbf{x}_i - \mathbf{x}_j)\}_{i,j \neq I}$$

$$\mathbf{c}(\mathbf{x}_I) = \{C(\mathbf{x}_i - \mathbf{x}_I)\}_{i \neq I}.$$

Clearly this corresponds to replacing the I:th row and column in **D** with zeros leaving an arbitrary nonzero element at the location (I, I), and neglecting the value of λ_I . We will show that this is in fact a two dimensional perturbation of **D**. Let us denote the I:th column of **D** by $\mathbf{d}_I$. By symmetry the I:th row of **D** is equal to $\mathbf{d}_I^T$. Denoting the I:th (N+1)x1 unit vector $\mathbf{e}_I$, and letting p stand for an arbitrary nonzero number, it is clear that the matrix

$$\mathbf{P} = \mathbf{e}_I \mathbf{d}_I^T - \mathbf{e}_I(p + d_{k,k})\mathbf{e}_I^T + \mathbf{d}_I \mathbf{e}_I^T$$

has at most two dimensional rank since it may be written

$$\mathbf{P} = \mathbf{U}\mathbf{V}^T = [\mathbf{e}_I, \mathbf{d}_I][\mathbf{d}_I - (p + d_{I,I})\mathbf{e}_I, \mathbf{e}_I]^T$$

and that the I:th row and column of the matrix **D** - **P** has only one non-zero element namely the diagonal element which is equal to p. [18] Thus the jacknifing equation can be solved using Woodbury's formula, see appendix D. Denoting $\mathbf{D}^{-1}$ with **B** and the columns of **B** with $\mathbf{b}_I$ we have trivially

$$\mathbf{B}\mathbf{d}_I = \mathbf{e}_I$$

so that

---

[18] We note that the perturbation is one dimensional in the particular case $d_I = \sigma^2 e_I$.

$$(\mathbf{V}^T\mathbf{BU})_{1,1} = \mathbf{d}_I{}^T\mathbf{Be}_I - (p + d_{I,I})\mathbf{e}_I{}^T\mathbf{Be}_I = 1 - (p + d_{I,I})b_{I,I},$$

$$(\mathbf{V}^T\mathbf{BU})_{1,2} = \mathbf{d}_I{}^T\mathbf{Bd}_I - (p + d_{I,I})\mathbf{e}_I{}^T\mathbf{Bd}_I = d_{I,I} - (p + d_{I,I}) = -p,$$

$$(\mathbf{V}^T\mathbf{BU})_{2,1} = \mathbf{e}_I{}^T\mathbf{Be}_I = b_{I,I},$$

and

$$(\mathbf{V}^T\mathbf{BU})_{2,2} = \mathbf{e}_I{}^T\mathbf{Bd}_I = 1$$

and thus

$$\mathbf{I} - \mathbf{V}^T\mathbf{D}^{-1}\mathbf{U} = \begin{bmatrix} (p + d_{I,I}) & b_{I,I} & p \\ -b_{I,I} & & 0 \end{bmatrix}.$$

This matrix is nonsingular if $b_{I,I}$ is non zero since the number p is always chosen to be non-zero.

Now, we derive the solution of the original kriging system 7.1.1 when $\mathbf{x} = \mathbf{x}_I$ is $\mathbf{e}_k$. From Woodbury's formula D.2, the kriging weights λ_I for determining the kriging estimate at the measurement point $\mathbf{x}_I$ are

$$\lambda_I = \mathbf{e}_I + \mathbf{D}^{-1}\mathbf{U}(\mathbf{I} - \mathbf{V}^T\mathbf{D}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{e}_I. \qquad 7.1.2$$

Now by Cramer's rule

$$(\mathbf{I} - \mathbf{V}^T\mathbf{D}^{-1}\mathbf{U})^{-1} = \frac{1}{pb_{I,I}}\begin{bmatrix} 0 & -p \\ b_{I,I} & (p + d_{I,I})\ b_{I,I} \end{bmatrix}$$

so first

$$\mathbf{BU}(\mathbf{I} - \mathbf{V}^T\mathbf{D}^{-1}\mathbf{U})^{-1} = \left[\frac{1}{p}\mathbf{e}_I, -\frac{1}{b_{I,I}}\mathbf{b}_I + \left(1 + \frac{d_{I,I}}{p}\right)\mathbf{e}_I\right]$$

and second

$$\mathbf{V}^T\mathbf{e}_I = \begin{bmatrix} -p \\ 1 \end{bmatrix}$$

which together imply that

$$\mathbf{D}^{-1}\mathbf{U}(\mathbf{I} - \mathbf{V}^T\mathbf{D}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{e}_I = -\frac{1}{b_{I,I}}\mathbf{b}_I + \frac{d_{I,I}}{p}\mathbf{e}_I$$

so that from 7.1.2 we obtain

$$\lambda_I = -\frac{1}{b_{I,I}}\mathbf{b}_I + \left(1 + \frac{d_{I,I}}{p}\right)\mathbf{e}_I$$

and in particular,

$$\lambda_{I,I} = \frac{d_{I,I}}{p}$$

which we knew all along. The kriging weights are given by

$$\lambda_{I,i} = -\frac{1}{b_{I,I}}\mathbf{b}_{I,i} \qquad i \neq I$$

Now in order to calculate the jacknifing measures MRE and MSRE we need to calculate the normalized kriging error. That is we need to express the variance of the kriging error. From section 4 it follows that the variance of the kriging error in general is given by

$$E\left[\left(Y(\mathbf{x})^* - Y(\mathbf{x})\right)^2\right] = C(0) - \mu(\mathbf{x}) - \sum_{i \in D(\mathbf{x})} \lambda_i(\mathbf{x}) C(\mathbf{x}_i - \mathbf{x})$$

and thus with the current notation

$$E\left[\left(Y(\mathbf{x}_I)^* - Y(\mathbf{x}_I)\right)^2\right] = d_{I,I} - \sum_{i \neq I} \lambda_{I,i}\mathbf{d}_{I,i} = d_{I,I} + \frac{1}{b_{I,I}} \sum_{i \neq I} \mathbf{b}_{I,i}\mathbf{d}_{I,i} =$$

$$d_{I,I} + \frac{1}{b_{I,I}}(1 - d_{I,I}b_{I,I}) =$$

$$\frac{1}{b_{I,I}}.$$

8 PROVING MODEL CONSISTENCY USING JACKNIFING

As a preliminary step we derive the expected value and the variance of the MRE and MSRE, based on the simplifying assumption that the reduced errors are normally distributed and uncorrelated. Subsequent analysis without these assumptions show that these preliminary estimates are inaccurate, and therfore the following should be viewed only as a simplified demonstration of the analysis approach. The second part of this section gives more correct formulae.

Assume that our a priori model is correct, i.e. that the observed values of hydraulic conductivity together constitute a realization of the considered model. If we add the assumptions that the reduced kriging errors, as defined in 7.3, belongs to N(0, 1)[19] and that the error made when treating them as uncorrelated is not too large we have

$$MRE \in N\left(0, \frac{1}{\sqrt{N}}\right)$$

$$N * MSRE \in \chi^2(N)$$

where χ^2(N) indicates a χ^2-distribution with N degrees of freedom, N is as always the number of measurements. Since the number of measurements, in a geohydrological investigation in general, and at the Finnsjön site in particular, is quite large it is appropriate to use asymptotic theory. Thus it may be shown[20] , that as the number of degrees of freedom goes to infinity, the random variable $(2N * MSRE)^{1/2}$ has the asymptotic distribution $N((2N - 1)^{1/2}, 1)$. Thus we will henceforth assume that

$$MSRE^{1/2} \in N\left(\sqrt{1 - \frac{1}{2N}}, \frac{1}{\sqrt{2N}}\right).$$

In particular this leads to the result that

$$\frac{V[MRE]}{V[MSRE^{1/2}]} = 2.$$

Moreover the confidence intervals become independent of the model tested. This is an advantage over the simulation approach described below. Thus

---

[19]This clearly follows (by the linearity of the formula for the kriging error) from the original assumption that the stochastical function involved, that is in our case the logarithm of the hydraulic conductivity, is a Gaussian process.

[20][Fisz, 1963, p 342] gives the reference Fisher, R. A., Theory of statistical estimation, PCPS 22, 700 (1925).

$$MRE \in \left[-\frac{1.96}{\sqrt{N}}, \frac{1.96}{\sqrt{N}}\right]$$

$$MSRE^{1/2} \in \left[\sqrt{1-\frac{1}{2N}} - \frac{1.96}{\sqrt{2N}}, \sqrt{1-\frac{1}{2N}} + \frac{1.96}{\sqrt{2N}}\right]$$

are 95% confidence intervals for MRE and $MSRE^{1/2}$ respectively. In a large number of the calculations of the jacknifing measures shown in section 9 the number of measurements is N = 1815. In that case the confidence intervals are [-0.046, 0.046] for MRE and [0.967, 1.033] for $MSRE^{1/2}$.

Note that we cannot treat MSE in this way since its distribution is dependent on the tested model as well as the location of the data points.

If the above assumptions seems inadequate one may perform repeated simulation of a process corresponding to the examined model on the actual measurement points together with subsequent jacknifing. In that way we derive an experimental distribution for the jacknifing measures. Either the approximate analytical above or the experimental distribution, can then be used in conjunction with the actual jacknifing measures to test the hypothesis:

H0: The measurement values stem from the given stochastic model.

In principle this should be done by finding an event (domain) in the (MRE, MSRE, MSE) -space that has probability measure α. If our actual value of (MRE, MSRE, MSE) is outside this event we can reject the hypothesis H0 on the 1 - α level. However to regard the stochastic vector (MRE, MSRE, MSE) is too laborious and we settle for deriving the a and 1 - α quantiles of the distributions of MRE, MSRE MSE separately, as above, and reject the hypothesis H0 if some of the jacknifing measures lie outside these confidence intervals.

This idea could be applied for any measure whatsoever, or for instance for the sample vector itself or the model parameters. The point is however to find good measures as the above measures hopefully are. To regard the sample itself is obviously impractical .

A more down-to-earth reason for performing this analysis of the distribution of the jacknifing measures is of course that it is vital to determine what a mismatch in one of the jacknifing measures implies about the model inaccuracy. For instance say that two different models give the measures (MRE_1, $MSRE_1$, MSE_1) and (MRE_2, $MSRE_2$, MSE_2) where (say) $MSE_1 = MSE_2$ and

$$| MRE_1 | > | MRE_2 | > 0$$

implying that model 2 is the better but

$$| 1 - MRSE_2^{1/2} | > | 1 - MRSE_1^{1/2} | > 0$$

implying the converse. The straightforward solution is within reach if one knows the variances ($V[MRE_1]$, $V[MSRE_1^{1/2}]$, $V[MSE_1^{1/2}]$) and ($V[MRE_2]$, $V[MSRE_2^{1/2}]$, $V[MSE_2^{1/2}]$) of the measures because then a reasonable combined measure can be defined as for instance

$$J_i = \frac{|MRE_i|}{V[MRE]^{1/2}} + \frac{|1 - MSRE_i^{1/2}|}{V[MSRE^{1/2}]^{1/2}} + \frac{|MSE_i^{1/2} - E[MSE_i^{1/2}]|}{V[MSE_i^{1/2}]^{1/2}}$$

or more simply

$$J_i = \frac{|MRE_i|}{V[MRE]^{1/2}} + \frac{|1 - MSRE_i^{1/2}|}{V[MSRE^{1/2}]^{1/2}} \qquad 8.1$$

where we, in both of the formulas above, tacitly assumed that V[MRE] and V[MSRE] are independent of the model. Thus model 1 is taken to be better than model 2 if $J_1 < J_2$.

8.1 Simulations

The first requirement to fulfill in order to perform the task formulated above is to be able to simulate from all inferred models. This was not done because it was not deemed to be critical. Instead in order to get a grip on the situation simulations has been done for the intrinsic models only with the following results.

8.1.1 Spherical isotropic

Using the model in 9.1.1.1 we obtain after 100 realizations:

E[MRE]	=	0.230E-3
$V[MRE]^{1/2}$	=	0.139E-2
$E[MSRE^{1/2}]$	=	0.995
$V[MSRE^{1/2}]^{1/2}$	=	0.255E-1
$E[MSE^{1/2}]$	=	0.164
$V[MSE^{1/2}]^{1/2}$	=	0.439E-2

so that approximately

$$\frac{V[MRE]^{1/2}}{V[MSRE^{1/2}]^{1/2}} \approx \frac{1}{18.4}$$

Assuming that the jacknifing measures are normally distributed we get the following 95% confidence intervals.

MRE	$\in$	[-2.49E-3, 2.95E-3]
$MSRE^{1/2}$	$\in$	[0.945, 1.045]
$MSE^{1/2}$	$\in$	[0.155, 0.173]

8.1.2 Spherical anisotropic

Using the model in 9.1.1.2 we obtain after 100 realizations:

$V[MRE]^{1/2}$	=	0.131E-2
$E[MSRE^{1/2}]$	=	1.00
$V[MSRE^{1/2}]^{1/2}$	=	0.242E-1
$E[MSE^{1/2}]$	=	0.166
$V[MSE^{1/2}]^{1/2}$	=	0.460E-2

so that approximately

$$\frac{V[MRE]^{1/2}}{V[MSRE^{1/2}]^{1/2}} \approx \frac{1}{18.5}$$

Assuming that the jacknifing measures are normally distributed we get the following 95% confidence intervals.

MRE	$\in$	[-2.56E-3, 2.56E-3]
$MSRE^{1/2}$	$\in$	[0.953, 1.047]
$MSE^{1/2}$	$\in$	[0.157, 0.175]

8.1.3 Exponential isotropic

Using the model in 9.1.2.1 we obtain after 100 realizations:

E[MRE]	=	-0.285E-3
$V[MRE]^{1/2}$	=	0.172E-2
$E[MSRE^{1/2}]$	=	1.01
$V[MSRE^{1/2}]^{1/2}$	=	0.195E-1
$E[MSE^{1/2}]$	=	0.216
$V[MSE^{1/2}]^{1/2}$	=	0.480E-2

so that approximately

$$\frac{V[MRE]^{1/2}}{V[MSRE^{1/2}]^{1/2}} \approx \frac{1}{11,3}$$

Assuming that the jacknifing measures are normally distributed we get the following 95% confidence intervals.

