



Technische Universität München

Faculty of Civil, Geo and Environmental Engineering
Engineering Risk Analysis Group
Univ.-Prof. Dr. Daniel Straub

Application of FORM to Hydraulic Tomography

Josef Wildgruber

Master's thesis
for the Master of Science program Civil Engineering

Author:	Josef Wildgruber
Matriculation number:	3600357
Supervisors:	Ph.D. Iason Papaioannou Univ.-Prof. Dr. Daniel Straub
Date of issue:	March 01, 2015
Date of submission:	July 07, 2015



Involved Organisations



Engineering Risk Analysis Group
Faculty of Civil, Geo and Environmental Engineering
Technische Universität München
Arcisstraße 21
D-80333 München

Declaration

With this statement I declare that I have independently completed this Master's thesis. The thoughts taken directly or indirectly from external sources are properly marked as such. This thesis was not previously submitted to another academic institution and has also not yet been published.

München, July 07, 2015

Josef Wildgruber

Josef Wildgruber
Zeislstr. 8
D-85402 Kranzberg
e-Mail: jo.wildgruber@googlemail.com

Abstract

Hydraulic conductivity (K) and storage (S) are ground parameters and of high importance for many tasks in the fields of environmental and civil engineering. Their distribution in the subsurface is highly heterogeneous and their values are not directly measurable. In this work the First Order Reliability Method (FORM) is applied to estimate K and S from observations of the hydraulic head (H). A numerical model is used to represent the groundwater flow in a subsurface domain. K and S are assumed to be described by random fields. To reduce the computational effort, each random field is a Kahunen-Loève-Expansion (KL-Expansion). A finite differences approach determines the sensitivities of H with respect to the Expansion of K and S. With this information a design point for each observation of H is identified. A conditional mean and standard deviation of K and S can then be calculated from the design points. The results for a 2D model of The North Campus Research Site at the University of Waterloo, Ontario, Canada (a test site for hydraulic tomography) show that the FORM approach is able to estimate ground parameters. The quality of the outcome is dependent on the assumed measurement error for the observations of H, the number of observations and the number of standard normal random variable used in the KL-Expansion. For upcoming projects it is recommended to replace the finite differences approach to calculate sensitivities with a more advanced alternative.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Outline	2
2	Probability Theory	3
2.1	Random Variables	3
2.1.1	Continuous Random Variables	4
2.1.2	Random Vectors	6
2.1.3	Normal Distribution	7
2.2	Random Fields	10
2.2.1	Continuous Random Fields	10
2.2.2	Discretization of a Random Field	11
2.2.3	Gaussian Random Fields	12
2.3	The Karhunen-Loève-Expansion	12
2.3.1	KL Expansion of Continuous Random Fields	14
2.3.2	KL-Expansion of Discretized Random Fields	15
3	Linear Parameter Estimation	17
3.1	Numerical Model	17
3.2	Sensitivity Calculation with Finite Differences	17
3.3	Linear Estimation Methods	18
3.3.1	Sequential Linearisation	18
3.3.2	First Order Reliability Method (FORM)	20
4	Application of FORM	25
4.1	The Test Site	25
4.1.1	Geometry	25
4.1.2	Hydraulic Tomography	26
4.2	Finite Differences Model	28
4.3	Assumptions	30
4.3.1	Discrete Gaussian Random Field	30
4.3.2	KL-Expansion	31
4.4	Observations	32
4.5	Implementation of FORM	33
4.6	Results	34
4.6.1	Dependency on Measurement Error	35
4.6.2	Dependency on Number of Observations	37
4.6.3	Dependency on Number of Random Variables	39

5	Summary and Conclusions	43
A	The First Appendix	45
	List of Figures	49
	List of Tables	53
	Bibliography	55

Chapter 1

Introduction

1.1 Motivation

Ground parameters have a high influence on groundwater flows and therefore on the spread of contaminants in the subsurface. As a result, the estimation of ground parameters in order to predict groundwater flows is an important task in engineering risk analysis today. Ground parameters are highly heterogeneous and in most cases the amount of observations available is not exhaustive due to financial and technical restrictions. Dr. T.-C. "Jim" Yeh from the University of Arizona and his colleagues dealt with the problem by performing numerical simulations. This permitted them to perform a sequential linearization and therefore to calculate conditional means and the standard deviations of different ground parameters using measurements of hydraulic heads i.e. the pressure in the subsurface. These measurements are obtained by hydraulic tomography, technique that extracts groundwater from an aquifer at a certain location and measures the hydraulic head response in the subsurface at different observation locations. As it happens, the retrieved data procures information on the behaviour of the subsurface and its properties i.e. the ground parameters. Moreover, Yeh tested his sequential linearization approach with different models [Yeh and Zhang, 1996] [Zhang and Yeh, 1997] [Hughson and Yeh, 1998]. Logically, when analysing three dimensional domains and highly heterogeneous parameters, the computational effort to estimate ground parameters increases.

This work offers to investigate an alternative to the sequential linearization approach. It is based on the First Order Reliability Method (FORM) [Hasofer and Lind, 1973]. The pivot of FORM are the design points which maximise the probability of occurrence of an observed hydraulic head. Unlike the sequential linearization, conditioning in the FORM approach is done only once after the design points have been identified. Therefore the results thus collected are expected to be the best possible linear estimations based on the observations. Since the total number of linearizations is higher for the FORM approach than for the sequential linearization, the models uses a Kahunen-Loeve-Expansion (KL-Expansion) of the ground parameters to reduce computational effort and time. It represents randomness by eigenfunctions and allows to significantly reduce the number of random variables while minimizing the spawned mean square error. Naturally, the quality of the estimation depends on the number of random variables in the KL-Expansion along with the number of observations taken into account.

1.2 Outline

The estimation of ground parameters employing a combination of FORM and KL-Expansion is based on a linear updating of random fields. Hence, the knowledge about the underlying probability theory is an important precondition. The second chapter presents it, starting from the definition of random variables. Afterwards, different kinds of random fields and their properties are exposed. In the end the Kahunen-Loève-Expansion is introduced. A numerical parameter estimation can be performed based on this information. The third chapter introduces two different ways to achieve this based on a numerical model for groundwater flows and the finite differences method for sensitivity analysis. The first one is the state-of-the-art method named sequential linearization approach, the second one is the FORM approach developed in this work. In chapter four, the FORM approach is validated by estimating ground parameters for a 2D model of The North Campus Research Site at The University of Waterloo, Ontario, Canada. For this purpose, a numerical model of the domain has to be created and reasonable assumptions have to be made. In the end the results illustrate the performance of the FORM approach.

Chapter 2

Probability Theory

Chapter two presents the underlying probability theory used for the approximation methods in this work. The first section of this chapter presents the properties of random variables. The second section introduces random fields, sets of random variables, that allow to represent the uncertainties of a multidimensional space. In order to discretize such random fields, the Kahunen-Loève Expansion of a random field by a set of eigenfunctions can be utilized. It is the subject of the third and last section.

2.1 Random Variables

In science, random variables are frequently used to describe uncertain outcomes of experiments. A real random variable X is a function defined on the sample space Ω of an experiment such that there is a probability of the random variable corresponding to each possible outcome; i.e. each event in Ω :

$$X : \Omega \rightarrow \mathbb{R} \quad (2.1)$$

An upper-case letter always denotes the random variable X and the lower-case letter one possible value it may take x .

As a probability function each random variable respects the three axioms of probability:

$$(i) \ Pr[x] \geq 0, \text{ for every } x \in X \quad (2.2)$$

$$(ii) \ Pr[\Omega] = 1 \quad (2.3)$$

$$(iii) \ \text{If } x_1 \in X, x_2 \in X, \text{ and } x_1 x_2 = \emptyset, \text{ then } Pr[x_1 + x_2] = Pr[x_1] + Pr[x_2] \quad (2.4)$$

Where Pr donates the probability of a specific outcome x . The definitions above were taken from [Kottegoda and Rosso, 2008].

This chapter first introduces continuous random variables. Afterwards, the properties of a

vector of random variables are presented. The last part introduces the gaussian or normal distribution.

2.1.1 Continuous Random Variables

A random variable is called continuous if the underlying experiment has an infinite number of possible outcomes i.e. all possible real numbers between two limits. It is fully described by its Probability Density Function (PDF) that can be derived from samples of the random variable. Further information on random variables can be found in [Papaioannou, 2011],[Kottegoda and Rosso, 2008].

Probability Density Function

The PDF is a mathematical function defined over the range of values that a random variable can possibly take. In general, the graph of the PDF takes the form of a continuous curve and maps an infinite number of outcomes of the random variable to function values $f_X(x) \in [0, 1]$.

The probability of an experiment resulting in a value inside an interval $I = [x_1, x_2]$ is obtained by integrating the PDF over I:

$$Pr[x_1 \leq X \leq x_2] = \int_{x_1}^{x_2} f_X(x)dx \leq 1. \quad (2.5)$$

In case of $[x_1, x_2] = [-\infty, \infty]$ this equation is equal to 1.

Cumulative Distribution Function of a Continuous Random Variable

The Cumulative Distribution Function (CDF) is the probability of non-exceedance of a certain outcome value of a random variable.

$$F_X(x) = P(X \leq x) \quad (2.6)$$

It is obtained by integrating the PDF:

$$F_X(x_i) = \int_{-\infty}^{x_i} f_X(x)dx \quad (2.7)$$

Information on the random behaviour can be deduced by obtaining moments of the random variables. The most important moments in practice are the mean or expected value and the

standard deviation.

Mean or Expected Value

The first moment about the origin is called the mean μ_X , or statistical expected value $E(X)$. It gives the central tendency of the random variable and is therefore an important attribute of the phenomenon studied. For continuous random variables it is given by:

$$\mu_X = E[X] = \int_{all x_i} x_i f_X(x_i) \quad (2.8)$$

Additional moments can be derived from the mean value. Moments with respect to μ_X are called central moments and are written as:

$$\mu_r = E[(X - \mu_X)^r] = \sum_{all x_i} (x_i - \mu_X)^r f_X(x_i) \quad (2.9)$$

The most important of them is the variance of a random variable.

Variance and Standard Deviation

The variance of a random variable X is the second central moments and therefore an indicator of the spread i.e. the difference to the mean value. For continuous random variables the variance is obtained by:

$$\mu_2 = E[(X - \mu_X)^2] = \text{Var}_X = \int_{all x_i} (x_i - \mu_X)^2 f_X(x_i) \quad (2.10)$$

The standard deviation σ_X is defined as $\sqrt{\text{Var}}$. In practice it is usually preferred because it has the same units than the random variable. Moments of random variables give information about the shape of the PDF. In case of little knowledge on the random variable, the moments obtained from only a few observations can be used to determine the shape of an assumed probability distribution.

These are the basic properties of continuous random variables. A vector of random variables is a central element to estimate uncertain ground parameters in this work. Such a vector is called a random vector.

2.1.2 Random Vectors

A random vector \mathbf{X} with size n maps a infinite number of outcomes Ω of n random variables to probabilities. \mathbf{X} therefore contains n random variables $\mathbf{X} = [X_1, \dots, X_n]^T$, where $[\cdot]^T$ donates the transpose operator. The properties of random vectors presented in this subsection as well as further explanations can be found in [Papaioannou, 2011]. A random vector \mathbf{X} is completely defined by its joint CDF:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1 \cap \dots \cap X_n \leq x_n) \quad (2.11)$$

The corresponding joint PDF can be obtained by differentiation of the joint CDF:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{\delta^n F_{X_1, \dots, X_n}(x_1, \dots, x_n)}{\delta x_1, \dots, \delta x_n} \quad (2.12)$$

The mean value vector of a random vector \mathbf{X} contains the mean values of its components:

$$\boldsymbol{\mu}_{\mathbf{X}} = [\mu_{X_1}, \dots, \mu_{X_n}]^T \quad (2.13)$$

The linear dependency between two single random variables X_i, X_j is captured by the covariance $Cov(X_i, X_j)$ and the correlation coefficient. The covariance is defined as follows:

$$Cov(X_i, X_j) = E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] \quad (2.14)$$

and the dimensionless correlation coefficient as:

$$\rho(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sigma_i \sigma_j} \quad \rho \in [-1, 1] \quad (2.15)$$

where σ_i and σ_j are the standard deviations of X_i and X_j . For statistical independent random variables $Cov(X_i, X_j)$ and $\rho(X_i, X_j)$ become 0. For a random vector of size n , there are $(n - 1)^2$ covariances. All covariances of a random vector are stored in the covariance matrix $\Sigma_{\mathbf{X}\mathbf{X}}$ defined as:

$$\Sigma_{\mathbf{X}\mathbf{X}} = [Cov(X_i, X_j)]_{n \times n} \quad (2.16)$$

This also applies to the correlation coefficients:

$$\mathbf{R}_{\mathbf{X}\mathbf{X}} = [\rho(X_i, X_j)]_{n \times n} \quad (2.17)$$

A relationship between $\mathbf{\Sigma}_{\mathbf{X}\mathbf{X}}$ and $\mathbf{R}_{\mathbf{X}\mathbf{X}}$ is given by:

$$\mathbf{\Sigma}_{\mathbf{X}\mathbf{X}} = \mathbf{D}_{\mathbf{X}} \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{D}_{\mathbf{X}} \quad (2.18)$$

where $\mathbf{D}_{\mathbf{X}}$ is a diagonal matrix with the standard deviations of the component random variables:

$$\mathbf{D}_{\mathbf{X}} = \text{diag}[\sigma_i]_{n \times n} \quad (2.19)$$

Random vectors describe relations between random variables. They are applied to represent random systems with several uncertain components.