MRE	$\in$	[-3.66E-3, 3.09E-3]
$MSRE^{1/2}$	$\in$	[0.972, 1.048]
$MSE^{1/2}$	$\in$	[0.207, 0.225]

8.1.4 Exponential anisotropic

Using the model in 9.1.2.2 we obtain after 100 realizations:

E[MRE]	=	0.216E-4
$V[MRE]^{1/2}$	=	0.140E-2
$E[MSRE^{1/2}]$	=	1.01
$V[MSRE^{1/2}]^{1/2}$	=	0.220E-1
$E[MSE^{1/2}]$	=	0.224
$V[MSE^{1/2}]^{1/2}$	=	0.546E-2

so that approximately

$$\frac{V[MRE]^{1/2}}{V[MSRE^{1/2}]^{1/2}} \approx \frac{1}{15.7}$$

Assuming that the jacknifing measures are normally distributed we get the following 95% confidence intervals.

MRE	$\in$	[-2.72E-3, 2.77E-3]
$MSRE^{1/2}$	$\in$	[0.967, 1.053]
$MSE^{1/2}$	$\in$	[0.213, 0.235]

8.2 Conclusions

Since the number of measurements in the above simulations are 1815 the theoretical value for $V[MSRE^{1/2}]^{1/2}$ equals 0.166E-1 which is lower than what is obtained in the simulations above but of the same order whereas the value for $V[MRE]^{1/2}$ in the simulations above is one order of magnitude to small compared to the theoretical value 0.235E-1. The only explanation for this fact at hand is that the treatment of the kriging errors as uncorrelated and normally distributed gives totally incorrect results. Thus the

analytical treatment should only serve as a model for the analysis. The use of a normal distribution to derive a confidence interval in the first part of this section is thus of no importance, since what will used below is only the approximative relation between $V[MRE]^{1/2}$ and $V[MSRE^{1/2}]^{1/2}$.

Since the distribution of MSE cannot be considered to be independent of the tested model, its value as a measure of the model fit is limited. Also, as explained previously, MSE does not vary among the models. Thus it seems consistent to choose

$$J = 15|MRE| + |1 - MSRE^{1/2}|$$

as a comparative measure between models analogously to 8.1. The value 15 is the approximate average of the values of $V[MSRE^{1/2}]^{1/2}/V[MRE]^{1/2}$ obtained above.

9 RESULTS

Since the objective of this work was to find the best model for the stochastic conductivity field at the Finnsjön site, it was necessary to test a lot of models. The results of this testing are shown below, with a section for each case. We stress that all results shown refer to measurements regularized to the 36m scale. Some miscellaneous results on other scales are found in appendix F.

As pointed out previously, some arbitrary choices have been included in the analysis. The most important of these are

Kriging neighbourhoods.

Weights used in the variogram fitting

The division into lag classes.

Regularization tolerances.

The sensitivity for some of theses choices are illustrated in section 9.2. It may again be pointed out that the division into lag classes is in some respects unfortunate, since it introduces arbitrariness into the analysis both directly and also indirectly through the choice of weights in the fitting, see section 6.1.

In all the cases (sections) below the following input was used, if the contrary is not explicitly stated.

Normal of the kriging neighbourhoods = (0, 0, -1).

Width of the kriging set = 20m.

Distance from the kriging set to the boundary of the kriging neighbourhood = 150m.

Minimal number of points in a kriging neighbourhood = 2.

Number of radial lag classes = 91.

Number of angular lag classes = 3 (in the anisotropic cases, in the isotropic one angular lag class was used.) The division was [0°, 30°], [30°, 70°], [70°, 90°]. Note that the angles are measured from the horizontal plane.

Weights in the variogram fitting: $1/n_{i,j}^{1/2}$ where $n_{i,j}$ is the number of lags in a lag class.

Positive regularization tolerance = 0.02.

Negative regularization tolerance = 0.041.

9.1 Intrinsic models

9.1.1 Spherical type variogram

9.1.1.1 Isotropic model

Fitted model:

Range	=	267.6
Variance	=	1.74

As previously stated in chapter 5.1 the range of a spherical model is defined as a, see 5.1.1 . In general the range will be the preferred way to present the results below instead of using the parameters a and λ.

Jacknifing measures:

MRE	=	1.80E-3
$MSRE^{1/2}$	=	1.24
$MSE^{1/2}$	=	0.1917
J	=	0.267

For a look at these variograms see figure 9.1.1.1.1.

9.1.1.2 Anisotropic model

Fitted model:

Variance	=	1.75
Horizontal range	=	307.7
Vertical range	=	257.8

jacknifing measures:

MRE	=	1.66E-3
$MSRE^{1/2}$	=	1.23
$MSE^{1/2}$	=	0.192
J	=	0.2549

For a look at these variograms see figure 9.1.1.2.1.

9.1.2 Exponential type models

9.1.2.1 Isotropic model

Fitted model:

(Practical) range	=	107.2 (321.6)
Variance	=	1.76

The range for a exponential model is defined, somewhat illogically but in correspondence with the spherical model as $1/\lambda$, using the notation of 5.1. This cannot be compared with the corresponding range for the spherical model. Instead to compare with the spherical range use the value inside the parenthesis which refers to the practical range i.e. in the notation of 5.1 $3/\lambda$ See figure 5.1.1.

Jacknifing measures:

MRE	=	1.27E-3
$MSRE^{1/2}$	=	0.957
$MSE^{1/2}$	=	0.192
J	=	0.062

For a look at these variograms see figure 9.1.2.1.1.

9.1.2.2 Anisotropic models

Fitted model:

Variance	=	1.88
(Practical) Horizontal range	=	377.3 (1132.1)
(Practical) Vertical range	=	94.3 (283.0)

As for the isotropic case the ranges are defined as $1/\lambda_h$ and $1/\lambda_v$ for the horizontal and vertical ranges respectively and the practical ranges three times these values. See section 5.1. jacknifing measures:

MRE	=	0.804E-3
$MSRE^{1/2}$	=	0.905
$MSE^{1/2}$	=	1.92
J	=	0.107

For a look at these variograms see figure 9.1.2.2.1.

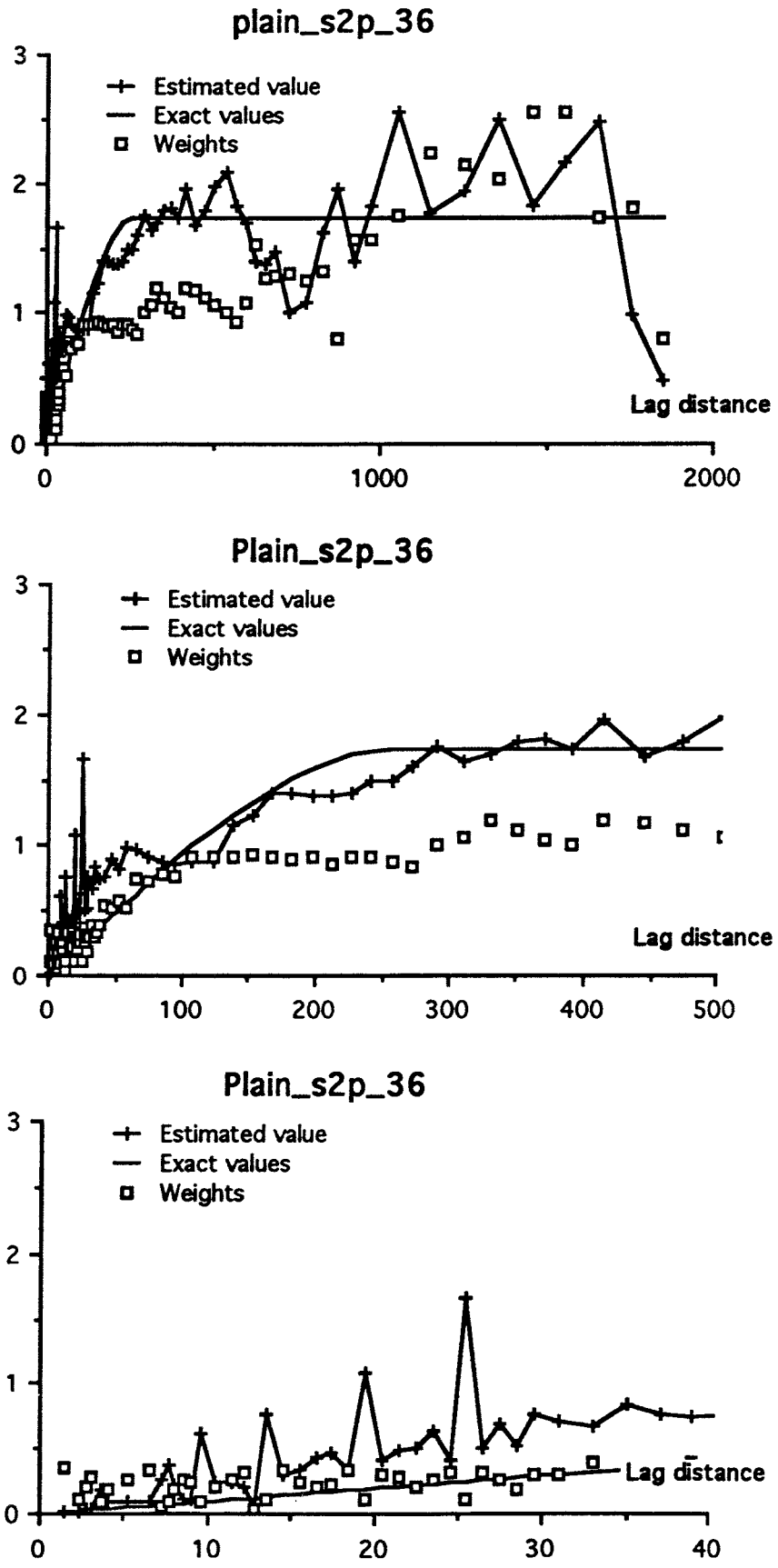


Figure 9.1.1.1.1 *Variogram fit in the spherical isotropic model. Note that the "peaks" corresponds to a low number of lags in the corresponding lag class.*

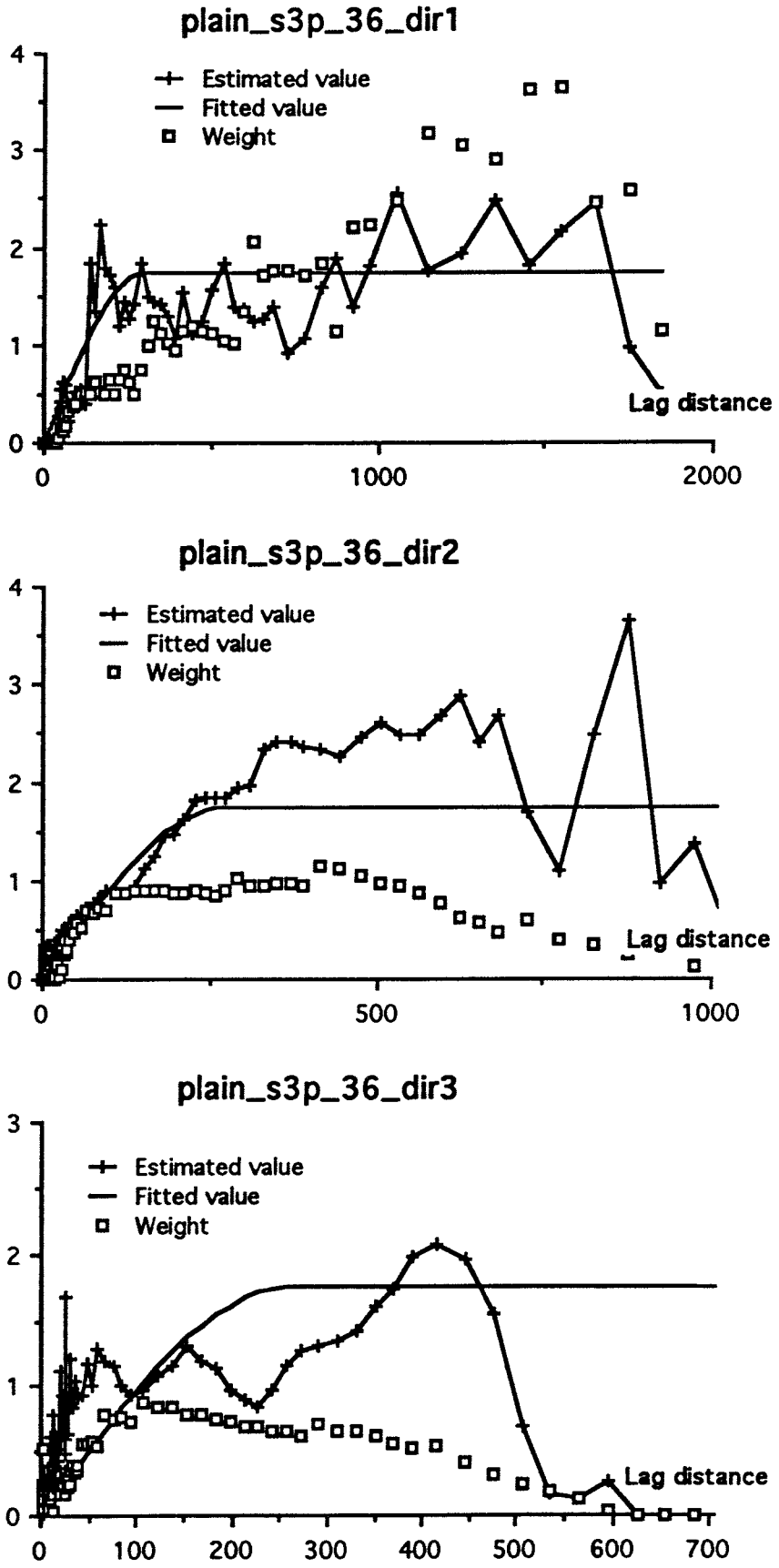


Figure 9.1.1.2.1 *Variogram fit in the spherical anisotropic model. The plots show the three different lagclasses dir1 = [0°, 30°), dir2 = [30°, 70°) and dir3 = [70°, 90°].*

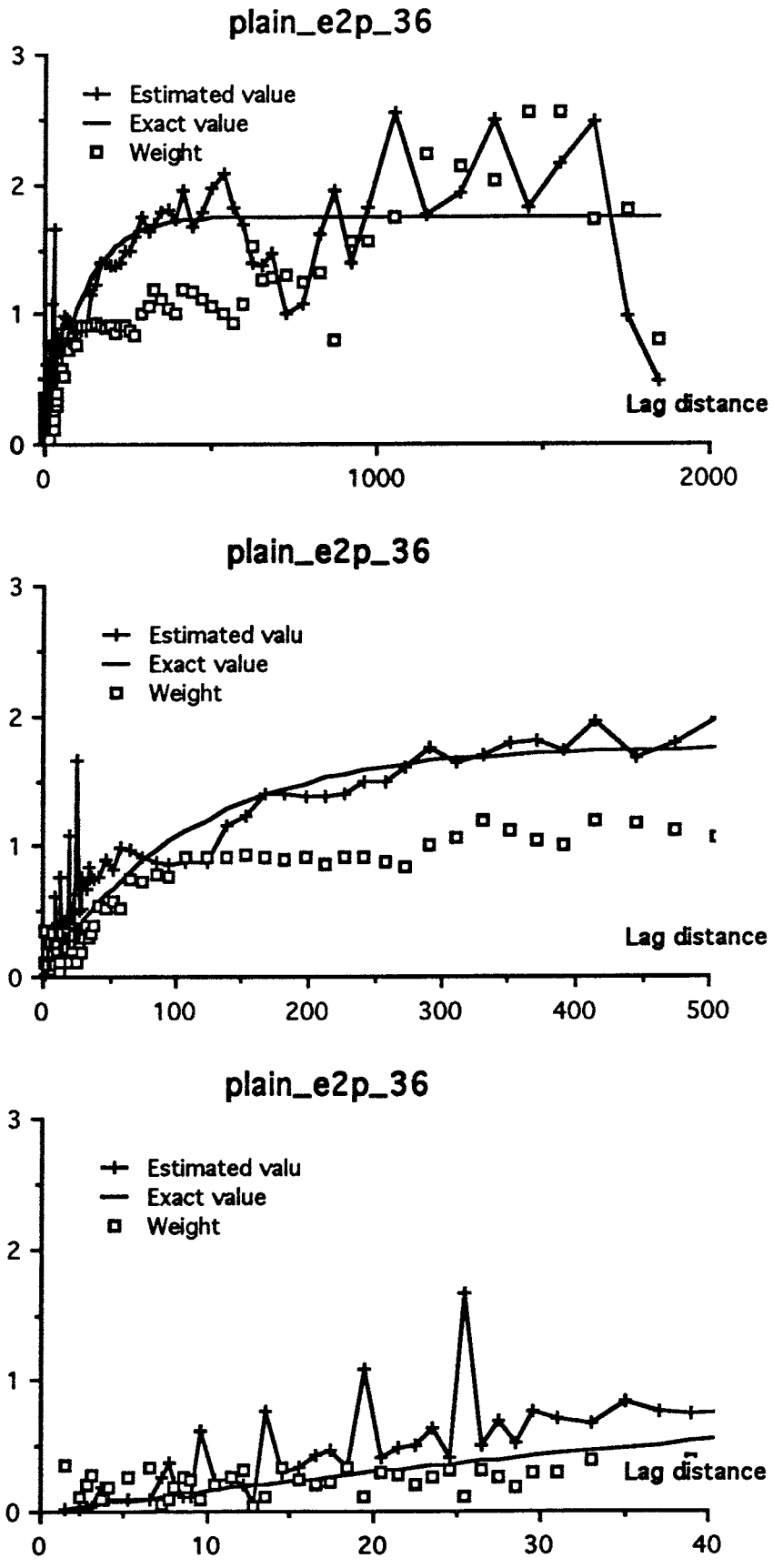


Figure 9.1.2.1.1 Variogram fit in the exponential isotropic model.

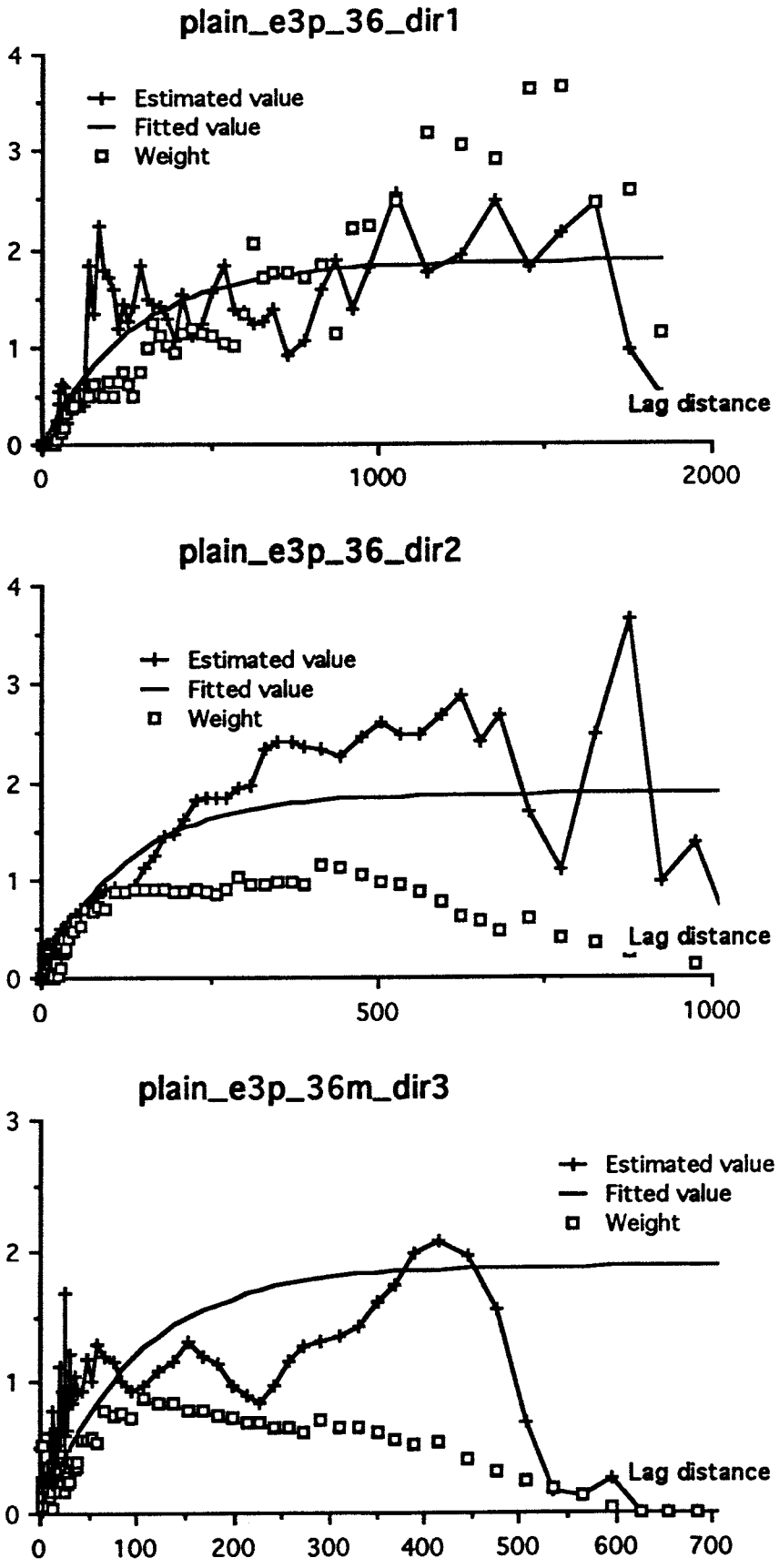


Figure 9.1.2.2.1 *Variogram fit in the exponential anisotropic model. The lagclasses as before.*

9.1.3 Conclusions for the intrinsic models

In the isotropic case the exponential model displays a much better fit than the spherical model. This is also reflected in the jacknifing measures. Note also that the peaks of the nonparametric variogram estimates at short lag distances is accompanied with rather low weights.

The calculated jacknifing measures together with the results obtained from the simulations shown in section 8 make it possible to reject all the intrinsic models on the 95% level.

The fitting of the anisotropic variogram is a delicate matter. The fitted spherical variogram shows almost no anisotropy at all whereas the fitted exponential variogram shows a anisotropy ratio of four. The horizontal range of exponential anisotropic variogram is approximately four times the range of the fitted isotropic variogram. Such a difference is likely to have a very large impact on the results of a simulation.

Since the fitting of a anisotropic variogram model includes any isotropic model the fit should be better. Nevertheless the jacknifing measures are worse. This shows that the process of finding the variogram as described in sections 6.1 and 6.1.1 does not lead to the best model from a jacknifing point of view. Unfortunately the idea to minimize the jacknifing measure with respect to the model parameters is untractable for large inference problems as the one described in this report.

9.2 Sensitivity

There are unfortunately some arbitrary choices involved in the results described above. First, and also most important, is the choice for kriging neighbourhoods.

9.2.1 Sensitivity for kriging neighbourhoods

Since the choice of kriging neighbourhoods is an important part of the model it seemed well worth an effort to try to find good choices of these. However variations performed using alternate inclination parallel to zone 2 and varying the width of the kriging neighbourhoods from 255m down to 70m do not give any significant change in the jacknifing measures. In fact the following results were obtained:

The input:

> Normal of the kriging neighbourhoods = (-0.239, -0.149, -1.00). This is the approximate normal of zone2. [Andersson et al, 1991]
>
> Width of the kriging set = 85m
>
> Distance from the kriging set to the boundary of the kriging neighbourhood = 85m.

Minimal number of points in a kriging neighbourhood = 32.

Tested model: Exponential isotropic, variance = 1.76, range = 107.2

Gave the output:

MRE	=	0.139E-2
$MSRE^{1/2}$	=	0.957
$MSE^{1/2}$	=	0.192
J	=	0.064

The input:

Normal of the kriging neighbourhoods = (0.0, 0.0, -1.00).

Width of the kriging set = 10m

Distance from the kriging set to the boundary of the kriging neighbourhood = 35m.

Minimal number of points in a kriging neighbourhood = 2.

Tested model: Exponential isotropic, variance = 1.76, range = 107.2

Gave the output:

MRE	=	0.131E-2
$MSRE^{1/2}$	=	0.957
$MSE^{1/2}$	=	0.191
J	=	0.063

9.2.2 Sensitivity for different fitting strategies

Because of the particular form of the data support i.e. the strong clustering of data in the boreholes, and the large information voids in between, one might suspect that what is important is the variogram fitting for short lags only. To check this assumption we performed the jacknifing analysis using the model that behaved the so far most favorable i.e. the isotropic exponential model with the fitting only taking the lags shorter than 130 m into account. As seen below in figure 9.2.2.1 this gives a nice fit for the variogram parameters

Variance	=	0.94	
(Practical) Range	=	25.3	(75.9)

but from a jacknifing point of view this model behaves poorly since the calculated jacknifing measures are

MRE	=	-0.145E-2
$MSRE^{1/2}$	=	0.638
$MSE^{1/2}$	=	0.195
J	=	0.384.

The kriging neighbourhoods used for this result was as previously 320m wide, which is perhaps too wide, so we repeated the analysis with the width instead being 160m which yielded

MRE	=	-0.387E-3
$MSRE^{1/2}$	=	0.638
$MSE^{1/2}$	=	0.193
J	=	0.369.

i.e. no substantial improvement.

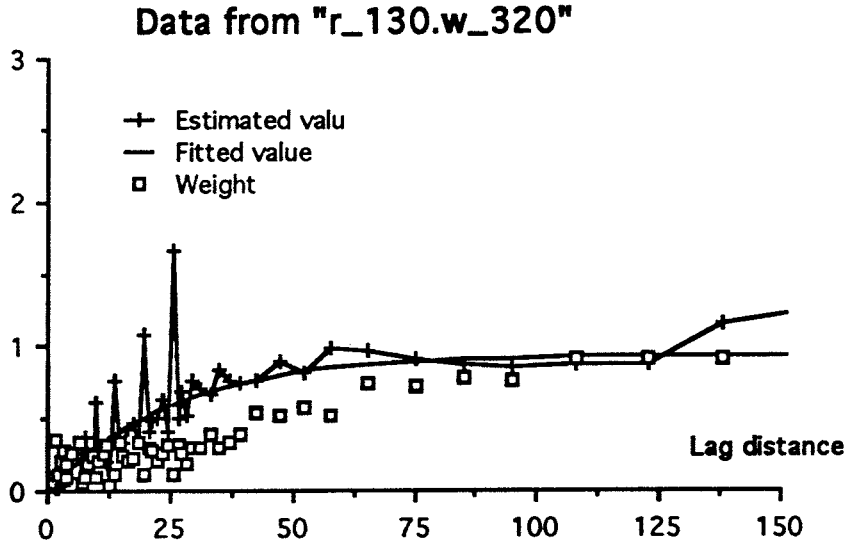


Figure 9.2.2.1 Variogram fit using only the lags up to 130 m.

Thus it seems as if the advantage of having a close fit at the short lags is totally nullified by not having the possibility of representing the true variance of the stochastic function. Thus the obvious generalization of the above study would be to introduce nested models. [Journel and Huijbregts, 1978].