Each random variable follows a certain probability distribution. The gaussian distribution is the central distribution used in this work due to its practicality and is explained in the following subsection.

2.1.3 Normal Distribution

Probability distributions describe the outcomes of different random processes and to make them mathematically analysable. This subsection introduces one of them, the normal distribution, which has been applied for continuous random variables in this work. Properties of this and other probability distributions can be found in [Kottegoda and Rosso, 2008].

Properties of the Normal Distribution

The normal distribution was originally used to study errors in the outcomes of an experiment repeated under similar conditions. Since the causes of the errors are not easily identifiable, the normal distribution offers a practicable way to represent such errors.

It is specified by two parameters: μ and σ , where μ is the mean, locating the mode, and σ the standard deviation, governing the spread, of the random variable. The notion for the normal distribution is $N(\mu, \sigma)$ and its PDF is given by:

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right], \quad \text{for } -\infty < x < +\infty \quad (2.20)$$

Figures 2.1 and 2.2 illustrate how different values for μ and σ change the form of the distribution. Normally, the CDF of the normal distribution can only be evaluated numerically.

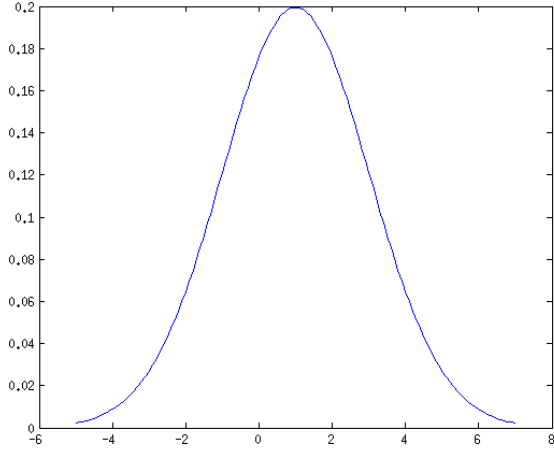


Figure 2.1: PDF of normal distribution with $\mu=1$ and $\sigma=2$

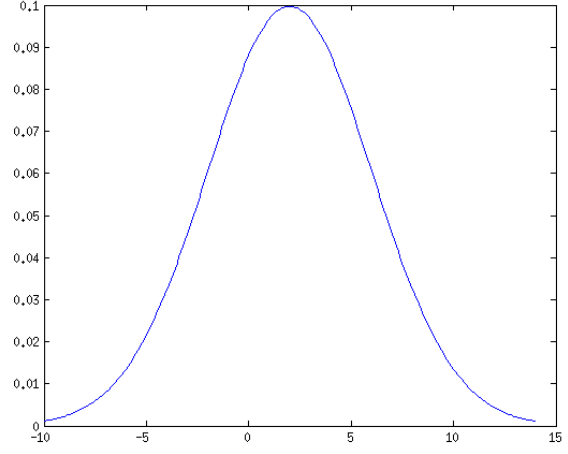


Figure 2.2: PDF of normal distribution with $\mu=2$ and $\sigma=4$

Nevertheless, for the normal distribution, the calculation of the CDF can be simplified by the transformation of the random variable into a standard normal space. Therefore, in practice a standardized curve that transforms the variate X into Z is used. The transformation is:

$$Z = \frac{X - \mu}{\sigma} \quad (2.21)$$

Thus, Z is called a standard normal or $N(0,1)$ variate with PDF given by the following equation:

$$\Phi(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right). \quad (2.22)$$

The CDF of X can then be written as:

$$\begin{aligned} F_X(x) &= Pr[X \leq x] = Pr\left[Z \leq \frac{x - \mu}{\sigma}\right] = Pr[Z \leq z] = \\ &= \Phi(z) = \int_{-\infty}^z \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt, \quad \text{for } -\infty < z < +\infty \end{aligned} \quad (2.23)$$

The properties of the normal distribution can easily be transferred to the multivariate case.

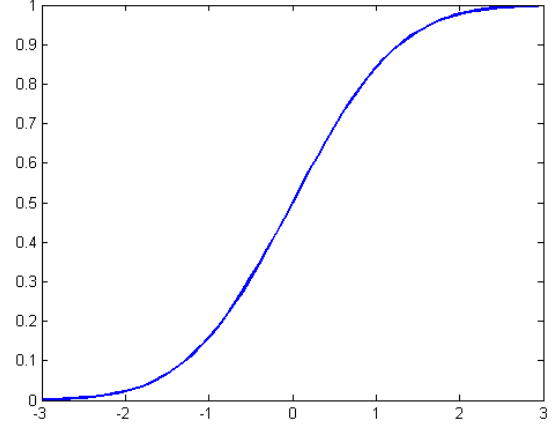
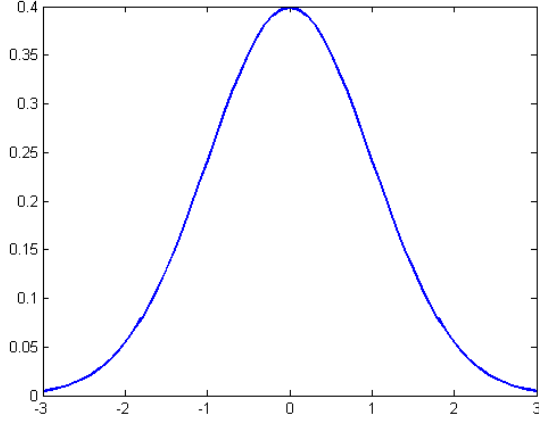


Figure 2.3: PDF of standard normal distribution **Figure 2.4:** CDF of standard normal distribution

Properties of Multivariate Normal Distribution

Resulting from the properties of the normal distribution, a normal random vector \mathbf{X} is completely defined by its mean value vector $\boldsymbol{\mu}_X$ and its covariance matrix $\boldsymbol{\Sigma}_{XX}$. The PDF of a n dimensional normal random vector \mathbf{X} is defined as:

$$F_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi^{n/2}(\det\boldsymbol{\Sigma}_{XX})^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^T \boldsymbol{\Sigma}_{XX}^{-1}(\mathbf{x} - \boldsymbol{\mu}_X)\right] \quad (2.24)$$

In the standard normal case the PDF is expressed on terms of the correlation matrix and the vector \mathbf{Z} of n standard normal random variables $\mathbf{Z} = [Z_1, \dots, Z_n]$:

$$F_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{2\pi^{n/2}(\det\mathbf{R}_{XX})^{1/2}} \exp\left[-\frac{1}{2}\mathbf{z}^T \mathbf{R}_{XX}^{-1}\mathbf{z}\right] \quad (2.25)$$

Furthermore, normal random vectors allow to calculate the unknown mean and covariance function i.e. the conditional distribution, of a subset of random variables \mathbf{X}_1 within the vector, from known outcomes of a subset \mathbf{X}_2 where $\mathbf{X}_1 \cap \mathbf{X}_2 = \emptyset$. This method can be found in [Adler, 1981]. The mean vector and the covariance matrix become:

$$\boldsymbol{\mu}(t) = (\boldsymbol{\mu}_1(t), \boldsymbol{\mu}_2(t)) \quad (2.26)$$

and

$$\boldsymbol{\Sigma}(t) = \begin{bmatrix} \boldsymbol{\Sigma}_{11}(t) & \boldsymbol{\Sigma}_{12}(t) \\ \boldsymbol{\Sigma}_{21}(t) & \boldsymbol{\Sigma}_{22}(t) \end{bmatrix} \quad (2.27)$$

The conditional distribution of \mathbf{X}_1 given \mathbf{X}_2 is also normal with mean:

$$\boldsymbol{\mu}_{1|2}(t) = \boldsymbol{\mu}_1(t) + \boldsymbol{\Sigma}_{12}(t)\boldsymbol{\Sigma}_{22}(t)^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2(t))^T \quad (2.28)$$

and covariance matrix:

$$\boldsymbol{\Sigma}_{1|2}(t) = \boldsymbol{\Sigma}_{11}(t) - \boldsymbol{\Sigma}_{12}(t)\boldsymbol{\Sigma}_{22}(t)^{-1}\boldsymbol{\Sigma}_{21}(t) \quad (2.29)$$

The possibility of calculating a conditional distribution for unknown components of a normal random vector is a central element for the estimation of of a random field.

In the context of this work, continuous normal random vectors are applied to uncertain ground parameters at multiple locations inside some domain of interest. The variation of the subsurface conditions at different locations can be represented by a so-called random field. The next section introduces random fields and how to analyse them.

2.2 Random Fields

This section starts by showing how uncertainties in a multidimensional space can be described by continuous random fields. Afterwards, it focuses on the advantages of discretizing a random field by a finite differences mesh and the benefits of using a gaussian random field for further calculations.

2.2.1 Continuous Random Fields

A random field is a collection of random variables $X(t)$. It means, that $X(t)$ is a random variable for each t , where t is a continuous parameter and $t \in T$. In general $X(t)$ is called a random field if $T = R^2$ or $T = R^3$. In these cases, t can be imagined as a spacial location and $X(t)$ as a random variable describing an uncertain property at this point. In theory each domain contains an infinite number of locations t_i and therefore infinite random variables $X(t_i)$. Then, $x(\mathbf{t})$ is a single realization of the random field $X(\mathbf{t})$ where \mathbf{t} is a vector of locations. The PDF for some $\mathbf{t} \in T$ is written $f_{X(\mathbf{t})}(x, \mathbf{t})$. For further reading on random fields, the reader is referred to [Adler, 1981] and [Papaioannou, 2011]

The random fields of ground parameters in this work can be fully described by its mean function and either its correlation or covariance function. The mean function is:

$$\mu_X(\mathbf{t}) = E[X(\mathbf{t})] = \int_{-\infty}^{\infty} x f_{X(\mathbf{t})}(x, \mathbf{t}) dx \quad (2.30)$$

and the variance function becomes:

$$\sigma_X^2(\mathbf{t}) = E[(X(\mathbf{t}) - \mu_X(\mathbf{t}))^2] = \int_{-\infty}^{\infty} (x(\mathbf{t}) - \mu_X(\mathbf{t}))^2 f_{X(\mathbf{t})}(x, \mathbf{t}) dx \quad (2.31)$$

where σ_X is the standard deviation function. The autocorrelation function can be expressed in terms of:

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X(t_1)X(t_2)}(x_1, t_1; x_2, t_2) dx_1 dx_2 \quad (2.32)$$

where $f_{X(t_1)X(t_2)}(x_1, t_1; x_2, t_2)$ is the joint PDF of the random variables $X(t_1), X(t_2)$. The auto covariance function is then:

$$\begin{aligned} \Gamma_{XX}(t_1, t_2) &= Cov[X(t_1)X(t_2)] = \\ E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))] &= R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) \end{aligned} \quad (2.33)$$

Random fields are often used to describe spacial variation by a set of random variables. In the continuous case, every domain contains an infinite number of random variables. Since the results of this work are based on a finite differences representation of an area of interest i.e. a random field, the latter has to be discretized by a mesh.

2.2.2 Discretization of a Random Field

A discretized random field $X(\mathbf{t})$ is a finite number of random variables with $t \in T = \Omega$. Therefore Ω describes a domain by a non infinite set of locations t_i and corresponding random variables $X(t_i)$. One possible approach to discretize a random field is to represent the random field by a finite differences mesh, where the property value of each element are set to be equal to the property value at its center of gravity t_c . This approach is named the midpoint method. This and other methods are presented in [Papaioannou, 2011]

The total number of random variables applied depends on the mesh used for the discretization since a parameter for a mesh element is represented by one random variable. Hence, a constant value is assigned to each mesh element and there are as many elements in the mesh as there are random variables.

More accurate results can be produced, by increasing the number of elements ergo decreasing their size.

In conclusion, the choice of a discrete random field gives the opportunity to perform numerical simulations of the behaviour of an uncertain real world domain. The outcome of theses simulations are the basis for the estimation methods in this work.