9.2.3 Sensitivity for lag class division

Using five lag classes each having an opening angle of 18° the results changed drastically compared with the results in section 9.1.2.2. We obtained the fitted model

Variance	=	2.14
(Practical) Horizontal range	=	2290.0
(Practical)Vertical range	=	354.0

which corresponds to a anisotropy ration of 6.5. These results further underline the difficulties in estimating the correlation lengths for anisotropic covariances and the large arbitrariness introduced by the introduction of lag classes.

9.2.4 Conclusions

It is dissapointing to conclude from section 9.2.1 that the the jacknifing measures are insensitive to the choice of kriging neighbourhoods. This may be due to the fact that the measurements are so strongly clustered and that thus the estimation for each removed measurement is dominated by its nearest neighbours. A possible remedy would be to remove a larger set of measurements.

From section 9.2.2 we point out that the type of fitting shown in figure 9.2.2.1 is what one would risk being close to if one restricted the lags to interhole instead of using also intrahole lagvectors. As the corresponding jacknifing measures show this results in a worse model.

Finally we note again that it is difficult to estimate correlation lengths in the anisotropic case and that the estimates are very sensitive to the lagclass division.

9.3 Residual models

9.3.1 Potential type trend function

The trend function is of type

$$\log(K) = \beta_0 - \beta_1 \log(\beta_3 + z) + \beta_2 I\{(x, y, z) \in zone\ 2\}$$

with z being positive downward (m. b. s. l). Alternatively, which is the reason for the name of the trend type,

$$K = \begin{cases} b_0 b_2 (b_3 + z)^{-b_1} & \text{if } (x, y, z) \in zone\ 2 \\ b_0 (b_3 + z)^{-b_1} & \text{if } (x, y, z) \notin zone\ 2 \end{cases}$$

with

$$\begin{cases} b_0 = 10^{\beta_0} \\ b_1 = \beta_1 \\ b_2 = 10^{\beta_2} \\ b_3 = \beta_3 \end{cases}.$$

Note carefully however that we fit the function E[logK] to the measured values $\log(K_i)$ and that this is not equivalent with fitting the model for E[logK] to the measured values K_i.

The regression performed with this trend function is not complete in that the parameter β_3 is set to the value 40.0 which is approximately the least value to keep ($\beta_3 + z$) strictly greater than zero. The inclusion of β_3 as an ordinary regression parameter would lead to a nonlinear regression problem. This generalization is of course highly desirable but has not been included in this work.

9.3.1.1 Spherical type variogram, isotropic model

Fitted model, semivariogram parameters:

Range	=	106.2
Variance	=	1.25

Trend parameters:

β_0	=	-2.07
β_1	=	2.07
β_2	=	0.074
β_3	=	40.0

For a look at these variogram see figure 9.3.1.1.1. The jacknifing measures are:

MRE	=	1.40E-3
$MSRE^{1/2}$	=	0.923
$MSE^{1/2}$	=	0.192
J	=	0.098

9.3.1.2 Spherical type variogram, anisotropic model

Fitted model, semivariogram parameters:

Variance	=	1.20

Horizontal range = 241.4

Vertical range = 79.4

Trend parameters:

β_0 = -2.22

β_1 = 2.05

β_2 = 0.054

β_3 = 40.0

Jacknifing measures:

MRE = 7.34E-4

$MSRE^{1/2}$ = 0.847

$MSE^{1/2}$ = 0.192

J = 0.164

9.3.1.3 Exponential type variogram, isotropic model

Fitted model, semivariogram parameters:

Variance = 1.29

(Practical)Range = 61.9 (185.7)

Trend parameters:

β_0 = -2.49

β_1 = 1.93

β_2 = 0.064

β_3 = 40.0

For a look at these variogram see figure 9.3.1.3.1. The jacknifing measures are:

MRE = 0.141E-2

$MSRE^{1/2}$ = 0.849

$MSE^{1/2}$ = 0.191

J = 0.172

9.3.1.4 Exponential type variogram, anisotropic model

Fitted model, semivariogram parameters:

Variance = 1.22

(Practical) Horizontal range = 80.8 (242.3)

(Practical) Vertical range = 44.0 (132.0)

Trend parameters:

β_0 = -2.465

β_1 = 1.966

β_2 = 0.075

β_3 = 40.0

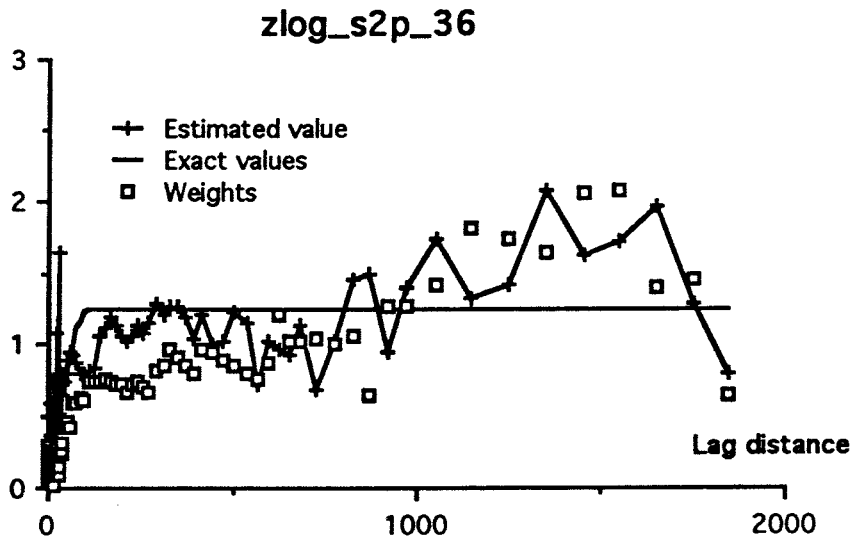
Jacknifing measures:

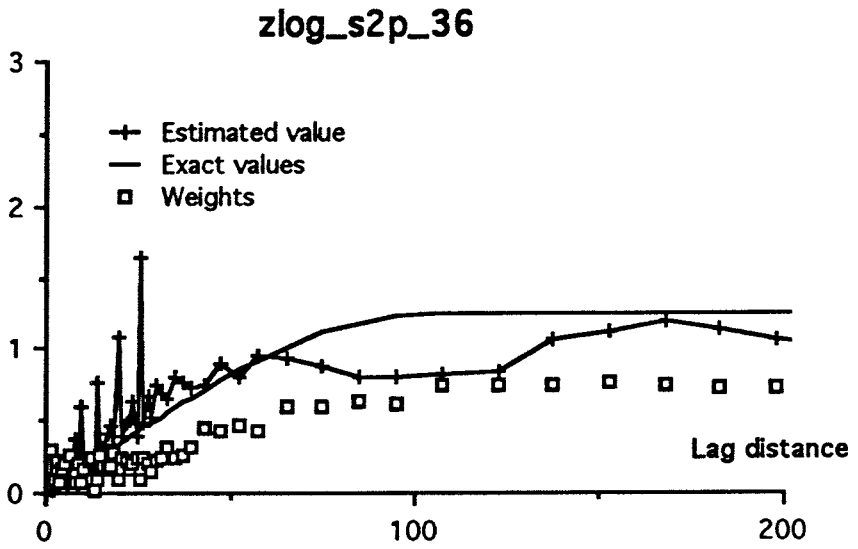
MRE = 0.106E-2

$MSRE^{1/2}$ = 0.759

$MSE^{1/2}$ = 0.191

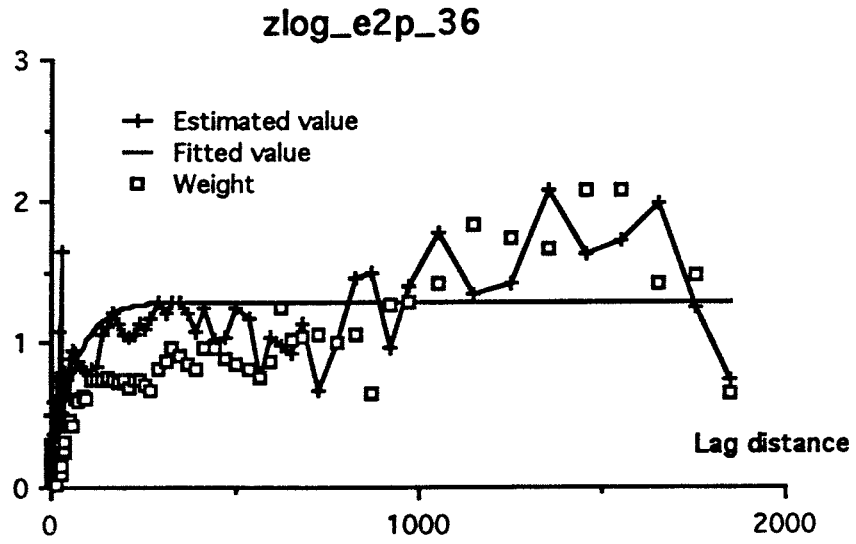
J = 0.257

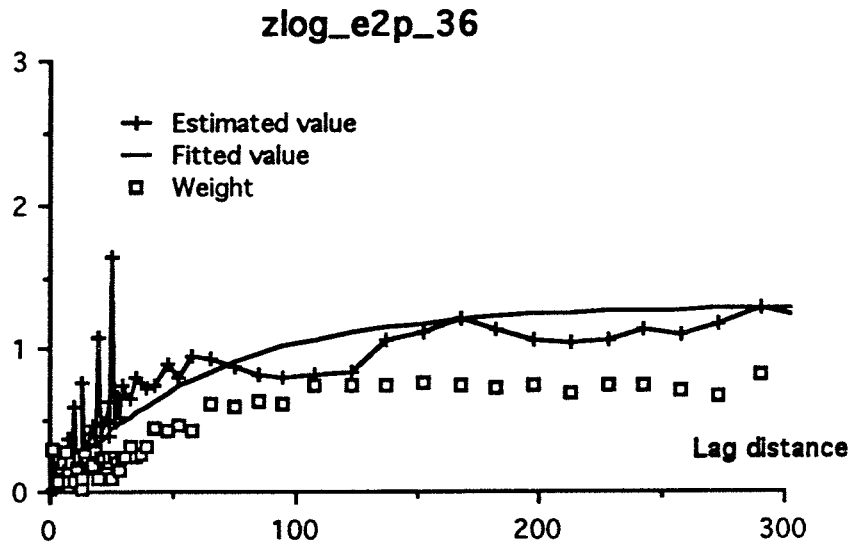




	Extremals for β_0		Extremals for β_1		Extremals for β_2	
β_0	2.43	-8.46	2.35	-8.37	-3.02	-3.00
β_1	3.94	-0.56	3.98	-0.59	1.70	1.69
β_2	0.025	0.026	0.026	0.025	0.26	-0.21

Figure 9.3.1.1.1 *Showing spherical isotropic variograms for the residuals after removal of a potential type trend. The table shows the extremal points on the 95% confidence ellipsoid according to section 6.2.4. Note in particular that this results implies that the hypothesis: "The contrast parameter β2 is zero." cannot be rejected on the 95% level.*





	Extremals for β_0		Extremals for β_1		Extremals for β_2	
β_0	0.27	-5.25	0.24	-5.22	-2.50	-2.49
β_1	3.07	0.78	3.08	0.77	1.94	1.92
β_2	0.064	0.065	0.066	0.063	0.27	-0.14

Figure 9.3.1.3.1 *Showing exponential isotropic variograms for the residuals after removal of a potential type trend. The table shows the extremal points on the 95% confidence ellipsoid according to section 6.2.4. Note in particular that this results implies that the hypothesis: "The contrast parameter β2 is zero." cannot be rejected on the 95% level.*

9.3.2 Exponential type trend function

The trend function is of type

$$\log(K) = \beta_0 - \beta_1 z + \beta_2 I\{(x, y, z) \in zone\ 2\}$$

or alternatively, which is the reason for the name of the trend type

$$K = \begin{cases} b_0 b_2 b_1^{-z} & \text{if } (x, y, z) \in zone\ 2 \\ b_0 b_1^{-z} & \text{if } (x, y, z) \notin zone\ 2 \end{cases}$$

with

$$\begin{cases} b_0 = 10^{\beta_0} \\ b_1 = \beta_1 \\ b_2 = 10^{\beta_2} \\ b_3 = \beta_3 \end{cases}$$

Note again that we fit the function E[logK] to the measured values $\log(K_i)$ and that this is not equivalent with fitting the model for E[logK] to the measured values K_i.

9.3.2.1 Spherical type variogram, isotropic model

Fitted model, semivariogram parameters:

Range	=	97.4
Variance	=	1.22

Regression parameters:

β_0	=	-5.85
β_1	=	0.0046
β_2	=	0.072

For a look at these variograms and the confidence intervals of the regression parameters see figure 9.3.2.1.1. The jacknifing measures are:

MRE	=	0.655E-3
$MSRE^{1/2}$	=	0.897
$MSE^{1/2}$	=	0.192
J	=	0.113

9.3.2.2 Exponential type variogram, isotropic models

Semivariogram parameters:

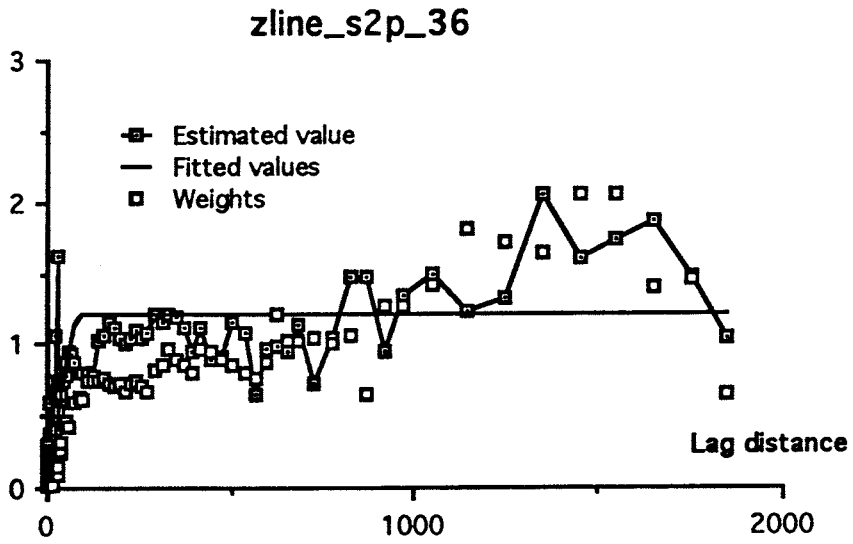
(Practical) Range	=	58.2 (174.6)
Variance	=	1.25

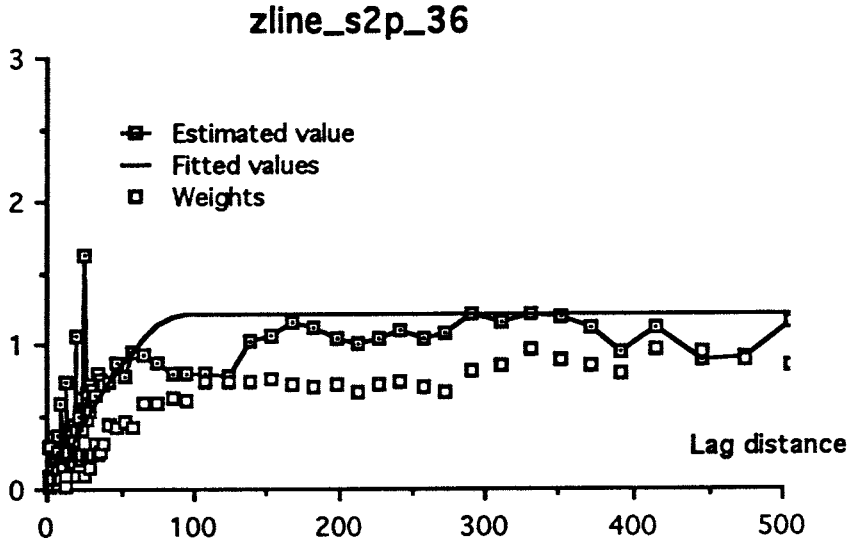
Regression parameters:

β_0	=	-5.95
β_1	=	0.0044
β_2	=	0.063

For a look at these variograms and the confidence intervals of the regression parameters see figure 9.3.2.2.1. The jacknifing measures:

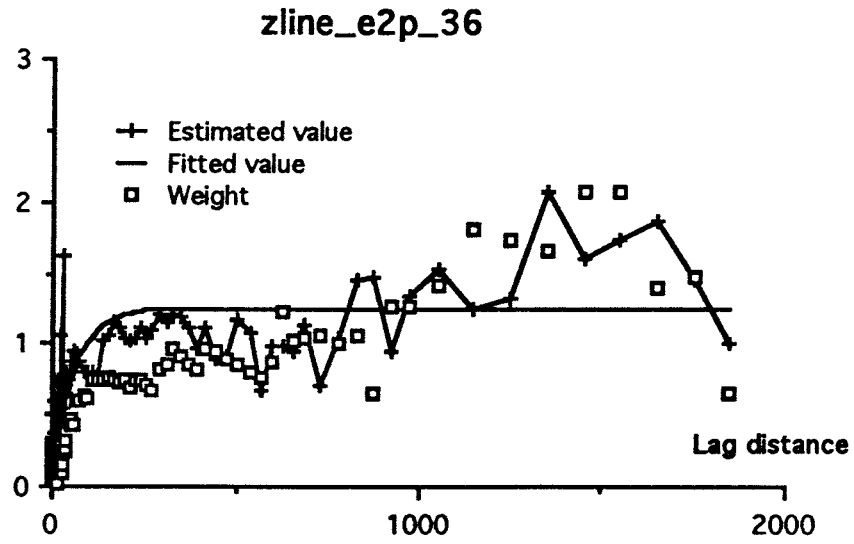
MRE	=	0.869E-3
$MSRE^{1/2}$	=	0.837
$MSE^{1/2}$	=	0.191
J	=	0.176

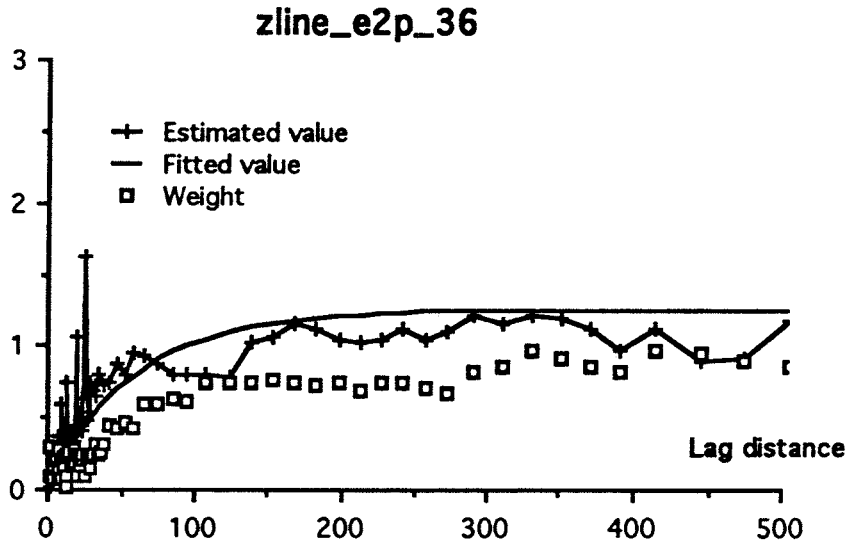




	Extremals for β_0		Extremals for β_1		Extremals for β_2	
β_0	-5.26	-6.451	-5.351	-6.357	-5.888	-5.820
β_1	0.006	0.003	0.007	0.003	0.005	0.005
β_2	0.060	0.084	0.068	0.076	0.281	-0.137

Figure 9.3.2.1.1 *Showing spherical isotropic variograms for the residuals after removal of a exponential type trend. The table shows the extremal points on the 95% confidence ellipsoid according to section 6.2.4. Note in particular that this results implies that the hypothesis: "The contrast parameter $\beta 2$ is zero." cannot be rejected on the 95% level.*





	Extremals for β_0		Extremals for β_1		Extremals for β_2	
β_0	-5.25	-6.65	-5.37	-6.53	-5.98	-5.92
β_1	0.006	0.003	0.007	0.002	0.004	0.004
β_2	0.054	0.071	0.06	0.066	0.27	-0.15

Figure 9.3.2.2.1 *Showing exponential isotropic variograms for the residuals after removal of a exponential type trend.The table shows the extremal points on the 95% confidence ellipsoid according to section 6.2.4.Note in particular that this results implies that the hypothesis: "The contrast parameter β2 is zero." cannot be rejected on the 95% level.*

9.3.2.3 Conclusion for the residual models

Looking at the variogram fits in figures 9.3.1.1.1 and 9.3.1.3.1 one sees, admittively with some effort, that the spherical model gives the better fit. This is reflected in the jacknifing measures which gives the spherical model as the best. As in the intrinsic case the anisotropic models behave worse than the isotropic. The potential type trend function gives overall better results than the exponential type.

10 CONCLUSIONS AND CAVEATS

This report has shown that :

It is possible to perform regularization of packer measurements given the assumption that the leakage due to the test is negligible, see section 2.

It is possible to use jacknifing as a criterion on model "goodness", see section 7. Woodbury's formula is a useful tool when calculating the jacknifing measures. This gives hypothesis tests far superior to those obtained by analytical formulae, which require adoption of invalid assumptions.

It is possible to perform hypothesis testing of model consistency with jacknifing, see section 8.

Some problems with the approach has also been identified:

It is difficult to estimate correlation lengths in the case of anisotropic models.

One should not divide the lag space into lag classes since this introduces a lot of arbitrariness in the analysis, see section 6.1 and 9.2.2.

In using models like the potential type one should be able to perform nonlinear regression, see section 9.3.1.

In the specific case of the Finnsjön site the choice of basis functions for regression could be improved upon, in particular with respect to the description of zone 2, dividing it into a upper and lower part. See the figures 2.2.4 to 2.2.6.

The jacknifing procedure is certainly disturbed by the fact that the measurement points are so strongly clustered. One result that points in this direction is the unsensitivity of the jacknifing measures for variations in the choice of kriging neighbourhoods, see section 9.2.1. Clearly what one is interested in is the prediction capabilities of a model over long ranges. In the current approach, the close-range estimates totally dominate the picture. A possible remedy would be to remove more than one point at a time.

The use of intrahole lag vectors is preferred to the sole use of interhole lagvectors, see section 9.2.2 and 9.2.4.

The process of finding the variogram as described in sections 6.1 and 6.1.1 does not lead to the best model from the jacknifing point of view, see section 9.1.3.

The results of the calculations on the 36 m scale gives the following conclusions:

The best model found is the exponential isotropic one without an explicit trend. (J = 0.062)

The best residual model found is the spherical isotropic one using an potential type trend. (J = 0.098)

The anisotropic models are in general worse than the isotropic.

The exponential-type trend is worse than the potential type trend.

The hypothesis that the expectation value in zone 2 is equal to the expectation value in the rock mass cannot be rejected on the 95% confidence level. No other level of confidence has been studied.

The hypothesis that the estimated model is correct can be rejected at the 95% level of confidence for all isotropic models and thus all models.

Especially the last point is not satisfactory since it means that we have not found any sufficiently good model. The following measures could be taken in the future to improve the situation:

Covariance model fit is rather poor and could be improved with nested models. Non-transition models could also be tried, see however 4.4.

As noted in section 2.2 the distributions are not especially normally distributed. Therefore one might try to find a suitable transform other than the simple logarithmic transform used that produces values which better fit a normal distribution.

The kriging neighbourhoods could be improved by first dividing the rock into rock mass and zone 2 before applying the slicing procedure described in section 5.2.

11 ACKNOWLEDGEMENTS

The author wishes to thank:

Mr Anders Ström SKB AB for his constant support and ability to get the author going.

Mr Nils Kjellbert SKB AB for his good spirits and fruitful discussions.

Mr Lars Lovius Starprog AB who wrote large parts of the code in INFERENS and the appendixes B and E.

Mr Lars Eriksson Starprog AB for enlightening conversations.

Mr. Joel Geier GOLDER for proofreading the entire manuscript.

Mrs Marita Norman for her patience, Mr. Viktor Norman and Ms. Erika Norman for being.

12 REFERENCES

[Ahlbom and Tiren, 1991] Ahlbom, K., Tiren, S. Overview of the geological and geohydrological conditions at the Finnsjön site and its surroundings, SKB TR 91-08.

[Ahlbom et al, 1988] Ahlbom, K et al, Characterization of Fracture Zones in the Brändan Area Finnsjön Study Site, Central Sweden SKB AR 88-09.

[Andersson et al, 1991] Andersson, J. E. et al , Hydrogeological Conditions in the Finnsjön Area, SKB TR 91-24.

[Braester and Thunvik] Braester, C., Thunvik, R., Numerical simulation of double packer tests. Calculation of rock permeability., SKB TR 82 - 06.

[Chiles, 1977] Géostatistique des phénomènes non stationnaries. Thèse, Univ. of Nancy 1, France.

[Clifton and Neuman, 1982] Clifton, P. M., and Neuman, S. P., Effects of Kriging and Inverse Modelling on Conditional Simulation of the Avra Valley Aquifer in Southern Arizona, Water Resources Research, Vol 18, No 4, pp 1215 - 1234, August 1982.

[Cressie and Hawkins, 1980] Cressie, N., Hawkins, D., Robust estimation of the variogram I, J. Int. Assoc. Math. Geol., 12, 115-125, 1980]

[Darcy, 1856] Darcy, H., Les fontaines publiques de la ville de Dijon, Dalmont, Paris.

[de Marsily, 1986] de Marsily, G., Quantitative Hydrogeology, Academic Press, 1986.

[Delfiner, 1976] Delfiner, P., Linear Estimation of Non Stationary Spatial Phenomena, in M. Guarascio et al. (eds.), Advanced Geostatistics in the Mining industry, 49-68, D. 1976, Reidel publishing Company, Dordrecht-Holland.

[Delhomme, 1979] Delhomme, J. P., Spatial Variability and Uncertainty in groundwater Flow Parameters: a Geostatistical Approach. Water Resour. Res. 15(2), pp. 269 - 280.

[Draper and Smith] Draper, N. R. and Smith, H., Applied Regression Analysis, Wiley.

[Fisz, 1963] Fisz, M., Probability theory and mathematical statistics, Wiley, 1963.

[Gray and O'Neill, 1976] Gray, W. and O'Neill, K., On the General Equations for Flow in Porous Media and their reduction to Darcy´s law, Water Resources Research April 1976, pp. 148 - 154.

[Journel and Huijbregts, 1978] Journel, A. Huijbregts, Ch. J. , Mining Geostatistics, Academic Press, 1989.

[Kitanidis, 1983] Kitanidis, P. K. , Statistical estimation of polynomial generalized covariance functions and hydrologic applications, Water Resour. Res., 19, 909-921, 1983.

[Lindbom et al, 1991] Lindbom, B., Boghammar, A., Lindberg H., Bjelkås, J., Numerical groundwater flow calculations at the Finnsjön site, SKB TR 91-12, 1991.

[Lovius et al, 1990] Lovius, L., Norman, S., Kjellbert N., Uncertainties in repository performance from spatial variability of hydraulic conductivities-statistical estimation and stochastic simulation using PROPER, SKB TR 90-03.

[Marle, 1967] Marle, C., Ecoulements monophasiques en mileu poreux, Rev. Inst. Fr. Pe´t. 22(10), 1471-1509.