The fact that all random variables are assumed to follow a normal distribution result in ran-

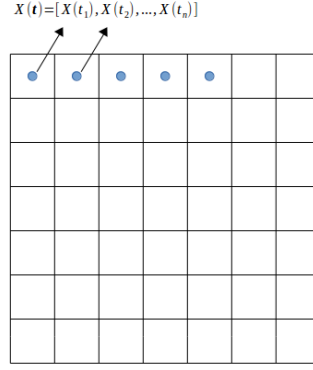


Figure 2.5: Example for midpoint discretization

dom fields with special conditions. They are denoted Gaussian Random Fields and presented in the following.

2.2.3 Gaussian Random Fields

The Gaussian Random Field (GRF) is a special class among them. The convenient analytic form of the multivariate Gaussian density makes it possible to calculate explicit results for problems generated by GRF. For other kinds of random fields this is in general not possible. [Adler, 1981]

Since each random variable within a GRF follows a normal distribution, it can be completely described by its mean function $\mu(t)$ and either its auto covariance function $\Gamma_{XX}(t_1, t_2)$, or its autocorrelation function $R_{XX}(t_1, t_2)$. The marginal PDF of a GRF is given by:

$$f_X(x(t)) = \frac{1}{\sigma_X(t)\sqrt{2\pi}} \exp \left[-\frac{1}{2} \frac{(x(t) - \mu_X(t))^2}{\sigma_X^2(t)} \right] \quad (2.34)$$

The main focus of this work is to approximate uncertain conditions by performing a bayesian updating on a discrete GRF.

As previously mentioned, the number of mesh elements has a great influence on the computational effort needed to run simulations of a discretized random field. In order to get more accurate results, the mesh can be broken down into rather small elements leading to a high number of unknown random variables and therefore to high computational effort. One way to drastically reduce the number of unknown random variables, while conserving a high precision is by describing the random field through its Kahunen-Loève expansion.

2.3 The Karhunen-Loève-Expansion

It is possible to obtain a realization of a Gaussian random field by the means of an eigenfunction representation of the process. In general, this expansion is referred to as the

Karhunen – Loève – Expansion or KL-Expansion. It offers an approximation of a Gaussian random field while maintaining the mean and standard deviation functions. With an increasing number of eigenfunctions, the mean square error is minimized, so the field can be precisely represented even with a small number of unknowns. Explanation of the KL-Expansion can also be found in [Papaioannou, 2011] [Adler, 1981].

In order to explain the KL-Expansion, some notations are required. Let T be a compact interval in \mathbb{R}^n , and $\Gamma(t_1, t_2)$ be a continuous, real valued, non-negative definite function, e.g. the autocorrelation function of a random field. Firstly, the integral equation can be notated as follows:

$$\int_T \Gamma(t_1, t_2) \phi(t_2) dt_2 = \lambda \phi(t_1) \quad (2.35)$$

If,

$$\int_T |\phi(t_2)|^2 dt_2 < \infty \quad (2.36)$$

then the non-zero number λ is called *eigenvalue* and the corresponding function $\phi(t_2)$ is called *eigenfunction*. In general, there is an infinite number of eigenvalues $\lambda_1, \lambda_2, \lambda_3, \dots$ with corresponding eigenfunctions $\phi_1, \phi_2, \phi_3, \dots$. It can be assumed that the eigenvalue sequence is non-increasing and that the eigenfunctions' form is an orthonormal sequence, in such a way that:

$$\int_T \phi_i(t_2) \phi_j(t_2) dt_2 = \begin{cases} 1 & \text{for } i = j. \\ 0 & \text{for } i \neq j. \end{cases} \quad (2.37)$$

According to *Mercer's theorem* [Zaanen, 1960] the autocovariance function $\Gamma(t_1, t_2)$ can be represented exactly by its eigenvalues and eigenfunctions. This works for all continuous, non-negative definite functions on a compact interval $T \times T \subset \mathbb{R}^n$. Then the series:

$$\Gamma(t_1, t_2) = \sum_{j=0}^{\infty} \lambda \phi_j(t_1) \phi_j(t_2), \quad (2.38)$$

converges absolutely and uniformly on $T \times T$. This fact further implies that the error decreases with an increasing number of eigenvalues and eigenfunctions.

$$\lim_{m \rightarrow \infty} \int_{TxT} \left| \Gamma(t_1, t_2) - \sum_{j=0}^m \lambda \phi_j(t_2) \phi_j(t_1) \right|^2 dt_1 dt_2 = 0, \quad (2.39)$$

A Gaussian random field $\hat{X}(\mathbf{t})$ can therefore be represented by its KL-Expansion of the form:

$$\hat{X}(\mathbf{t}) = \mu_X(\mathbf{t}) + \sum_{i=0}^{\infty} \sqrt{\lambda_i} \phi_i(\mathbf{t}) U_i \quad (2.40)$$

where U_i are standard normal random variables.

Using this result, a Gaussian random field can be represented by a rather small number of standard normal random variables. The realization is based on a combination of eigenfunctions, or eigenvectors. Depending on the underlying mesh, both approaches have advantages and disadvantages, which are exposed hereafter.

2.3.1 KL Expansion of Continuous Random Fields

In the continuous case the KL-Expansion uses eigenfunctions to discretize a random field. For most random fields these eigenfunctions have to be calculated numerically by solving the integral equation in (2.37), only for rectangular domains and some auto covariance functions the solution can be obtained analytically, which decreases the necessary computation significantly.

A combination of eigenfunctions describes the variation of a continuous random field in one direction. Hence, for multidimensional domains one set of eigenfunctions per dimension has to be found. The KL expansion for a 3 dimensional continuous random field therefore becomes:

$$\hat{X}(x, y, z) = \mu_X(x, y, z) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sqrt{\lambda_i} \phi_i(x) U_i \sqrt{\lambda_j} \phi_j(y) U_j \sqrt{\lambda_k} \phi_k(z) U_k. \quad (2.41)$$

It is obvious that the number of combinations of eigenfunctions is high even for only a few eigenfunctions per dimension since their number has to be taken to the cube. Also the quality of the approximation with this method is equal for the whole domain. Therefore, it does not allow to define areas of interest with higher resolution i.e. use of a larger number of eigenfunctions. In praxis one has to decide if it makes sense to use this approach. The criteria should be the number of dimensions of the domain of interest and if some areas within the domain are of higher interest than others. An alternative for multidimensional random fields, discretized by heterogeneous meshes is shown in the following.

2.3.2 KL-Expansion of Discretized Random Fields

Instead of solving the integral equation (2.35) or getting an analytical solution of the eigenfunctions, the KL-Expansion can be performed on discretized random fields with eigenvectors and eigenvalues of the covariance-matrix Σ . Therefore an eigenvalue problem of the following form has to be solved:

$$\Sigma \nu_i = \lambda_i \nu_i \quad \text{with } i=1, \dots, n \quad (2.42)$$

The KL-Expansion in the discretized case becomes:

$$\hat{X} = \mu_X + \sum_{i=0}^{\infty} \sqrt{\lambda_i} \nu_i U_i \quad (2.43)$$

The advantage of using eigenvectors instead of eigenfunctions for the KL expansion is the possibility of defining areas of interest inside a discrete random field by refining the mesh. If the covariance between the mesh elements is based on their distance from each other, the eigenvectors of the resulting covariance matrix inherit this information which leads to a more accurate approximation for precisely meshed areas. Furthermore, the eigenvectors do not have to be calculated for each dimension. Therefore higher order eigenvectors can still be taken into account, which leads to higher precisions. The disadvantage of this approach is the calculation of the eigenvalue problem. With increasing number of random variables i.e. mesh elements, the complexity of eigenvalue solvers increases exponentially. Therefore, the computational effort often rapidly exceeds the available resources.

The KL expansion is a powerful tool that allows to drastically reduce the number of random variables needed in order to obtain realizations of a GRF. This is useful since the computational effort of random field updating methods highly depend on the number of unknowns. The choice of using eigenfunction or eigenvalues in the KL expansion should be made after analysing the GRF and deciding if a discretized field can be used or refined areas of interest have to be taken into account.

In the second chapter discretizing a random field permits the running of numerical simulations. Combined with a KL-Expansion it is capable of efficiently calculate realizations of uncertain properties in a real world domain.

The goal is to simulate a realization that behaves like a real world domain. This would allow to make predictions for different scenarios. It can be achieved by gathering information on a domain of interest and adjust the numerical simulation based on them. In chapter 3 two methods for linear estimation of uncertain ground parameters are presented.

Chapter 3

Linear Parameter Estimation

Before the methods to update a gaussian random field, chapter 3 introduces in its first section a numerical model to simulate groundwater flows dependent on gaussian random fields and in its second section the finite differences method to obtain the sensitivities of a discretized random field, in other words to linearize it. This makes possible the adjustment of a random field, so that it matches observations taken from a corresponding real world domain. Afterwards, the second section puts forth two different approaches to achieve this.

3.1 Numerical Model

One way to calculate groundwater flows is given by the flow equation:

$$\nabla[K(x)\nabla H] + Q(x) = S_s(x)\frac{\partial H}{\partial t} \quad (3.1)$$

Where $K(x)$ is the hydraulic conductivity of the soil, $S_s(x)$ is the storage of the soil, $H(x)$ is the hydraulic heads and $Q(x)$ is the flux. For discretized domains, a numerical solution for H given K , S_s and Q exists.

The ground parameters K and S_s are supposed to be unknown in this work. Therefore they are replaced by random variables. As a result, the distributions of K and S_s in a discretized domain are discretized random fields $K(\mathbf{t})$ and $S_s(\mathbf{t})$, and H is a function of them. The sensitivities capture how $K(\mathbf{t})$ and $S_s(\mathbf{t})$ change H . One way to calculate them is presented in the following.

3.2 Sensitivity Calculation with Finite Differences

Sensitivity analysis determines the influence of input parameters on the outcome of a function. For the linear estimation of ground parameters, the relationship between H and $K(\mathbf{t})$ (H and $S_s(\mathbf{t})$) is of special interest. Since the procedure is the same for $K(\mathbf{t})$ and $S_s(\mathbf{t})$, a general notation $P(\mathbf{t})$ will refer to a random field for a ground parameter in the logarithmic scale ($\log K(\mathbf{t})$ and $\log S(\mathbf{t})$ respectively).

The finite differences method is a straight forward method to calculate sensitivities. First each random variable $P(t_i)$ in $P(\mathbf{t})$ is perturbed by dt_i . The differences in H with respect to one perturbed $P(t_i)$ divided by dt_i gives the sensitivity.

$$\nabla H(P(t_i)) = \frac{\Delta H(P(t_i))}{\Delta t_i} = \frac{H(P(t_i + dt_i)) - H(P(t_i))}{dt_i} \quad (3.2)$$

For a linearization, one sensitivity with respect to each random variable has to be calculated. For n random variables this results in $n + 1$ solutions for $H(P(\mathbf{t}))$. The finite differences approach is computational expensive on the one hand side. On the other, the advantage of this method is an easy implementation since the source code of the solver remains unchanged.

The sensitivities of a random field are the key element of linear updating methods and affect their efficiency and accuracy. Their understanding is necessary before explaining these methods. Their explanation is the objective of the next section.

3.3 Linear Estimation Methods

This second section exposes two linear methods to approximate discrete gaussian random fields:

- the sequential linearization method, the state-of-the-art method for estimating soil parameters;
- and the First Order Reliability Method (FORM), method originally developed for structural reliability analysis that has been applied to soil parameters estimation in this work.

3.3.1 Sequential Linearisation

Jim Yeh from the University of Waterloo was the first to apply the sequential linearization method to estimate uncertain subsurface conditions [Zhang and Yeh, 1997]. Its attempt is to estimate the grounds parameters from observations of the hydraulic head. A solution is found based on a numerical model for groundwater flows and a linearization of the outcome of the hydraulic heads. Conditional mean values and standard deviation of a gaussian random field representing a ground parameters $P(\mathbf{t})$ are found in each iteration step. The iteration stops, once the outcome of $H(P(\mathbf{t}))$ matches the observations.

Linearisation Procedure

Initially, $P(\mathbf{t})$ is set to a constant mean vector $\boldsymbol{\mu}_{Pprior}$. The function of the hydraulic heads $H(P(\mathbf{t}))$ can be linearised at the point $\boldsymbol{\mu}_{Pprior}$ by the First-Order-Taylor-Series-Expansion:

$$H(\boldsymbol{\mu}_{Pposterior}) \approx H(\boldsymbol{\mu}_{Pprior}) + \nabla H(\boldsymbol{\mu}_{Pprior})^T * (\boldsymbol{\mu}_{Pposterior} - \boldsymbol{\mu}_{Pprior}) \quad (3.3)$$

where $\nabla H(\boldsymbol{\mu}_{Pprior})$ is the gradient of $H(P(\mathbf{t}))$ at the point $\boldsymbol{\mu}_{Pprior}$. $\boldsymbol{\mu}_{Pposterior}$ is obtained

by solving the inverse equation: hydraulic head \hat{H} :

$$\boldsymbol{\mu}_{P\text{posterior}} \approx \boldsymbol{\mu}_{P\text{prior}} + \nabla H(\boldsymbol{\mu}_{P\text{prior}})^{-T} * (H(\boldsymbol{\mu}_{P\text{posterior}}) - H(\boldsymbol{\mu}_{P\text{prior}})) \quad (3.4)$$

where $\nabla H(\boldsymbol{\mu}_{P\text{prior}})^{-1}$ is the inverse of the gradient. Observations \hat{H} can be included by setting:

$$H(\boldsymbol{\mu}_{P\text{posterior}}) = \hat{H} \quad (3.5)$$

Then $\boldsymbol{\mu}_{P\text{posterior}}$ is an estimation of $\boldsymbol{\mu}_P$ that in the end produces a model outcome equal to \hat{H} . The inverse equation is equal to a Bayesian Updating of a gaussian random field.