[Matheron, 1973] Matheron, G., The Intrinsic Random Functions and their Application, Adv. Appl. Prob. 5, 439-468, 1973.

[Moye, 1967] Moye, D. G., Diamond drilling for foundation exploration, Civil engineering transactions, 1967, pp 95 - 100.

[Neuman and Jacobson, 1984] Neuman, S. P., and Jacobson, E. A., Analysis of Nonintrinsic Spatial Variability by residual kriging with Application to Regional Groundwater levels, Math. geol, Vol. 16, No 5, 1984.

[Norman and Geier, 1991] Norman, S., and Geier, J., Interface between discrete fracture network and stochastic continuum models for the SKB91 safety assesment study. Proposal to SKB 1991.

[Omre, 1984] Omre, H. , The variogram and its estimation in Geostatistics for Natural Resources characterization, vol. 1, NATO ASI Ser. C,

ed. G. Verly et al. , p. 122, D. Reidel, Hingham, Mass. 1984.

[Placket, 1960] Placket, R. L., Regression Analysis, Clarendon Press, Oxford, 1960.

[Press et al, 1989] Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., Numerical Recipies, The art of scientific computing, Cambridge University Press 1989.

[Russo and Jury, 1987] Russo, D. and Jury, W. A., A Theoretical Study of the Estimation of the Correlation Scale in Spatially Variable fields, 1. Stationary fields, Water resour. res., Vol. 23, No 7, pp 1257-1268, July 1987.

[Whitaker, 1985] Whitaker, S. Flow in Porous Media 1: A Theoretical Derivation of Darcy´s law, Transport in Porous Media 1 (1986) 3-25, D. Reidel Publishing Company.

[Winberg, 1989] Winberg, A., Analysis of spatial variability of hydraulic conductivity in the SKB database GEOTAB, SKI TR 89:12, 1989.

APPENDIX A INTRINSIC RANDOM FUNCTIONS

The objective of this appendix is to explore the theory of intrinsic random functions in more detail. In order to do that we have to start with some more exact definitions of a stochastic function.

A.1 Mathematical definition of a stochastic function

First we define what we will mean with a **stochastic function**. In order to do that we first introduce the event space Ω which is a so-called measure space. The meaning of that notion is that there is defined an σ–algebra $\aleph$[21] of subsets to Ω. The sets in this algebra are known in the language of probability theory as **events**. Finally there also exists a positive probability measure P defined on the members of $\aleph$ such that $P(\Omega) = 1$ and that

$$P\left(\bigcup_i O_i\right) = \sum_i P(O_i)$$

if the sets O_i are disjoint, that is the events O_i are independent.

We define the space $L(\Omega, \aleph, P)$ as the real (or complex) valued functions X that are measurable with respect to $\aleph$[22]. In the language of probability theory X is a **stochastic variable**. We also define $L^2(\Omega, \aleph, P)$ as the subset of $\{X \in L(\Omega, \aleph, P)\}$ whose members satisfy

$$\int_{\Omega} X(\omega)^2 dP(\omega) < \infty \quad . \qquad \text{A.1.1}$$

Then X is a stochastic variable with the extra quality of having a finite variance.

Now define a **stochastic function** on $\mathbf{R}^3$ as a mapping

$$Y : \mathbf{R}^3 \rightarrow L^2(\Omega, \aleph, P) \qquad \text{A.1.2}$$

---

[21] A collection $\aleph$ of subsets of a set Ω is said to be a σ-algebra in Ω if $\aleph$ has the following properties

(i) $\Omega \in \aleph$

(ii) If $O \in \aleph$ then $O^C \in \aleph$ where O^C denotes the complement of O relative to Ω.

(iii) If $O = \bigcup_{i=1}^{\infty} O_i$ and if each $O_i \in \aleph$ then $O \in \aleph$.

[22] A realvalued function F is measurable with respect to $\aleph$ if for every open set O in R the inverse image $F^{-1}(O)$ belongs to $\aleph$.

this is to say that a stochastic function is a mapping that assigns a stochastic variable to each point $\mathbf{x} \in \mathbf{R}^3$. This can of course also be written in the same manner as in chapter 3 i.e. as $Y(\mathbf{x}, \omega)$.

In order to avoid misunderstanding it is appropriate to insert a short discussion of the event space Ω. The element or "points" ω of Ω can be taken to consist of mappings $\omega:\mathbf{R}^3 \rightarrow \mathbf{R}^1$ [23]. The mapping A.1.2 is then the map $Y:\mathbf{x} \rightarrow Y(\mathbf{x})(\omega) = \omega(\mathbf{x})$. Thus in particular the distribution function of the stochastic variable $Y(\mathbf{x})$ is $P(Y(\mathbf{x}) \leq \alpha) = P(\{\omega : \omega(\mathbf{x}) \leq \alpha\})$ and the requirement A.1.1 is reformulated as

$$\int_{\Omega} \omega(\mathbf{x})^2 dP(\omega) < \infty \qquad \text{A.1.3}$$

for all $\mathbf{x} \in \mathbf{R}^3$. The formulas A.1.3 and A.1.1 are usually expressed by the phrase that the stochastic function is of second order. Note that this has nothing to do with stationarity.

After these preliminaries we define the expectation value function or trend of a second order stochastic function as

$$m_Y(\mathbf{x}) = E[Y(\mathbf{x})] = \int_{\Omega} Y(\mathbf{x}, \omega) dP(\omega) < \infty$$

and the covariance

$$C_F(\mathbf{x}_1, \mathbf{x}_2) = E\left[(Y(\mathbf{x}_1) - m_Y(\mathbf{x}_1))(Y(\mathbf{x}_2) - m_Y(\mathbf{x}_2))\right] =$$

$$\int_{\Omega}\int_{\Omega} (Y(\mathbf{x}_1, \omega_1) - m_Y(\mathbf{x}_1))(Y(\mathbf{x}_2, \omega_2) - m_Y(\mathbf{x}_2)) dP(\omega_1) dP(\omega_2) < \infty$$

That these integrals exist as finite values is a consequence of the second-order assumption and the Cauchy Schwarz inequality. Finally $Y(\mathbf{x})$ is a **weakly (second order) stationary** stochastic function if $m_Y(\mathbf{x})$ is a constant and $C_Y(\mathbf{x}_1, \mathbf{x}_2)$ depends only on the difference $\mathbf{x}_1 - \mathbf{x}_2$.

A.2 Mathematical definition of intrinsic random functions

To make the definition of an intrinsic random function precise we follow [Matheron, 1973] and start by introducing the class Λ of measures on $\mathbf{R}^3$ with **finite support** i.e. measures λ such that

---

[23] That is we take Ω to consist of all possible realizations. This is equivalent to take the sample space equal $N_6 = \{1,2,3,4,5,6\}$ in the standard example case of tossing of a die and $N_6 \otimes N_6 \otimes \ldots$ in the case of an infinite, or undetermined, number of tosses.See [Doob, 1990].

$$\int f(\mathbf{x})d\lambda(\mathbf{x}) = \sum_i f(\mathbf{x}_i)\lambda_i$$

holds for any continuous function f:$\mathbf{R}^3 \rightarrow \mathbf{R}$. The above sum contains only finitely many terms. Next we define a translation operator τ_ξ on this class of measures by the requirement that

$$\int f(\mathbf{x})d(\tau_\xi\lambda)(\mathbf{x}) = \sum_i f(\mathbf{x}_i + \xi)\lambda_i$$

should hold for all continuous functions f.[24]

Now define Λ' as the class of all finitely supported measures λ which annihilate the constant functions, i.e. measures λ such that

$$\int_{\mathbf{R}^3} d\lambda = \sum_{i=1}^{n} \lambda_i = 0 \quad . \qquad \text{A.2.1}$$

We will say that Z is a **generalized second order random function** on Λ' if it is a linear map $Z: \Lambda' \rightarrow L^2(\Omega, \aleph, P)$.

A generalized random function on Λ' is an **intrinsic random function of order zero** if the stochastic function (of $\mathbf{x}$) $Y(\tau_{\mathbf{x}}\lambda)$ is second-order stationary for any choice of λ in Λ'. That this stochastic function is second-order stationary is equivalent to the requirement that

$$E[Z(\lambda)Z(\lambda')] = E[Z(\tau_\xi\lambda)Z(\tau_\xi\lambda')]$$

for any choices of measures λ and λ' in Λ'. Expanding this gives

$$E\left[\sum_{i=1}^{n}\lambda_i Z(\mathbf{x}_i)\sum_{j=1}^{n'}\lambda'_j Z(\mathbf{x}'_j)\right] = E\left[\sum_{i=1}^{n}\lambda_i Z(\mathbf{x}_i + \xi)\sum_{j=1}^{n}\lambda'_j Z(\mathbf{x}'_j + \xi)\right]$$

and thus in particular

$$E\left[(Z(\mathbf{x}_1) - Z(\mathbf{x}_2))^2\right] = E\left[(Z(0) - Z(\mathbf{x}_2 - \mathbf{x}_1))^2\right] = 2\gamma(\mathbf{x}_2 - \mathbf{x}_1)$$

where $\gamma(\mathbf{h})$ is the semivariogram of Z. This shows that the above definition of an intrinsic random order function implies the definition in section 3.1.

---

[24]The choice of the continuous functions is natural since the measures of compact support is the dual space of the continuous functions.

Conversely let us show that this definition of an intrinsic random function follows from the definition of an intrinsic random function as contained in 3.2.1 and 3.2.2 in the case of zero order. In fact

$$E[Z(\lambda)Z(\lambda')] =$$

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda'_j Z(\mathbf{x}_i)Z(\mathbf{x}'_j)\right] = -\frac{1}{2}E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda'_j\left(Z(\mathbf{x}_i) - Z(\mathbf{x}'_j)\right)^2\right] =$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda'_j\gamma(\mathbf{x}_i - \mathbf{x}_j) \qquad \text{A.2.2}$$

where we used A.2.1 and 3.2.2 in the second equality. Clearly this and 3.2.1 proves our case and it also shows in particular by choosing $\lambda = \lambda'$ that the semivariogram is conditionally negative semi-definite, i.e. that

$$-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\mathbf{x}_i - \mathbf{x}_j) \geq 0$$

for any λ in Λ' and that the semivariogram is the generalized covariance of order zero as defined in [Matheron, 1973, p 450] and [de Marsily, 1986, p 314].

A.3 Representations of an intrinsic stochastic function

As seen there is a major difference between stochastic functions as defined in section A.1 and intrinsic stochastic functions in that the latter are defined on a space of finite measures. Now we say that a stochastic function Y(x) is a **representation** of the intrinsic random function Z(λ) if

$$Z(\lambda) = \int Y(\mathbf{x})d\lambda(\mathbf{x}) \qquad \forall\ \lambda \in \Lambda'$$

First we consider the question of the existence of a representation. It would be tempting to put $\lambda = \delta(\mathbf{x} - \mathbf{x}')$ in the above formula, where $\delta(\cdot)$ is the Dirac measure, but this measure does not belong to Λ'. However this is easily rectified by taking

$$\lambda(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}') - \delta(\mathbf{x})$$

where $\mathbf{x}'$ is an arbitrary point and thus deriving as a necessary condition

$$Z(\delta(\mathbf{x} - \mathbf{x}') - \delta(\mathbf{x})) = Y(\mathbf{x}') - Y(\mathbf{0}).$$

We easily see that this condition is also sufficient since taking any measure λ in Λ' then integrating the above formula with respect to $\mathbf{x}'$ we have

$$\int Y(\mathbf{x}')d\lambda(\mathbf{x}') = \int Z(\delta(\mathbf{x} - \mathbf{x}') - \delta(\mathbf{x}))d\lambda(\mathbf{x}') \qquad \text{A.3.1}$$

since obviously

$$\int Y(\mathbf{0})d\lambda(\mathbf{x}') = Y(\mathbf{0})\int d\lambda(\mathbf{x}') = 0$$

Now since λ is a finite measure the right hand side of A.3.1 is written

$$\sum_j Z\left(\delta\left(\mathbf{x}-\mathbf{x}'_j\right)-\delta(\mathbf{x})\right)\lambda_i$$

for some set of points $\{\mathbf{x}'_j\}$ and moreover since, by definition, Z is a linear mapping this is equal to

$$Z\left(\sum_j\left(\delta\left(\mathbf{x}-\mathbf{x}'_j\right)-\delta(\mathbf{x})\right)\lambda_i\right).$$

Finally it is clear that

$$\sum\left(\delta\left(\mathbf{x}-\mathbf{x}'_j\right)-\delta(\mathbf{x})\right)\lambda_i = \sum\lambda_i\delta\left(\mathbf{x}-\mathbf{x}'_j\right) = \lambda(\mathbf{x})$$

and thus

$$\int Y(\mathbf{x}')d\lambda(\mathbf{x}') = Z(\lambda)$$

which was what we wanted to show.

Now the above reasoning can also be used to obtain all representations of an intrinsic function Z. To that end let us assume that Y_1 and Y_2 are two representations of the intrinsic random function Z. Thus by definition

$$\int Y_1(\mathbf{x})d\lambda(\mathbf{x}) = \int Y_2(\mathbf{x})d\lambda(\mathbf{x}) \qquad \forall\ \lambda \in \Lambda'$$

and in particular choosing $\lambda \in \Lambda'$ as

$$\lambda(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}') - \delta(\mathbf{x})$$

we have

$$g_1(\mathbf{x}') = g_2(\mathbf{x}') + X$$

for all points $\mathbf{x}'$ and where X is a stochastic variable given by i.e

$$X = \left(Y_2(\mathbf{0}) - Y_1(\mathbf{0})\right) \in L^2(\Omega, \aleph, P)$$

Note that the integrals written here all refer to functions with values in the Hilbert space $L^2(\Omega, \aleph, P)$. A very nice discussion of such (abstract) integrals is found in [Ladas and Lakshmikantham, p 1- 20].

A.4 Relations to kriging

As pointed out in [Delfiner, 1976, p. 57] and as shown in section A.2 an intrinsic function (of order zero) is an equivalence class of stochastic functions that differ only

by an arbitrary trivial random function $X(\omega)\cdot 1$. Furthermore the key equation is the special case of A.2.2 with $\lambda' = \lambda$

$$E\left[Y(\lambda)^2\right] = -\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\mathbf{x}_i - \mathbf{x}_j)$$

since this may be interpreted as expressing estimation error solely in terms of the variogram. In fact choosing λ such that

$$\int f(\mathbf{x}')d\lambda(\mathbf{x}') = f(\mathbf{x}) - \sum_{i\in D(x)}\lambda_i(\mathbf{x})f(\mathbf{x}_i)$$

for any continuous function $f(\mathbf{x})$ where the weights λ_i satisfy the nonbias condition, we see that the kriging error is expressible as

$$E\left[\left(Y(\mathbf{x}) - \sum_{i\in D(x)}\lambda_i(\mathbf{x})Y(\mathbf{x}_i)\right)^2\right] =$$

$$2\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i(\mathbf{x})\gamma(\mathbf{x}-\mathbf{x}_j) - \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i(\mathbf{x})\lambda_j(\mathbf{x})\gamma(\mathbf{x}_i - \mathbf{x}_j)$$

Thus the point, [Delfiner, 1976], hat if the interest is to predict the values of the intrinsic random function, we do not need the covariance of a representation but only the variogram (generalized covariance) which is the covariance for the intrinsic random function.

A.5 References

[Ladas and Lakshmikantham, 1972] Ladas, G. E., Lakshmikantham, V., Differential equations in abstract spaces, Academic Press 1972

[Matheron, 1973] Matheron, G., The Intrinsic Random Functions and their Application, Adv. Appl. Prob. 5, 439-468, 1973.

[de Marsily, 1986] de Marsily, G., Quantitative Hydrogeology, Academic Press, 1986.

[Doob, 1990] Doob, J. L., Stochastic Processes, Wiley 1990.

[Delfiner, 1976] Delfiner, P., Linear Estimation of Non Stationary Spatial Phenomena, in M. Guarascio et al. (eds.), Advanced Geostatistics in the Mining industry, 49-68, D. 1976, Reidel puplishing Company, Dordrecht-Holland.

APPENDIX B. RETRIEVING DATA FROM GEOTAB

The borehole information was retrieved from the SKB database GEOTAB.

The table SHSINJCD contains conductivity measurements data. Here we chos the columns IDCODE, , SECLEN, SECUP and K. IDCODE contains the name of current hole SECLEN the length of the packer section (c.f. next appendix), SECUP gives the length to the upper boundary of the measurement section along the hole and K contains stationary values of the measured conductivity.

To get the coordinates of the measurement points in the so-called "local grid" system the table mentioned above has to be combined with the table BOREHOLE which gives the coordinates in intervals of ten meters along the holes. From this table we selects the columns XCOORD YCOORD and ZCOORD. The value of these coordinates are interpolated to upper measurement section using the GEOTAB "#" command when issuing the WHERE statement. The selected columns are then stored as a temporary table (here called TMPTB1). To perform this we type (our command are given in bold face, comments inside "<>"):

R>**select shsinjcd bhcoord**

.

.

select columns:**1, 6, 14-16, 3, 2** <the choice are given in a menu not listed here>

where: **:1 like *FI* and :1=:12 and :3 # :13** <data from Finnsjön has FI in the name>

Output to ?(Screen, Table, Both(S&T), Formatted file, Printfile)**T**

Name of temporary table:**tmptb1**

No, of columns in primary key ? (all) <just hit return>

.

.

Now , the column ZCOORD in table BHCOORD does not contain the depth in the "local grid" system but rather the depth below the ground level. The relations to the fixed entity - meters below sea level - are given in the table BOREHOLE where the column Z gives the ground level for each borehole. We now combine the tables TMPTB1 and BOREHOLE in a new temporary table (TMPTB2) containing the offset value Z not only once per borehole (as in table BOREHOLE) but for each measurement point. This temporary table is now selected. A new column is opened using the "+" command when selecting column. This column are then specified to give the depth

below the sea level i.e. ZCOORD-Z. The final result is then saved as an ascii file, which can be split in one file per borehole and sorted by ascending values in the SECCUP column to obey the HYDRASTAR format. The remaining GEOTAB steps to generate the output ascii file are as follows:

R>**select tmptb1, borehole**

.

.

select columns: **1-7, 15**

where: **:1=:8**

Output to ?(Screen, Table, Both(S&T), Formatted file, Printfile)**T**

Name of temporary table:**tmptb2**

No, of columns in primary key ? (all) <hit return>

.

.

R>**select tmptb2**

.

.

select columns:**1-4**

select columns:+

Enter column heading: **mbsl**

Enter expression: **:3 - :8**

.

select columns:**6-7**

where: <no condition just hit return>

Output to ?(Screen, Table, Both(S&T), Formatted file, Printfile)**F**

.

.

Select: **1** <the choice from menu is for ascii file>

Outputfile:**out**

.

.

Ascii conversion completed, output file : out.asc

APPENDIX C. MOYE'S FORMULA

This appendix derives Moye's formula [Moye, 1967] and is inserted solely for the conveniance of the reader.

We study the packer test by using the mathematical model expressed by the formulas 2.1.4 and 2.1.5. In particular we repeat the approximation that the flow per difference head q_1 is written using the notation of section 2.1 as

$$q_1 := \frac{Q_1}{\Delta h_1} = \int_{z_1}^{z_1+L_1} K(\mathbf{x})\left(\frac{1}{\Delta h_1}\frac{\partial h_0(\mathbf{x})}{\partial \rho} + \frac{\partial h_1(\mathbf{x})}{\partial \rho}\right)\rho_w dz \approx$$

$$\int_{z_1}^{z_1+L_1} K(\mathbf{x})\frac{\partial h_1(\mathbf{x})}{\partial \rho}\rho_w dz \qquad \text{C.1}$$

Now the flow resulting from the overpressure is assumed to be purely cylindrical up to a radial distance of L/2, see fig C.2. Assuming the conductivity to be a constant K_1 in a neighbourhood of the packer test we solve the steady state hydrology equation in a cylindrical system of coordinates shown in 2.1.1 to obtain for the head

$$h_1(\rho) = A\ln(\rho) + B,$$

where A and B are constants to be determined, and for the radial component U_ρ of the Darcy velocity

$$U_\rho(\rho) = -K_1\frac{A}{\rho}.$$

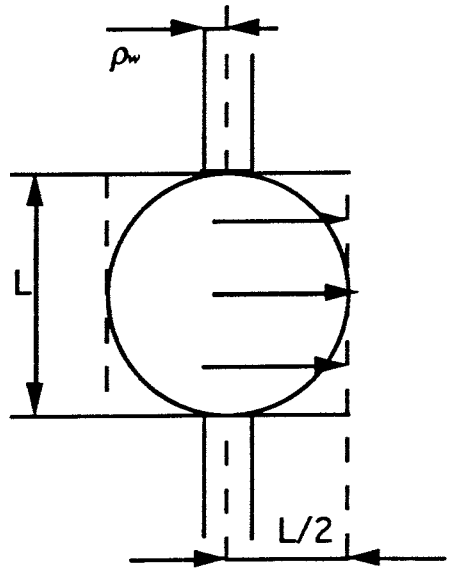


Figure C.2 *Illustrating a packer test and the corresponding derivation of Moye's formula.*

Since the total flux Q_1 is known we have

$$A = -\frac{Q_1}{2\pi L K_1}$$

and the head on the boundary of the pressurized section is by the boundary conditions 2.1.5 given as

$$h_1(\rho_w) = \Delta h_1 .$$

This gives

$$B = \Delta h_1 + \frac{Q_1}{2\pi L_1 K_1} \ln(\rho_w)$$

and thus

$$h_1(\rho) = \Delta h_1 + \frac{Q_1}{2\pi L_1 K_1} \ln\left(\frac{\rho_w}{\rho}\right) .$$

Assume that the flux changes abruptly from radial to spherical at $\rho = r = L/2$. For spherical flux in a homogeneous medium we have for the head

$$h_1(r) = \frac{A}{r} + B$$

and for the Darcy velocity

$$U_r(r) = K\frac{A}{r^2} .$$

It follows from the boundary conditions 2.1.5 and the underlying assumption that the position of the measurement is far from the the boundary that

$$\lim_{r \to \infty} h_1(r) = 0$$

giving B = 0 above. Next transferring the head from the radial case at $\rho = r = L/2$ gives

$$A = \frac{L_1}{2}\Delta h_1 + \frac{Q_1}{4\pi K_1} \ln\left(\frac{2\rho_w}{L_1}\right)$$

and thus if we equate the total flow in the radial and spherical regions we get

$$Q_1 = 4\pi K_1 \frac{L_1}{2}\Delta h_1 + Q_1 \ln\left(\frac{2\rho_w}{L_1}\right)$$

from which we derive Moye´s formula

$$K_1 = \frac{Q_1\left(1 - \ln\left(\frac{2\rho_w}{L_1}\right)\right)}{2\pi L_1 \Delta h_1} = q_1\frac{\left(1 - \ln\left(\frac{2\rho_w}{L_1}\right)\right)}{2\pi L_1} \qquad \text{C.2}$$

APPENDIX D WOODBURY'S FORMULA

First let us define an r-dimensional perturbation of a matrix C as any matrix C' such that

$$C' = C - UV^T$$

where U and V both have r linearly independent columns. The reason for the name r-dimensional is that the matrix UV^T with the stated restrictions is a general form of a linear map with r-dimensional range. [25]

Woodbury's formula concerns the solution of a linear system of equations

$$C'\lambda = c$$

in terms of the inverse of the matrix C. Generally speaking this is very useful when we have some knowledge of C such as for instance a factorization and want to solve a system of equations involving C'. Thus let us try to solve

$$(C - UV^T)\lambda = c.$$

Multiplication from the left with C^{-1} gives

$$\lambda - C^{-1}U(V^T\lambda) = C^{-1}c. \qquad \text{D.1}$$

Now the idea needed is to see that we may express $V^T\lambda$ in known quantities. Thus multiplying with V^T from the left we have

$$V^T\lambda - V^T C^{-1}U(V^T\lambda) = V^T C^{-1}c$$

so that

$$V^T\lambda = (I_r - V^T C^{-1}U)^{-1} V^T C^{-1}c$$

where I_r is the rxr identity matrix. Inserting this in D.1 we obtain Woodbury's formula

---

[25] In order to see this assume that $B:R^n \to R^n$ and dim(Range(B)) = r. Take any nxr matrix U such that the columns of U spans Range(B), denote the columns of U by u_j. Since $\{u_j\}$ spans Range(B), $B\lambda = \Sigma b_j(\lambda)u_j$ for any λ and it is easy to see that the the b_j:s must be linear functionals. It follows by Riesz theorem that there exist vectors v_j such that $b_j(\lambda)=v_j\lambda^T$ for all j in [1,r]. Since the u_j:s are linearly independent the kernel of B, N(B) is equal to the orthogonal complement of the span of the v_j denoted $\text{span}[\{v_j\}]^{\perp}$. Hence dim(N(B))=n - dim(span[{v_j}]) and since it is always true for linear mappings that dim(N(B)) + dim(Range(B)) = n it follows that dim(span[{v_j}]) = r i.e that the vectors v_j are linearly independent.

$$\lambda = \mathbf{C}^{-1}\mathbf{c} + \mathbf{C}^{-1}\mathbf{U}\left(\mathbf{I}_r - \mathbf{V}^T\mathbf{C}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}^T\mathbf{C}^{-1}\mathbf{c} \qquad \text{D.2}$$

or equivalently

$$\left(\mathbf{C} - \mathbf{U}\mathbf{V}^T\right)^{-1} = \mathbf{C}^{-1}\left[\mathbf{I}_n + \mathbf{U}\left(\mathbf{I}_r - \mathbf{V}^T\mathbf{C}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}^T\right].$$

A necessary and sufficient condition for the inverse of C' to exist is that C itself together with the rxr matrix $\mathbf{I}_r - \mathbf{V}^T\mathbf{C}^{-1}\mathbf{U}$ are nonsingular.

APPENDIX E FITTING PLANES TO ZONE BOUNDARY

In [Andersson et al, 1989, Table 4.3] the estimated intersection with the boundaries of zone2 are given for the boreholes at the Finnsjö site. These estimates are only based on conductivity measurements and therefore differ from the results in [Ahlbom et al, 1988], where also other methods of localizing the zone boundaries, such as radar and salinity measurements, are taken into account. We use the conductivity measurement based method, because it is the trend of conductivity we wish to model for the zone.

The upper and lower zone boundaries are estimated by fitting planes to the intersections using linear regression. Using only one plane for each of the two boundaries gives a poor fit, however.

Instead the boundaries are modelled with two different planes each. One plane contains all the borehole except the peripheral KFI07 and a second hole. This second hole is chosen so as to maximally reduce the sum of the square residual. The borehole found by this criterion was KFI06 for the upper boundary and HFI01 for the lower.

For KFI07 and the KFI06 we construct two planes that intersect the holes in the positions specified. The observed and calculated values of intersection for upper and lower boundaries are given in the following tables.

Borehole	Observed z	Regression z
BFI01	210	207
BFI02	168	163
HFI01	80	73
KFI05	98	103
KFI09	82	88
KFI11	191	193

Table E.1 *Upper boundary of zone2 excluded holes KFI06 and KFI07 regression expression: $z = 538.8 - 0.259x - 9.235*10^{-2}y$. Here (x, y, z) denote the local grid coordinate system and in particular z is m. b. s. l.* [26]

[26]Note that the the hole KFI10 does not participate in the calculation due to an flaw in the communication between the author and one of his coworkers. However this do probably not influence the results in any larger extent.