Bayesian Updating

The bayesian updating approach explained in chapter 2 can be applied here. The covariance matrix of the field therefore can be written as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{P,P} & \boldsymbol{\Sigma}_{P,\hat{H}} \\ \boldsymbol{\Sigma}_{\hat{H},P} & \boldsymbol{\Sigma}_{\hat{H},\hat{H}} \end{bmatrix} \quad (3.6)$$

Since \hat{H} and P are related through the sensitivities, the covariance matrix can be expressed as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{P,P} & \nabla H(P)^T \boldsymbol{\Sigma}_{P,P} \\ \boldsymbol{\Sigma}_{P,P} \nabla H(P) & \nabla H(P)^T \boldsymbol{\Sigma}_{P,P} \nabla H(P) \end{bmatrix} \quad (3.7)$$

A conditional mean can be calculated by (2.29)(2.30):

$$\boldsymbol{\mu}_{P\text{posterior}} = \boldsymbol{\mu}_{P\text{prior}} + \nabla H(\boldsymbol{\mu}_{P\text{prior}})^T \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{P\text{prior}}, \boldsymbol{\mu}_{P\text{prior}}} \boldsymbol{\Sigma}_{H(\boldsymbol{\mu}_{P\text{prior}}), H(\boldsymbol{\mu}_{P\text{prior}})}^{-1} (\hat{H} - H(\boldsymbol{\mu}_{P\text{prior}}))^T \quad (3.8)$$

where $\boldsymbol{\mu}_{P\text{posterior}}$ is the updated mean and the updated covariance matrix of P becomes:

$$\boldsymbol{\Sigma}_{P,P'} = \boldsymbol{\Sigma}_{P,P} - \nabla H(P)^T \boldsymbol{\Sigma}_{P,P} \boldsymbol{\Sigma}_{\hat{H},\hat{H}}^{-1} \nabla H(P) \boldsymbol{\Sigma}_{P,P} \quad (3.9)$$

For a perfectly linear system the method would deliver this solution in one step. For the non-linear case the solution is found iteratively by setting the obtained $\boldsymbol{\mu}_{P\text{posterior}}$ as the new $\boldsymbol{\mu}_{P\text{prior}}$ in each iteration step.

3.3.2 First Order Reliability Method (FORM)

FORM is a method originally used in the field of structural reliability and was applied by [Hasofer and Lind, 1973], [Rackwitz and Fiessler, 1978] and [Der Kiureghian et al, 2005]. A so-called limit state function $G(\mathbf{U})$ is used to describe the situation where the capacity of a structure $C(\mathbf{U})$ with respect to different parameter \mathbf{U} and the demand D , i.e. an applied load, are equal. Hence, the limit state function separates the failure domain F with higher D than $C(\mathbf{U})$ from the non-failure domain, where $C(\mathbf{U})$ exceeds D . The problem is expressed in the standard normal space. If the capacity and the demand are not standard normal distributed, they have to be transformed to the standard normal space. In the standard normal space the point on the limit state function with minimum distance to the origin is called design point \mathbf{u}^* . It is therefore the point with the highest failure probability. The distance between the origin and the design point is the reliability index denoted with β . Figure 3.1 illustrates these objects positions in the standard normal probability space for $\mathbf{U} = [U_1, U_2]$.

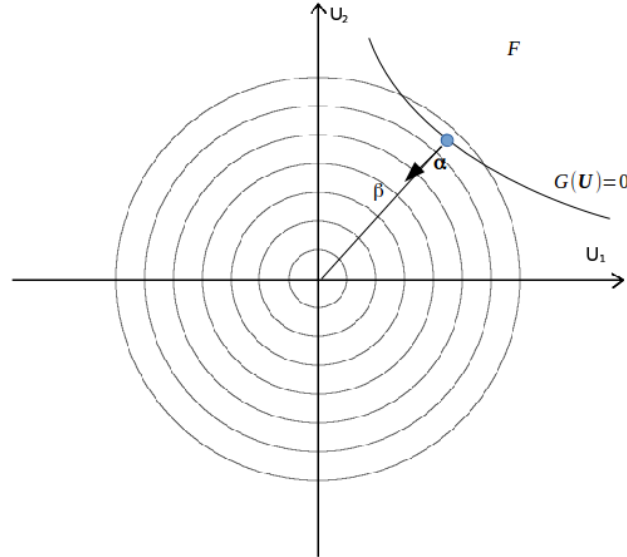


Figure 3.1: Design point location in the standard normal space

The design point can be identified by linearizing the system multiple times.

FORM applied to hydraulic tomography

In this work, the ability of FORM to estimate ground parameters is investigated. The goal is to show the advantages of FORM compared with the state-of-the-art method, i.e. the sequential linearization. In the context of updating ground parameters a different definition of the limit state function is needed. In this work, ground parameters are represented by a KL-Expansion with m standard normal random variables \mathbf{U} . A limit state function is defined to be the surface in the standard normal space where the outcome of a finite difference model $H(\mathbf{U})$ at a certain location, i.e. the hydraulic head are equal to an observation of the hydraulic head at this location \hat{H} :

$$G(\mathbf{U}) = \hat{H} - H(\mathbf{U}) \quad (3.10)$$

It is easy to see that a linearization of G at some point is the negative linearization of H at the same point. The design point is the point with the highest probability for the model outcome $H(\boldsymbol{\mu}_U)$ to be equal to the observation \hat{H} . In the end, a conditioning based on design points gives the most likely realization of \mathbf{U} that matches each observation as good as possible. The following linear equation describes the relationship between the limit state function and the means of the variables \mathbf{U} :

$$G(\boldsymbol{\mu}_{U\text{posterior}}) \approx G(\boldsymbol{\mu}_{U\text{prior}}) + \nabla G(\mathbf{u}^*)^T * (\boldsymbol{\mu}_{U\text{posterior}} - \boldsymbol{\mu}_{U\text{prior}}) \quad (3.11)$$

With (3.11) it becomes

$$\hat{H} - G(\boldsymbol{\mu}_{U\text{posterior}}) \approx G(\boldsymbol{\mu}_{U\text{prior}}) + \nabla G(\mathbf{u}^*)^T (\boldsymbol{\mu}_{U\text{posterior}} - \boldsymbol{\mu}_{U\text{prior}}) \quad (3.12)$$

The limit state function can be expressed by setting $G(\boldsymbol{\mu}_{U\text{posterior}}) = 0$. Therefore (3.13) changes to:

$$G(\boldsymbol{\mu}_{U\text{prior}}) + \nabla G(\mathbf{u}^*)^T (\boldsymbol{\mu}_{U\text{posterior}} - \boldsymbol{\mu}_{U\text{prior}}) = 0 \quad (3.13)$$

In this case $\boldsymbol{\mu}_{U\text{posterior}}$ is a point on the limit state surface, $\nabla G(\mathbf{u}^*)^T$ is the gradient vector at the design point and $G(\boldsymbol{\mu}_{U\text{prior}})$ equals the difference in the observed hydraulic head \hat{H} and the calculated hydraulic head for $\boldsymbol{\mu}_{U\text{prior}}$. For the inverse updating (3.13) becomes:

$$\boldsymbol{\mu}_{U\text{posterior}} = \boldsymbol{\mu}_{U\text{prior}} - (\nabla G(\mathbf{u}^*)^T)^{-1} G(\boldsymbol{\mu}_{U\text{prior}}) \quad (3.14)$$

In the standard normal space $\nabla G(\mathbf{u}^*)^T$ and $G(\boldsymbol{\mu}_{U\text{prior}})$ are expressed by $\boldsymbol{\alpha}$ and β . Both are introduced in the next part.

Linearization Procedure

Before $\boldsymbol{\mu}_U$ can be updated, a design point for each observation has to be identified. Since the limit state function is supposed to be non-linear, this has to be done by solving an optimization problem of the form:

$$\mathbf{u}_{i+1}^* = \mathbf{u}_i^* + \nabla H(\mathbf{u}_i^*)^T (\hat{H} - H(\mathbf{u}_{i+1}^*)) \quad (3.15)$$

where \mathbf{u}_{i+1}^* is an approximation of the design point. In the standard normal space this can be written as:

$$\mathbf{u}_{i+1}^* = \boldsymbol{\alpha}_i \beta_i \quad (3.16)$$

where α_i is the normalized gradient of the limit state function evaluated at the design point:

$$\alpha_i = \frac{\nabla G(\mathbf{u}_i^*)}{\|\nabla G(\mathbf{u}_{i+1}^*)\|} \quad (3.17)$$

and β_i is the distance from the origin to \mathbf{u}_{i+1}^* . According to fundamental geometry this equals the function value at the origin divided by the norm of the gradient vector:

$$\beta_i = \frac{G(0)}{\|\nabla G(\mathbf{u}_i^*)\|} = \frac{G(\mathbf{u}_i^*) - \nabla G(\mathbf{u}_i^*)^T \mathbf{u}_i^*}{\|\nabla G(\mathbf{u}_i^*)\|} = \frac{G(\mathbf{u}_i^*)}{\|\nabla G(\mathbf{u}_i^*)\|} \alpha_i^T \mathbf{u}_i^* \quad (3.18)$$

This state-of-the-art procedure to find a design point is named iHLRF-Algorithm [Sudret and Der Kiureghian, 2000]. Given the design points, a conditional mean vector $\mu_{U_{posterior}}$ and covariance matrix can be calculated in one step by solving the inverse problem.

Inverse Bayesian Updating

For the conditioning, $\mu_{U_{prior}}$ is a standard normal vector with all entries set to the mean value 1. The covariance matrix can be divided into four parts.

$$\Sigma = \begin{bmatrix} \Sigma_{U,U} & \Sigma_{U,y} \\ \Sigma_{y,U} & \Sigma_{y,y} \end{bmatrix} \quad (3.19)$$

where

$$y = \alpha^T * U \quad (3.20)$$

It can also be written as:

$$\Sigma = \begin{bmatrix} \Sigma_{U,U} & A^T \Sigma_{U,U} \\ \Sigma_{U,U} A & A^T \Sigma_{U,U} A \end{bmatrix} \quad (3.21)$$

Where $\Sigma_{U,U}$ is the identity matrix \mathbf{I} and the matrix \mathbf{A} contains the vectors α_n of each observation $[1, \dots, n]$. Hence, the covariance matrix of the design points is the dot product of the matrix \mathbf{A} with itself:

$$\Sigma_{y,y} = A^T \bullet A \quad (3.22)$$

And the cross covariances contain \mathbf{A} :

$$\Sigma_{U,y} = \Sigma_{y,U}^T = \mathbf{A} \quad (3.23)$$

The posterior mean of the random variables is then calculated in one step by (3.14) by setting $y=\beta$:

$$\mu_{U|posterior} = \mu_{U|prior} - \Sigma_{y,U} \Sigma_{y,y}^{-T} \beta \quad (3.24)$$

and the conditional covariance becomes:

$$\Sigma_{U,U'} = I - I\mathbf{A}[\mathbf{A}^T I \mathbf{A}]^{-T} \mathbf{A}^T I \quad (3.25)$$

The third chapter introduced the sequential linearization and the FORM to update a discrete gaussian random field based on observations within the field. Both are iterative, linear methods and need to obtain the sensitivities of the system multiple times. These can be determined by a finite differences approach, an expensive but easy to implement way to do so. The FORM approach was developed in this work as an alternative to the sequential linearization method to estimate ground parameters. In the forth chapter, FORM is applied to an example created with information from a real world domain.

Chapter 4

Application of FORM

This chapter verifies FORM's ability to estimate soil parameters by applying it to a 2D domain with Monte Carlo simulated values for the hydraulic conductivities (K) and the storage (S). The geometry of the 2D domain, its functionality, and the information on the distribution of K and S correspond to the North Campus Research Site at the University of Waterloo. Before going deeper in the explanation of the application of FORM, the first section of this chapter gives an overview of the domain. Thereafter, the second section expounds how a finite differences mesh was derived from the real world situation. Finally, the last two sections present the assumptions that had to be made, the performed experiments and the results.

4.1 The Test Site

The North Campus Research Site (NCRS) is located on the University of Waterloo (UW) campus, in Waterloo, Ontario, Canada. The geology of the region was highly influenced by the Wisconsin glacial period that ended approximately 10.000 years ago. Investigations on the geology beneath the UW started in 1979. Over the years, research projects on the NCRS disclosed the highly heterogeneous character of the subsurface. For detailed information on the experiments, the reader can consult [Karrow, 1979], [Karrow, 1993] or [Sebol, 2000].

Sophisticated instrumentation was constructed on the NCRS to gather information on the ground parameters through hydraulic tomography. Results can be found in [Alexander et al, 2011] and [Berg, 2011].

In order to present the relevant NCRS's properties for this work, this section first depicts the geometry of the area and then expounds the hydraulic tomography.

4.1.1 Geometry

On the NCRS 9 wells are installed in a square shaped area of 15m x 15m. Each of them is around 18m deep and allows either an observation of the hydraulic heads at different depths (CMT) or both observation of the hydraulic heads and extracting (injecting) groundwater at different depth (PW). Figure 4.1 lays out their location at the NCRS.

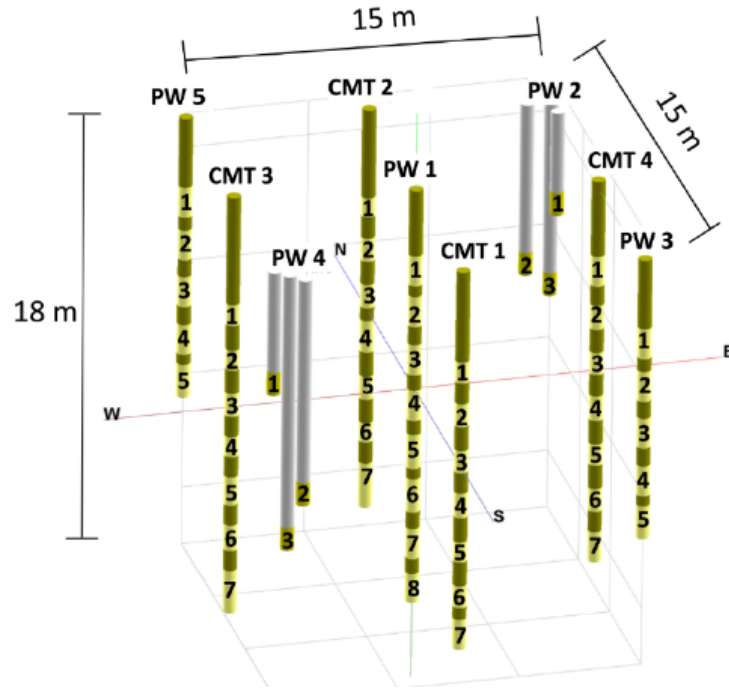


Figure 4.1: Well locations at NCRS

[Alexander et al, 2011] procures a more detailed description of the wells, nonetheless for this work only their locations are of interest.

The set-up allows to perform hydraulic tomography and, consequently, to glean information on the subsurface from its results.

4.1.2 Hydraulic Tomography

In the course of a pumping test, water is extracted or injected from the ground at a PW. The pumping is always applied at a certain depth. Afterwards, the effects on the hydraulic heads in the domain can be observed at the other well locations. An injection of water at a PW forces the hydraulic heads in the area to increase, whereas an extraction leads to their decline. In this example, water is extracted from different wells during a given period of time.

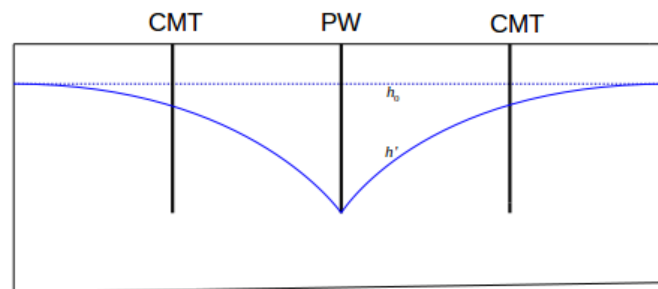


Figure 4.2: Hydraulic head change due to a pumping test

The differences between the initial values of the hydraulic heads and the values during and after the realization of the pumping is called the drawdown (ddn):

$$ddn(x, t) = h_0(x, t) - h'(x, t) \quad (4.1)$$

Performing a pumping test over a certain period of time (transient) brings about a so-called drawdown curve that can be obtained at an observation location. Since the difference in initial and observed hydraulic head increases, the drawdown curve escalates over time.

After a certain period, the drawdown reaches a so-called steady-state. It means that the surrounding aquifer manages to compensate for the water extracted through the well by maintaining a flux equal to the pumping rate. In this case, the hydraulic heads, and therefore the drawdowns, remain constant.

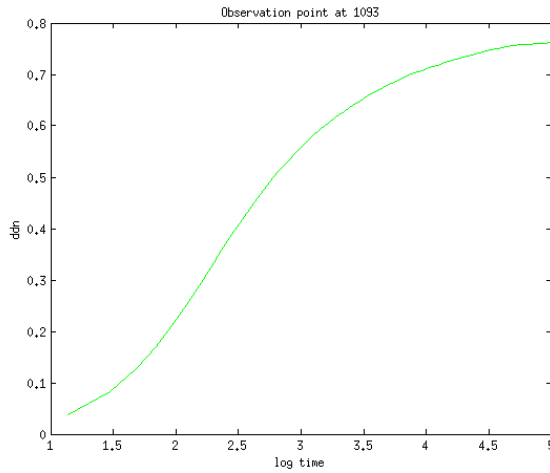


Figure 4.3: Example for drawdown curve (7.5m distance from PW)

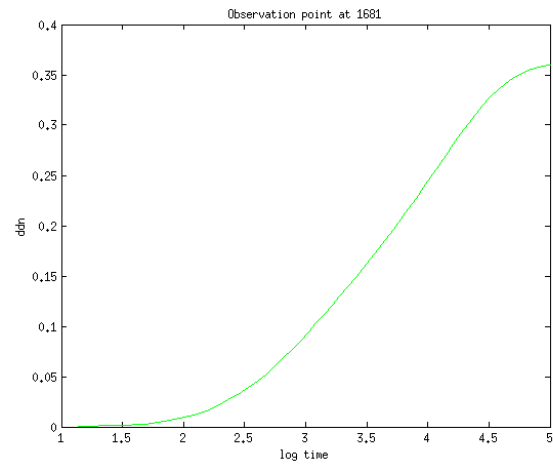


Figure 4.4: Example for drawdown curve (~ 21m distance from PW)

Figure 4.3 and 4.4 show drawdown curves at two different observation locations for one pumping test. The shape of the drawdown curve contains information on the soil parameters K and S inside the domain.

Hydraulic Conductivities

The hydraulic conductivities affect the drawdown curve by changing the steady-state. In fact, low conductivity areas request high water pressure to compensate for a gradient in the fluxes due to a pumping test and that leads to a low drawdown. On the other hand, high conductivity values cause a high drawdown. The drawdown curves therefore contain information on the hydraulic conductivities in a domain. The quality of this information varies within the domain, as it depends on the locations of the wells for pumping and observation. The area between the wells has the most influence on the drawdown, so more precise information on it can be derived from the curve.

Storage

The storage affects the drawdown curve by shifting it in time. Low storage areas allow water to traverse them faster than high storage areas. The consequence is a faster response at the observation location. Therefore, storage has a high influence on the shape and the position of the drawdown curve but does not affect the values. Similar to the hydraulic conductivities, the information on the storage derived from a curve is best for the area between a pumping well and an observation well.

This section gave an overview of the NCRS. Furthermore, it showed the influence of soil parameters on the drawdown and how to obtain information on them by performing pumping tests. The FORM approach uses this information to linearly update the mean and standard deviation of $\log K$ and $\log S$. The next section then presents a method to numerically simulate pumping tests at the NCRS.

4.2 Finite Differences Model

The FORM approach is based on numerical simulations of pumping tests at the NCRS. A flow equation has to be solved for each simulation. It is given by:

$$\nabla[K(x)\nabla H] + Q(x_p) = S_s(x)\frac{\partial H}{\partial t} \quad (4.2)$$

Where $Q(x_p)$ is the flux created by a pumping at a pumping location x_p and H is the outcome of the hydraulic heads in the domain.

In order to simulate pumping tests numerically, the domain has to be discretized. This was realized in a 2D, square shaped finite differences mesh with rectangular elements. The mesh is illustrated in Figure 4.5.

The pumping and observation wells were modelled in the central part of the mesh, where their positions correspond to the ones at the NCRS. This part of the mesh has been refined since the quality of the information on K and S obtained here is the highest.

The mesh has 47 rectangular elements per dimension and 2209 elements in total, but for the validation of the FORM approach only the area of interest in Figure 4.6. is taken into account.

The boundary was set to maintain a constant head value, meaning that a change of the fluxes inside the domain i.e. a pumping test, does not affect the hydraulic heads at the boundary. Therefore, the distance between the boundary and the area of interest has an impact on the result of a pumping test, as it forces the result to satisfy the fixed hydraulic heads at the boundary. Logically, this impact weakens when the distance between the boundary and the area of interest increases. The final dimensions were obtained by expanding the mesh until the solution of the finite differences model i.e. the drawdown due to a pumping test stopped varying.

In conclusion, the finite differences model can be utilized for the validation of the FORM approach applied to hydraulic tomography. The goal of this work is to determine mean values

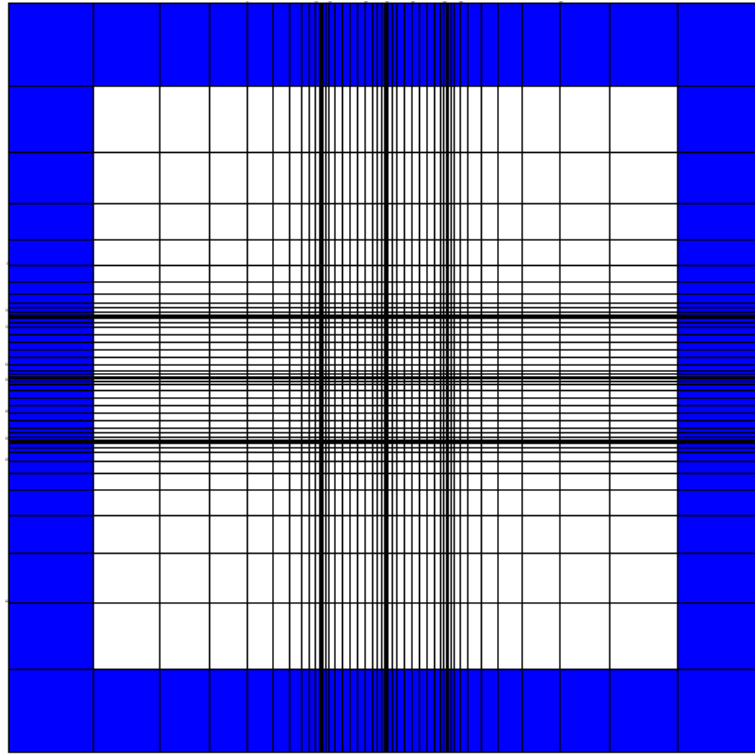


Figure 4.5: Finite Differences Mesh of NCRS

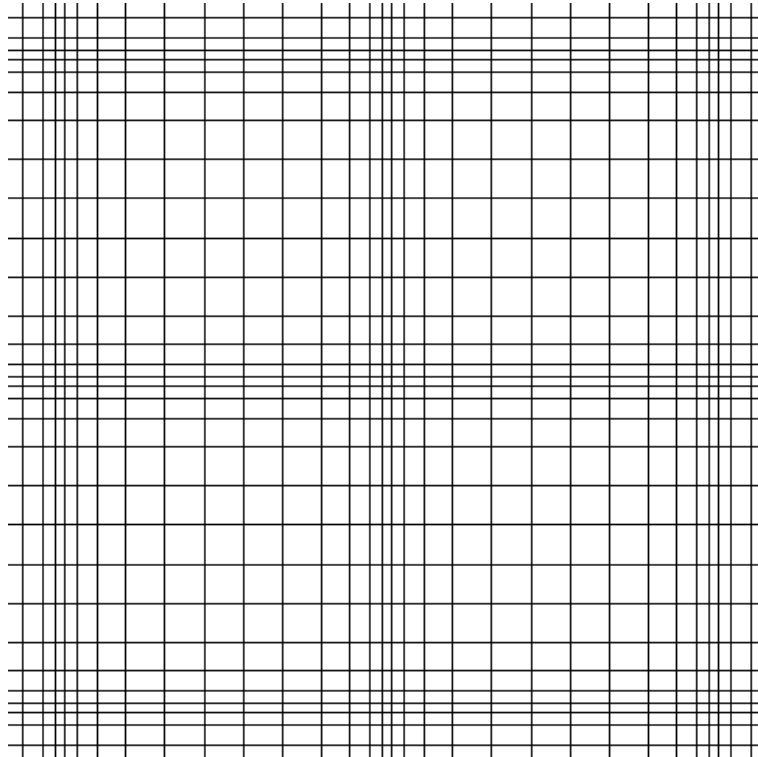


Figure 4.6: Finite Differences Mesh of Area of Interest

and standard deviations for $\log K$ and $\log S$ for each mesh element so that a simulated pumping test matches observations made during a pumping test with the same mesh and Monte Carlo simulated values for $\log K$ and $\log S$. The next section therefore introduces assumptions that are necessary to make the FORM results comparable.