Borehole	Observed z	Regression z
BFI01	321	317
BFI02	245	260
KFI05	204	193
KFI06	241	243
KFI09	153	156
KFI11	305	301

Table E.2 *Lower boundary of zone2 excluded holes HFI01 and KFI07 regression expression z = 725. 6 - 0. 3367x - 8. 947*10^{-2}y. Here (x, y, z) denotes the local grid coordinate system and in particular z is m. b. s. l.*

Note that in the actual inference procedure, the geometry in the . zone and . zonex files is defined so that, for the holes KFI06 and KFI07, use the planes intersecting the holes in the specified position rather than the planes derived by linear regression for **both** the upper and lower zone boundaries. This means that HFI01 should participate and KFI06 should be excluded from the regression calculation for the lower zone boundary. This error cannot, however, corrupt the results significantly for the following reasons:

In [Andersson et al, 1989] Table 4. 3 the lower intersection with the zone is specified to 125+ m indicating at least 125 m. Using the regression equation above for the lower boundary gives the value 148 m. Furthermore, due to great variation of the measurement section lengths in this hole, the regularization has succeeded only for sections above the upper boundary, so no part within the zone participates in the actual calculation.

Excluding KFI06 from regression calculation for the lower boundary of the zone just leads to slight modifications of the figures for the regression z in the table above. The greatest difference is for BFI01 where the estimated value changes by less that 3 m.

APPENDIX F REGRESSION ON FINNSJÖ DATA.

A simple regression analysis has been made on data from the Finnsjön site. Measurements was regularized to 6,18,36 and 48m. In the 6 and 18 m case data from all drill holes were used but in the 36 m and 48 m case the two meter measurements from kfi05 was excluded.

The trend funtions used are of two types, potential and exponential, for each of which ordinary and iterative generalized regression was performed. The covariance functions presented here and used in the iterative generalized regression is isotropic and of spherical type.

Potential trend funktion

The trend function is of type

$$\log(K) = \beta_0 - \beta_1 \log(\beta_3 + z) + \beta_2 I\{(x, y, z) \in zone\ 2\}$$

with z being positive downward (m.b.s.l). Alternatively which is the reason for the name of the trend type

$$K = \begin{cases} b_0 b_2 (b_3 + z)^{-b_1} & \text{if } (x, y, z) \in zone\ 2 \\ b_0 (b_3 + z)^{-b_1} & \text{if } (x, y, z) \notin zone\ 2 \end{cases}$$

with

$$\begin{cases} b_0 = 10^{\beta_0} \\ b_1 = \beta_1 \\ b_2 = 10^{\beta_2} \\ b_3 = \beta_3 \end{cases}$$

The regression performed with this trend function is not complete in that the parameter β_3 is set to the value 40.0 which is approximately the least value to keep ($\beta_3 + z$) strictly greater than zero. The inclusion of β_3 as an ordinary regression parameter would lead to a nonlinear regression problem. This generalization is of course highly desirable.

Results 6 m

Ordinary regression

Regression parameters (figures within parenthesis are excluding 3m measurements in kfi05):

β0 = -2.96(-3.08)
β1 = 2.06(2.01)
β2 = 1.19(1.05)
β3 = 40.0

Iterative generalized regression

Spherical type isotropic covariance: Variance = 1.82, Range = 40.49
Spherical type isotropic residual covariance: Variance = 1.33, Range = 14.5
Regression parameters:
β0 = -3.02
β1 = 2.02
β2 = 0.903
β3 = 40.0

Results 18 m

Ordinary regression

Regression parameters:
β0 = -2.13
β1 = 2.25
β2 = 1.51
β3 = 40.0

Iterative generalized regression

Spherical type isotropic covariance: Variance = 1.91, Range = 195.7
Spherical type isotropic residual covariance: Variance = 1.33, Range = 53.9
Regression parameters:
β0 = -1.82
β1 = 2.31
β2 = 0.261
β3 = 40.0

Results 36 m

Ordinary regression

Regression parameters (figures within parenthesis is from excluding all measurements in kfi05):
β0 = -1.95(-2.16)
β1 = 2.23(2.16)

β2 = 1.58(1.48)
β3 = 40.0

Extremal points on the 95% confidence ellipsoid:

Extremals for β0		Extremals for β1		Extremals for β2	
2.40	-6.31	2.38	-6.28	-2.28	-1.63
4.03	0.44	4.04	0.42	2.18	2.28
1.48	1.68	1.54	1.62	2.87	0.29

Important note: Since the derivation of a confidence ellipsoid rests entirely on the validity of the assumed model the above figures are of no value from a statisticians point of view since the assume model in ordinary regression is that the residuals are uncorrelated.

Iterative generalized regression

Spherical type isotropic covariance: Variance = 1.74(1.63), Range = 267.4(203.9)
Spherical type isotropic residual covariance: Variance = 1.25(1.22), Range = 106.2(87.4)
Regression parameters:
β0 = -2.07(-2.08)
β1 = 2.07(2.09)
β2 = 0.074(0.090)
β3 = 40.0

Extremal points on the 95% confidence ellipsoid:

Extremals for β0		Extremals for β1		Extremals for β2	
2.43	-8.46	2.35	-8.37	-3.02	-3.00
3.94	-0.56	3.98	-0.59	1.70	1.69
0.025	0.026	0.026	0.025	0.26	-0.21

Note in particular that this results implies that the hypothesis: "The contrast parameter β2 is zero." cannot be rejected on the 95% level.

Results 48 m

Ordinary regression

Regression parameters:
β0 = -1.89
β1 = 2.22

$\beta 2 = 1.64$
$\beta 3 = 40.0$

Iterative generalized regression

Spherical type isotropic covariance: Variance = 1.68, Range = 309.7
Spherical type isotropic residual covariance: Variance = 1.23 , Range = 182.4
Regression parameters:
$\beta 0 = -2.32$
$\beta 1 = 1.93$
$\beta 2 = 0.027$
$\beta 3 = 40.0$

Exponential trend function

The trend function is of type

$$\log(K) = \beta_0 - \beta_1 z + \beta_2 I\{(x, y, z) \in zone\ 2\}$$

or alternatively, which is the reason for the name of the trend type

$$K = \begin{cases} b_0 b_2 b_1^{-z} & \text{if } (x, y, z) \in zone\ 2 \\ b_0 b_1^{-z} & \text{if } (x, y, z) \notin zone\ 2 \end{cases}$$

with

$$\begin{cases} b_0 = 10^{\beta_0} \\ b_1 = \beta_1 \\ b_2 = 10^{\beta_2} \\ b_3 = \beta_3 \end{cases}$$

Results

Ordinary regression

Regression parameters:
β0 = -6.48
β1 = 0.00403
β2 = 1.358

Iterative generalized regression

Spherical type isotropic covariance: Variance = 1.91, Range = 195.7
Spherical type isotropic residual covariance: Variance = 1.32, Range = 52.6
Weighted regression parameters:
β0 = -6.17
β1 = 0.00456
β2 = 0.245

Compilation of results

Potential trend funktion

Ordinary regression

Scale	β_0	β_1	β_2
6m	-2.96	2.06	1.19
18m	-2.13	2.25	1.51
36m	-1.95	2.23	1.58
48m	-1.89	2.22	1.64

Iterative generalized regression

	Nonresidual		Residual				
Scale	V	R	V	R	β_0	β_1	β_2
6m	1.82	40.5	1.33	14.5	-3.02	2.02	0.903
18m	1.91	195.7	1.33	53.9	-1.82	2.31	0.261
36m	1.74	267.4	1.25	106.2	-2.07	2.07	0.074
48m	1.68	309.7	1.23	182.4	-2.32	1.93	0.027

LIST OF NOTATION

^	Signals that this is an estimate or signals this is a unit vector.
‘	A mean removed form.
*	Kriging estimate.
B	Inverse of the kriging matrix.
β^	An estimate of the parameter vector in regression.
b	Generic vector.
β	The parameter vector in regression.
C	Generic notation for a covariance matrix.
c	Generic notation for the covariance vector
$C(\cdot)$	Covariance function.
D	Kriging matrix.
d	Packer length.
D(**x**)	Kriging neighbourhood.
Δh	Packer test difference head.
Δp	Packer test overpressure.
E(**x**)	Kriging set.
$E[\cdot]$	Expectation value operator.
f(**x**), g(**x**)	A generic function
$F(\cdot)$	A F-distribution function.
ϕ_i	Basis functions for linear regression.
G	The matrix of geometrical anisotropy.
$\gamma(\cdot)$	The semivariogram.
η	Coordinate vector in the isotropic (lag)space.
h(**x**)	The hydraulic potential
J	Joint jacknifing measure.
$K(\cdot)$	The hydraulic conductivity for a certain averaging scale viewed as a stochastic process.

$L(\cdot)$	Lagrangian multiplicator function or Likelihood function
L_i	Packer interval.
M	The number of trends in a linear model.
$m(\cdot)$	Expectation value function.
MRE	Mean reduced error.
MSE	Mean square error.
MSRE	Mean square reduced error.
N	Number of measurements.
n	Normal of the kriging neighbourhoods and sets.
N_{kn}	Number of measurements in the kn:th kriging neighborhood.
o, w	Overlap and width in the definition of kriging neighbourhoods and sets.
Θ	Number of angular lag classes
Q	Packer test flow
q	$Q/\Delta h$
R	Number of radial lag classes
ρ_w	Borehole radius.
s	Generic real.
$\mathbf{U}(\cdot)$	The Darcy velocity for a certain averaging scale viewed as a stochastic process.
$V[\cdot]$	Centered variance operator.
Ω	Event space.
W_i	Weights
X	A stochastic variable.
ξ	Generic notation for a lag vector
x	Generic notation for a point in space.
x^, y^, z^	Basis system.
Y	The stochastic vector obtained by restricting the function $Y(\cdot)$ to a set of spatial points.
y	A realization of Y.

$Y(\cdot)$	The stochastic process defined by $Y(\cdot) = \log(K(\cdot))$. Also used as a generic notation for a stochastic process or a representation of an intrinsic random function.
z	Coordinate corresponding to **z^**.
z^	Normal vector pointing upward.
$Z(\cdot)$	An intrinsic random process.
{x:(·)}	The set of all **x** such that (·)

List of SKB reports

Annual Reports

1977-78
TR 121
KBS Technical Reports 1 – 120
Summaries
Stockholm, May 1979

1979
TR 79-28
The KBS Annual Report 1979
KBS Technical Reports 79-01 – 79-27
Summaries
Stockholm, March 1980

1980
TR 80-26
The KBS Annual Report 1980
KBS Technical Reports 80-01 – 80-25
Summaries
Stockholm, March 1981

1981
TR 81-17
The KBS Annual Report 1981
KBS Technical Reports 81-01 – 81-16
Summaries
Stockholm, April 1982

1982
TR 82-28
The KBS Annual Report 1982
KBS Technical Reports 82-01 – 82-27
Summaries
Stockholm, July 1983

1983
TR 83-77
The KBS Annual Report 1983
KBS Technical Reports 83-01 – 83-76
Summaries
Stockholm, June 1984

1984
TR 85-01
Annual Research and Development Report 1984
Including Summaries of Technical Reports Issued during 1984. (Technical Reports 84-01 – 84-19)
Stockholm, June 1985

1985
TR 85-20
Annual Research and Development Report 1985
Including Summaries of Technical Reports Issued during 1985. (Technical Reports 85-01 – 85-19)
Stockholm, May 1986

1986
TR 86-31
SKB Annual Report 1986
Including Summaries of Technical Reports Issued during 1986
Stockholm, May 1987

1987
TR 87-33
SKB Annual Report 1987
Including Summaries of Technical Reports Issued during 1987
Stockholm, May 1988

1988
TR 88-32
SKB Annual Report 1988
Including Summaries of Technical Reports Issued during 1988
Stockholm, May 1989

1989
TR 89-40
SKB Annual Report 1989
Including Summaries of Technical Reports Issued during 1989
Stockholm, May 1990

1990
TR 90-46
SKB Annual Report 1990
Including Summaries of Technical Reports Issued during 1990
Stockholm, May 1991

1991
TR 91-64
SKB Annual Report 1991
Including Summaries of Technical Reports Issued during 1991
Stockholm, April 1992

Technical Reports

List of SKB Technical Reports 1992

TR 92-01
GEOTAB. Overview
Ebbe Eriksson[1], Bertil Johansson[2], Margareta Gerlach[3], Stefan Magnusson[2], Ann-Chatrin Nilsson[4], Stefan Sehlstedt[3], Tomas Stark[1]
[1]SGAB, [2]ERGODATA AB, [3]MRM Konsult AB
[4]KTH
January 1992

TR 92-02
Sternö study site. Scope of activities and main results
Kaj Ahlbom[1], Jan-Erik Andersson[2], Rune Nordqvist[2], Christer Ljunggren[3], Sven Tirén[2], Clifford Voss[4]
[1]Conterra AB, [2]Geosigma AB, [3]Renco AB,
[4]U.S. Geological Survey
January 1992

TR 92-03
Numerical groundwater flow calculations at the Finnsjön study site – extended regional area
Björn Lindbom, Anders Boghammar
Kemakta Consultants Co, Stockholm
March 1992

TR 92-04
Low temperature creep of copper intended for nuclear waste containers
P J Henderson, J-O Österberg, B Ivarsson
Swedish Institute for Metals Research, Stockholm
March 1992

TR 92-05
Boyancy flow in fractured rock with a salt gradient in the groundwater – An initital study
Johan Claesson
Department of Building Physics, Lund University, Sweden
February 1992

TR 92-06
Characterization of nearfield rock – A basis for comparison of repository concepts
Roland Pusch, Harald Hökmark
Clay Technology AB and Lund University of Technology
December 1991

TR 92-07
Discrete fracture modelling of the Finnsjön rock mass: Phase 2
J E Geier, C-L Axelsson, L Hässler,
A Benabderrahmane
Golden Geosystem AB, Uppsala, Sweden
April 1992