4.3 Assumptions

The FORM estimation of $\log K$ and $\log S$ for the finite differences mesh depends on some assumptions. Before everything, these unknown parameters had to be defined as random variables and a reasonable relationships between them had to be chosen. Afterwards, the resulting random field could finally be discretized through a KL-Expansion.

4.3.1 Discrete Gaussian Random Field

First of all, $\log K$ and $\log S$ were assumed to be Gaussian distributed random variables for the center of each mesh element. As a result, this makes the finite differences mesh a discretized Gaussian random field. The general means and standard deviations of $\log K$ and $\log S$ for the whole domain were derived from observations at the NCRS. Furthermore, a reasonable correlation between the mesh elements has been supposed.

Mean and Standard Deviation

Measurements of K were taken and analysed at the NCRS. Information on $\log K$ was obtained by Slug Tests, Falling Head Parameter Tests and pumping tests, and by estimating it through the grain size distribution. A detailed explanation of these tests can be found in [Alexander et al, 2011]. Based on their results a mean and standard deviation of $\log K$ in the whole domain were identified and the University of Waterloo provided them to facilitate their use in this work. Table 4.1 shows their values.

	$\log K$
μ	-4.6779
σ	1

Table 4.1: Assumed mean and standard deviation of $\log K$

$\log S$ can only be observed by performing pumping test and getting accurate results is very difficult [Alexander et al, 2011]. Consequently, the same mean value and standard deviation of $\log S$ used for the sequential linearization method at the University of Waterloo were conjectured.

	$\log S$
μ	-4
σ	1

Table 4.2: Assumed mean and standard deviation of $\log S$

These informations are included to the FORM approach so that they make sure that the estimation respects the basic properties of the NCRS.

Correlation

Furthermore, the correlation function of the Gaussian random field is set to follow an exponential function of the form:

$$\text{Corr}(e_1, e_2) = \exp^{-\frac{|e_1 - e_2|}{l}} \quad (4.3)$$

Where $|e_1 - e_2|$ is the distance between the two mesh elements e_1 and e_2 and l is the correlation-length, an indicator that demonstrates how far the value at a certain mesh element affects the elements around it.

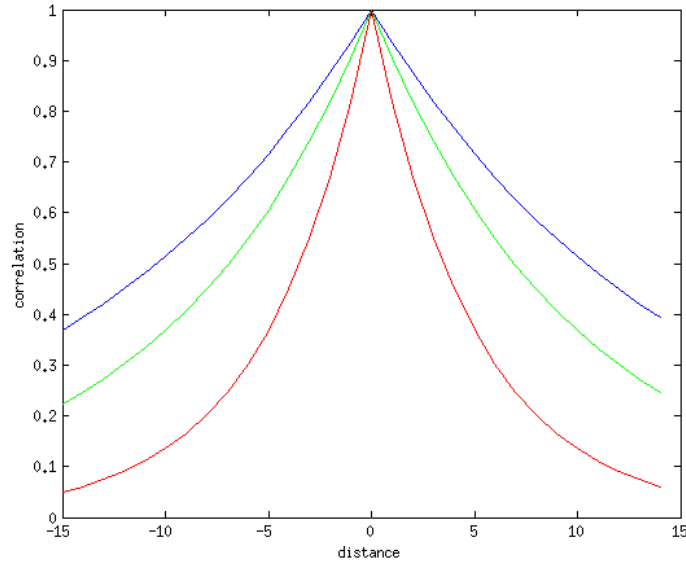


Figure 4.7: Realization of (4.3) for $l=5$ (red), $l=10$ (green) and $l=15$ (blue)

A correlation length of 15m has to be assumed For the FORM validation. It is equal to the dimensions of the area of interest and causes all elements within to be sufficiently dependent on each other. With smaller correlations the randomness of $\log K$ and $\log S$ in the domain would increase making it harder to estimate it. Furthermore, $\log K$ and $\log S$ are expected to be correlated in the same way.

4.3.2 KL-Expansion

Instead of simulating values for $\log K$ and $\log S$ in each mesh element, each of the two fields is created using a KL-Expansion with the first 75 eigenvalues and eigenvectors of the correlation matrix calculated by (4.3). Together with the mean values and standard deviations determined in section 4.3.1, the KL-Expansions of $\log K$ and $\log S$ are obtained by (2.4.1).

This allows a realization of the random fields that satisfies the assumed means, standard deviations and correlation and reduces the total number of random variables from 4418 to 150 at the same time. The relationship between the standard normal random variables in the KL-Expansion and the ground parameters is shown in Figure 4.8.

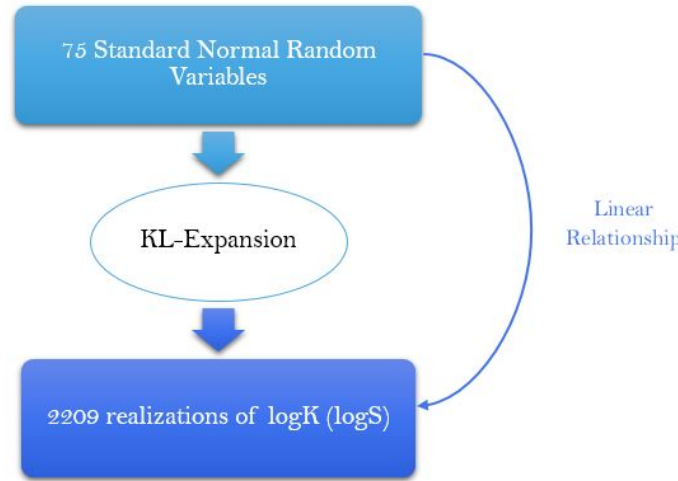


Figure 4.8: Linear Relationship between standard normal random variables and $\log K(\log S)$

These assumptions permit the FORM approach to estimate means and standard deviations for $\log K$ and $\log S$ throughout the finite differences mesh with reduced computational effort. The procedure depends on observations of the drawdown. The next section shows how to obtain them.

4.4 Observations

Observations on the systems behaviour have to be taken as targets for the FORM approach. 108 observations of the drawdown due to 8 different pumping locations are considered. The pumping rate of each pumping test is set to be $-0.001 \frac{m^3}{s}$. Figure 4.9 illustrates the numbering of the well locations in the area of interest. We are assuming that all of them can be used for pumping and observing. This is not true for the real domain at the NCRS but allows us to perform more pumping test in the simulated case.

From each drawdown curve three observations are picked as targets for the FORM approach. It is advised to choose points that best describe the shape of the curve, e.g. points with maximum first or second order derivative. All observations can be found in detail in appendix A. With the stated assumptions and observations the FORM approach can be run. Its implementation is illustrated in the following.

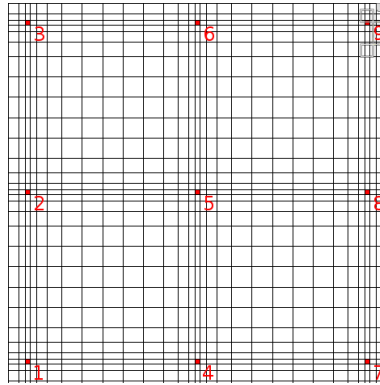


Figure 4.9: Numbering of locations for pumping and observation in the area of interest

4.5 Implementation of FORM

The FORM approach was implemented in Python. It is based on multiple solutions of the finite differences model. In the code, this is done by calling the software MODFLOW-USG. Figure 4.10 shows the relationship between the standard normal random variables used in the KL-Expansion and the model outcome of the hydraulic heads.

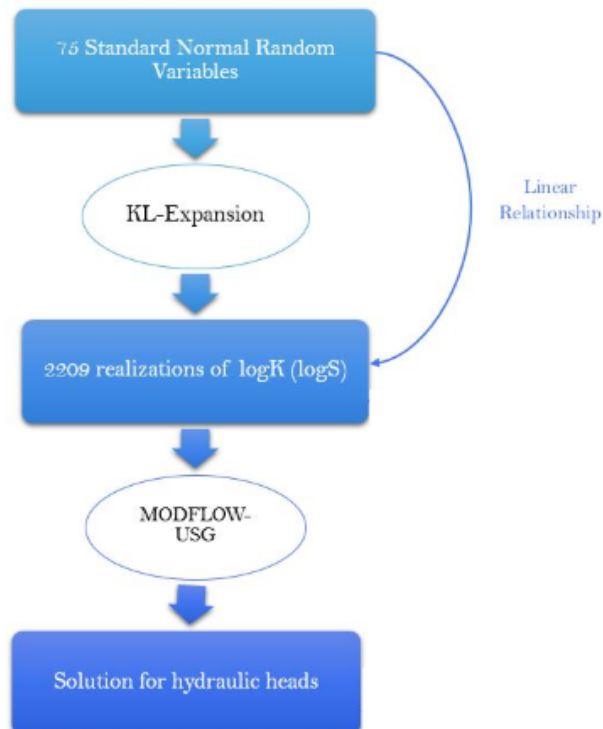


Figure 4.10: Linear Relationship between standard normal random variables and the hydraulic heads

A design point for an observed drawdown is determined by linearizing the model multiple times with respect to the standard normal random variables. The linearization procedure is based on finite differences and requires $m+1$ model runs.

Afterwards a conditioning on the design point gives a realization of the m standard normal random variables that matches the observations the best. In the end, means and standard deviations for $\log K$ and $\log S$ can be calculated from them using the KL-Expansion again. Figure 4.11 shows how this procedure was implemented in Python.

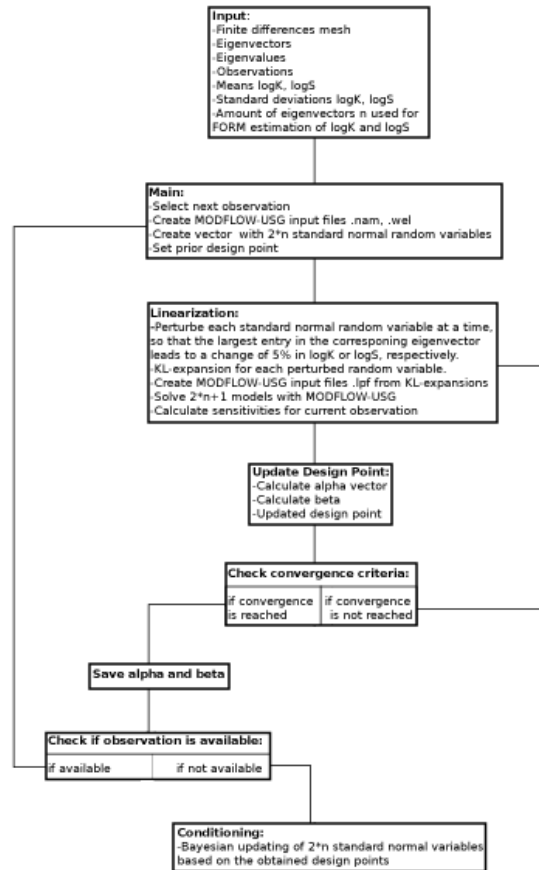


Figure 4.11: Structure of FORM implementation

The implementation was validated by a numerical example. The results are presented in the next part.

4.6 Results

The FORM approach was tested by applying it to the mesh introduced in 4.2 with simulated values for $\log K$ and $\log S$. This was done by creating outcomes for 300 standard random variables. 150 of them were used in the KL expansion of $\log K$ and the other 150 in the KL expansion of $\log S$. The result for the area of interest is presented in Figures 4.12 and 4.13

For the FORM approach only 75 standard random variables were taken into account for the

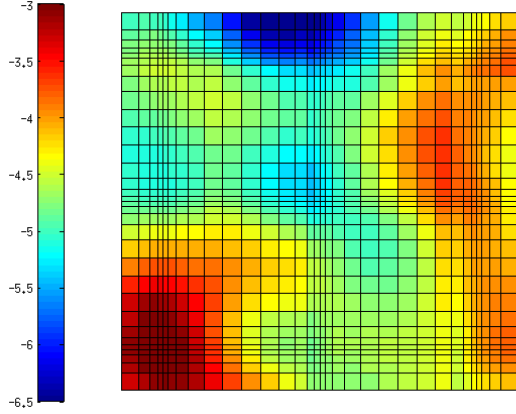


Figure 4.12: Monte Carlo Simulation of logK inside the area of interest

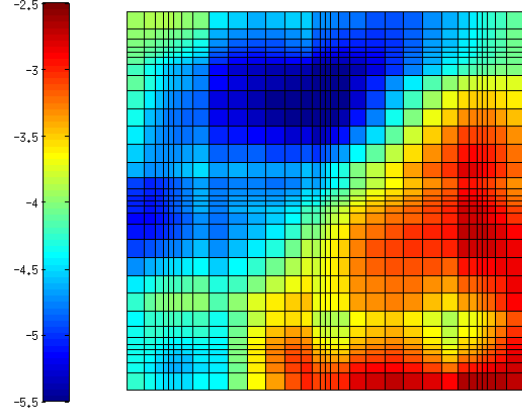


Figure 4.13: Monte Carlo Simulation of logS inside the area of interest

KL expansion of logK and logS respectively. Therefore FORM won't be able to completely represent the randomness of the model, which is the case for real world domains.

The results show the ability of FORM to estimate the means and standard deviations dependent on the number of observations, the number of eigenvectors applied to the KL expansion of the estimation i.e. random variables to be updated by FORM, and the assumed measurement error.

4.6.1 Dependency on Measurement Error

A measurement error introduces an uncertainty to the observations, i.e. the observations are assumed not to be perfect. This was done by adding a diagonal matrix ε with constant values to Σ_{u^*, u^*} in (3.21):

$$\Sigma_{u^*, u^*} = \Sigma_{u^*, u^*} + \text{diag}(\varepsilon) \quad (4.4)$$

Regarding ε is necessary for real world observations. In this example the observations taken from the simulation are actually perfect. Nevertheless, a measurement error is added in order to prevent Σ_{u^*, u^*} from becoming a singular matrix due to redundant information in the observations. The following Figures illustrates how different values for ε affect the FORM estimation of the means.

For $\varepsilon = 0$, the covariance matrix is almost singular because of redundant information in the observations (Figure 4.14 and 4.17). Then, the inversion creates large values for the updated means of logK and logS that do not match the Monte Carlo realization. On the other hand, a rather high error of 0.1 flattens the means by adding variability (Figure 4.15 and 4.18). A good estimation of the simulated values was produced by assuming ε to be 0.0075 (Figure 4.16 and 4.19).

The updated standard deviation of logK and logS increases with larger measurement errors

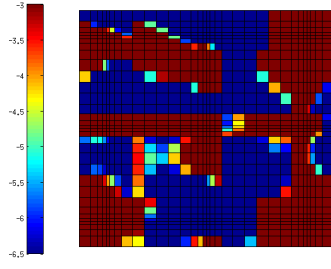


Figure 4.14: FORM solution of the mean of $\log K$ for $\varepsilon=0$

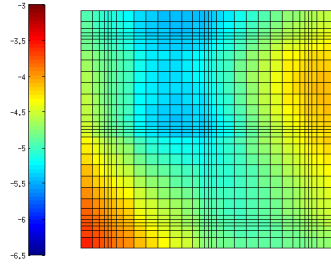


Figure 4.15: FORM solution of the mean of $\log K$ for $\varepsilon=0.1$

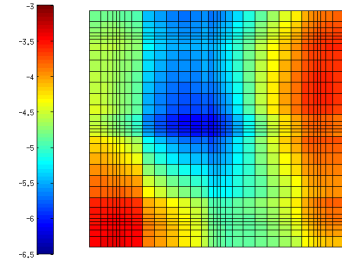


Figure 4.16: FORM solution of the mean of $\log K$ for $\varepsilon=0.0075$

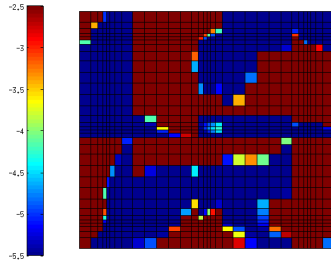


Figure 4.17: FORM solution of the mean of $\log S$ for $\varepsilon=0$

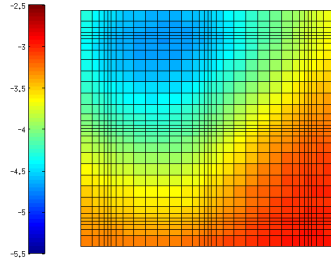


Figure 4.18: FORM solution of the mean of $\log S$ for $\varepsilon=0.1$

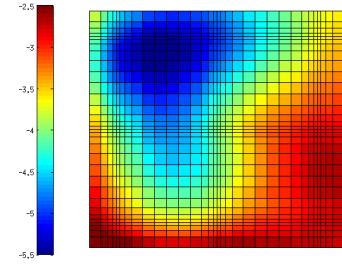


Figure 4.19: FORM solution of the mean of $\log S$ for $\varepsilon=0.0075$

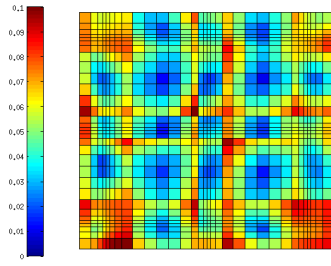


Figure 4.20: FORM solution of the standard deviation of $\log K$ for $\varepsilon=0$

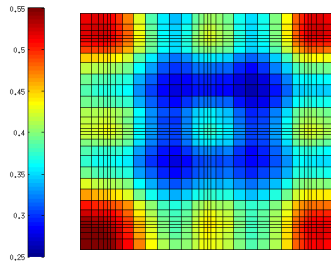


Figure 4.21: FORM solution of the standard deviation of $\log K$ for $\varepsilon=0.1$

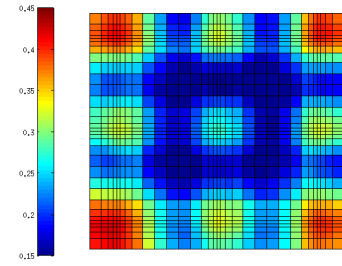


Figure 4.22: FORM solution of the standard deviation of $\log K$ for $\varepsilon=0.0075$

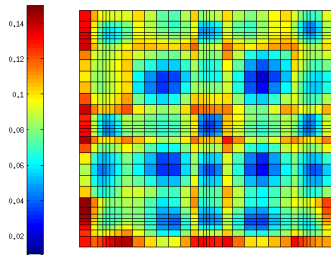


Figure 4.23: FORM solution of the standard deviation of $\log S$ for $\varepsilon=0$

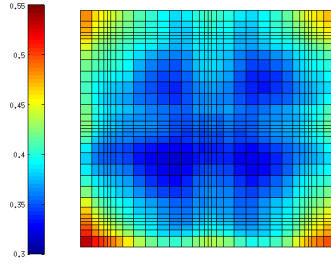


Figure 4.24: FORM solution of the standard deviation of $\log S$ for $\varepsilon=0.1$

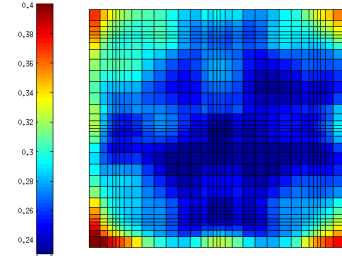


Figure 4.25: FORM solution of the standard deviation of $\log S$ for $\varepsilon=0.0075$

due to additional uncertainty (Figures 4.20-4.25). The standard deviation is higher in the

refined areas around the locations for pumping and observation. The reason might be that each observation contains most information on the area between itself and the corresponding pumping well. Larger elements in this area have more influence on the observed drawdown. Therefore their variability could be limited.

The computational effort stays the same for different values of ε .

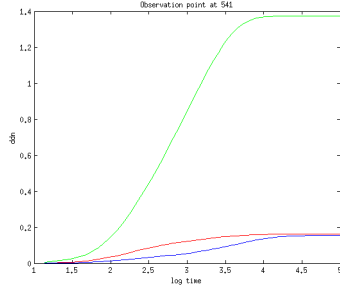


Figure 4.26: Drawdowns for initial means (green), simulated random variables (red) and FORM estimation (blue)

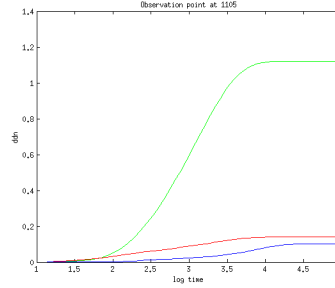


Figure 4.27: Drawdowns for initial means (green), simulated random variables (red) and FORM estimation (blue)

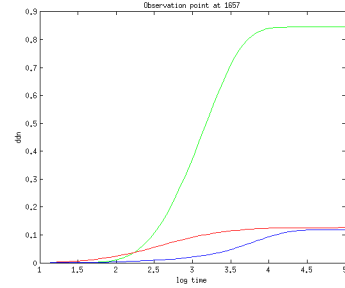


Figure 4.28: Drawdowns for initial means (green), simulated random variables (red) and FORM estimation (blue)

The drawdown curves in Figure 4.26, Figure 4.27 and Figure 4.28 are due to a pumping at location 1 and observations at locations 2, 5 and 7 in this order. The curves for the FORM estimation are shifted to the right what means, that the response occurs later than in the target case. This indicates a higher storage that actually can be observed in Figure 4.19 in the area around location 1.

Since a ε of 0.0075 is necessary in order to get reasonable mean values, it is applied in the following to analyse the influence of the number of observations and eigenvectors.

4.6.2 Dependency on Number of Observations

Each observation of the drawdown contains information on the unknown distribution of $\log K$ and $\log S$. In the following, three approximations of the Monte Carlo realization were calculated by the FORM approach with different amounts of observations n . The following table shows the three different cases. All observations are listed in appendix A for further information.

Table 4.3: Amounts of Observations

	Observations
Case 1	2,5,8,11,14,17,20,23,26,29
Case 2	1-50
Case 3	1-108

Table 4.4: Cases with different amounts of observations

The influence of an increasing n on the estimation of $\log K$ and $\log S$ is illustrated next.

Since case 1 only regards 8 observations due to a pumping at location 1 and 2 observations due to a pumping at location 2, the inversion affects the mean of $\log K$ in lower left area

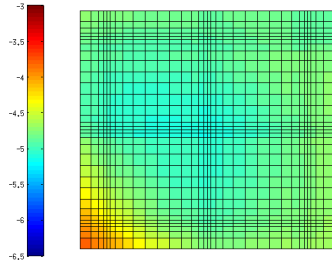


Figure 4.29: FORM solution of the mean of $\log K$ for $n=10$

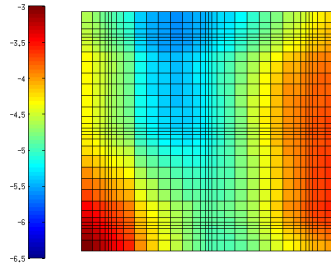


Figure 4.30: FORM solution of the mean of $\log K$ for $n=50$

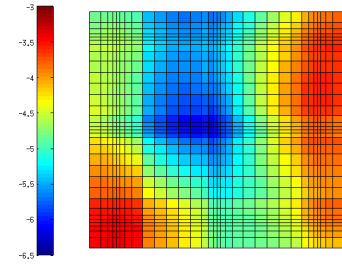


Figure 4.31: FORM solution of the mean of $\log K$ for $n=108$

the most (Figure 4.29). Increasing the number of observations then leads to more accurate results (Figure 4.30 and 4.31)

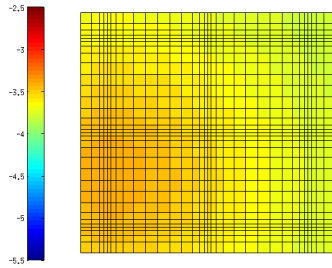


Figure 4.32: FORM solution of the mean of $\log S$ for $n=10$

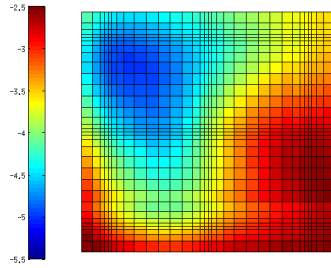


Figure 4.33: FORM solution of the mean of $\log S$ for $n=50$

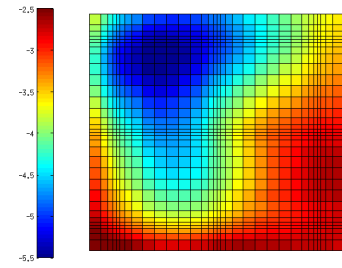


Figure 4.34: FORM solution of the mean of $\log S$ for $n=108$

In case 1, only one observation per location is given. Since $\log S$ is a variable of time, the information derived from each observations is 0. Nevertheless, a small effect can be seen on the left (Figure 4.32). This might be due to the observations taken from location 3 and 4 due to pumping test at location 1 and 2. Like for $\log K$, the accuracy increases for more observations (Figure 4.33 and 4.34).

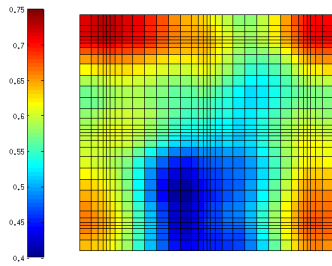


Figure 4.35: FORM solution of the standard deviation of $\log K$ for $n=10$

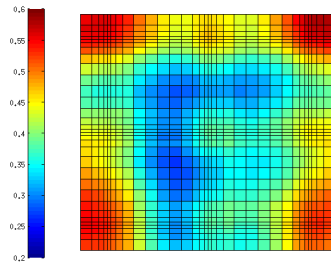


Figure 4.36: FORM solution of the standard deviation of $\log K$ for $n=50$

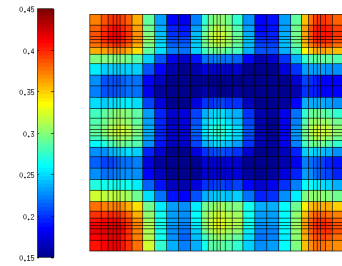


Figure 4.37: FORM solution of the standard deviation of $\log K$ for $n=108$

The updated standard deviation of $\log K$ decreases since more observations introduce more information (Figure 4.35-4.37). Case 1 gives most accurate results for the lower part of the domain. This is reasonable since both pumping tests in case 1 observe the drawdown at location 4 therefore more information is available in this area. More observations then lead to lower standard deviations.

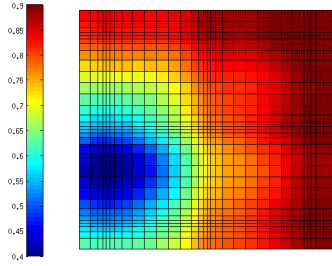


Figure 4.38: FORM solution of the standard deviation of $\log S$ for $n=10$

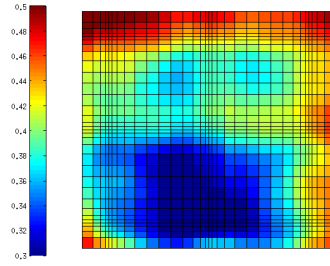


Figure 4.39: FORM solution of the standard deviation of $\log S$ for $n=50$

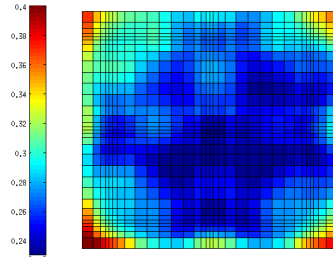


Figure 4.40: FORM solution of the standard deviation of $\log S$ for $n=108$

The standard deviation of $\log S$ shows similar behaviour then the one for $\log K$ (Figure 4.38-4.40). For case 1 two available observations at location 3 and 4 allow the derivation of information on $\log S$. The area between 1 and 2 is therefore the one with the lowest standard deviations. In general the variability of the standard deviation of $\log S$ throughout the mesh is smaller than for $\log K$.

The computational effort increases with the number of observations. In general, four linearizations had to be performed per observation to identify the design point. Each linearization with finite differences has to solve the model $n+1$ times where n is the number of random variable in the KL expansion. Table 4.5 contains the computational efforts of the FORM approach with 150 random variables for different amounts of observations. The code ran on a Intel(R) Xeon(R) CPU E5-2650 with 2.6GHz. Solving the model once took 0.5 sec.

	Number of Model Runs per Observation	Number of Observations	Computation time in s
Case 1	604	10	3020
Case 2	604	50	15100
Case 3	604	108	32616

Table 4.5: Computational effort for cases with different amounts of observations

4.6.3 Dependency on Number of Random Variables

The number of random variables updated by the FORM approach corresponds to the number of eigenvectors used in the KL expansion. Using less eigenvectors for the updating than for the Monte Carlo realization causes FORM to neglect some of the uncertainty. It is shown how different amounts of eigenvectors m change the quality of the outcome.

The updated mean values are almost not affected by the amount of random variables i.e. eigenvectors used in the KL expansion (Figure 4.41-4.46).

The posterior standard deviations increase with the amount of random variables since more possible solutions to match the observations occur (Figure 4.47-4.52). This result shows that FORM in combination with a KL expansion is able to produce good estimations of $\log K$ and $\log S$ even with a small number of random variables. This reduces the computational effort significantly.

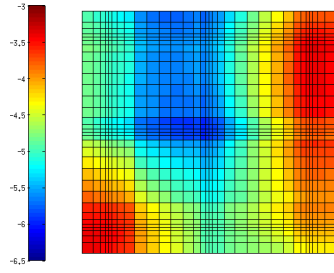


Figure 4.41: FORM solution of the mean of $\log K$ for $m=25$

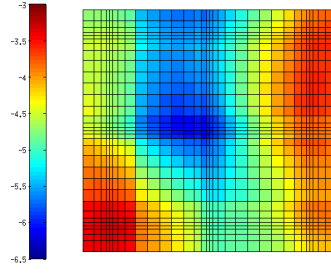


Figure 4.42: FORM solution of the mean of $\log K$ for $m=50$

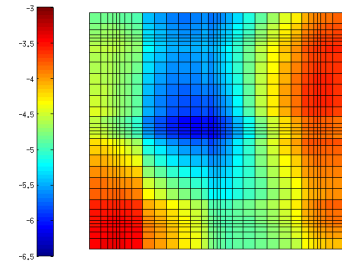


Figure 4.43: FORM solution of the mean of $\log K$ for $m=75$

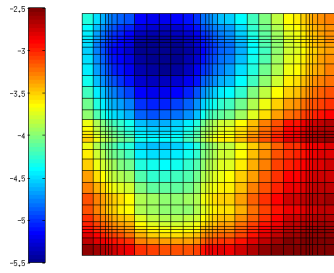


Figure 4.44: FORM solution of the mean of $\log S$ for $m=25$

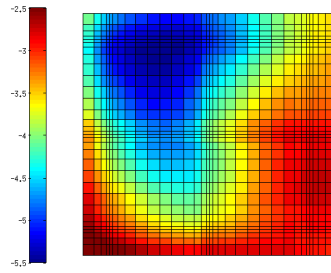


Figure 4.45: FORM solution of the mean of $\log S$ for $m=50$

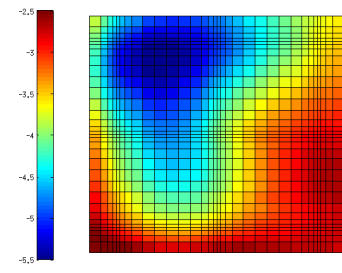


Figure 4.46: FORM solution of the mean of $\log S$ for $m=75$

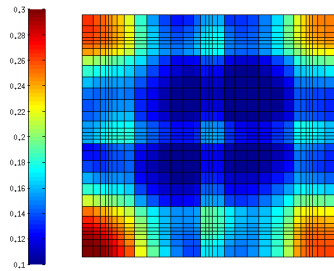


Figure 4.47: FORM solution of the standard deviation of $\log K$ for $m=25$

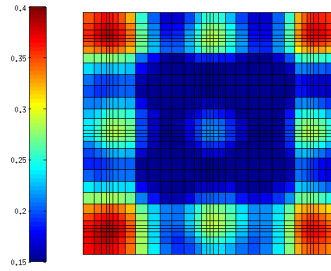


Figure 4.48: FORM solution of the standard deviation of $\log K$ for $m=50$

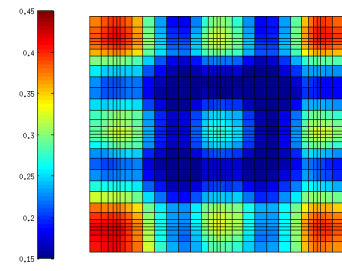


Figure 4.49: FORM solution of the standard deviation of $\log K$ for $m=75$

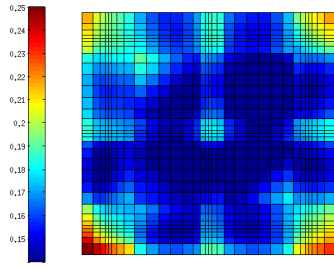


Figure 4.50: FORM solution of the standard deviation of $\log S$ for $m=250$

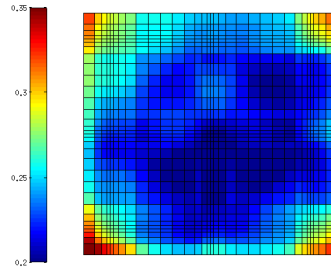


Figure 4.51: FORM solution of the standard deviation of $\log S$ for $m=50$

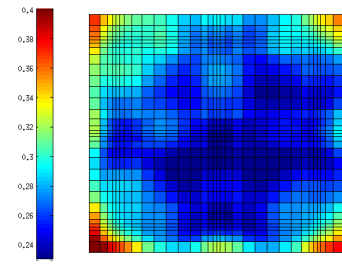


Figure 4.52: FORM solution of the standard deviation of $\log S$ for $m=75$

	Number of Model Runs per Observation	Number of Observations	Computation time in s
Case 1	204	108	11016
Case 2	404	108	21816
Case 3	604	108	32616

Table 4.6: Computational effort for cases with different amounts of random variables

Chapter 5

Summary and Conclusions

The thesis showed, that it is possible to use FORM in order to estimate uncertain ground parameters based on hydraulic tomography. This was done by first creating a numerical model of The North Campus Research Site at The University of Waterloo, in Waterloo, Ontario, Canada. The model uses a finite differences mesh and assumes a linear relationship between hydraulic heads and ground parameters. This is enough to simulate hydraulic tomography by solving a system of linear equations.

The ground parameters throughout the mesh are unknown and were represented by gaussian random variables for each mesh element. Therefore each ground parameter distribution in the mesh is a discretized gaussian random field and fully described by its mean vector and covariance matrix.

Information on the ground parameters can be gathered by observing the drawdown due to a pumping test for different locations and times. The idea of FORM is then to find a design point for each observation. The design point represents a realization of the ground parameters that most likely creates the observed drawdown. This is done by linearizing the system multiple times.

For each linearization sensitivities are obtained by a finite differences approach. The disadvantage of this method is the high computational effort. It was implemented because the software MODFLOW-USG was used for solving model and its source code could not be changed.

In order to reduce the computational effort, KL-expansions of the discretized gaussian random fields to represent ground parameters were used. They allows to describe a gaussian random field through a combination of a small number of eigenvectors, eigenvalues and standard normal random variables. A linearization of the system with finite differences can then be obtained in a reasonable amount of time.

After identifying design points in term of the standard normal random variables for all available observations, the mean values and the covariances of the ground parameters were updated. This was done by solving the inverse problem at the design points.

The results show that it is important to chose observations that do not lead to redundant information since this causes the covariance matrix of the design points to be almost singular. A large number increases the probability of redundant information while a small number might not be enough to get a good estimation. It is also illustrated, that for the analysed situation reasonable results are produced even with a small number of random variables. This reduces the computational effort of linearizing the system therefore the computation time.

For future studies of systems with higher randomness and 3 dimensions it is advised to change the linearization method. E.g. the adjoint sensitivity method [Mok, 1999] would significantly reduce the computation time for problems like this.

Appendix A

The First Appendix

Observation Number	Well Location	Oberservation Location	Drawdown	Time is s
1	1	2	0.0714511946	232.5646209717
2	1	2	0.1492122412	3404.2912597656
3	1	2	0.0133028664	48.110824585
4	1	3	0.0517075919	1906.8594970703
5	1	3	0.0942593813	9208.8095703125
6	1	3	0.0069970624	401.8125305176
7	1	4	0.1778849959	2551.609375
8	1	4	0.1516875029	401.8125305176
9	1	4	0.0901317745	48.110824585
10	1	5	0.0745601058	561.1848754883
11	1	5	0.1346445233	6023.30859375
12	1	5	0.0096891541	29.7878398895
13	1	6	0.053640563	1906.8594970703
14	1	6	0.0976988822	7995.6127929687
15	1	6	0.0089991419	401.8125305176
16	1	7	0.0634858087	401.8125305176
17	1	7	0.1197282746	5225.6118164063
18	1	7	0.0113274595	48.110824585
19	1	8	0.0596694909	1222.1575927735
20	1	8	0.1101742536	7995.6127929687
21	1	8	0.0152381258	121.2813720703
22	1	9	0.0512822755	1906.8594970703
23	1	9	0.0943825096	9208.8095703125
24	1	9	0.0099352431	401.8125305176
25	2	3	0.1964025795	1419.3359375
26	2	3	0.3816715479	7995.6127929687
27	2	3	0.0351780504	281.304107666
28	2	4	0.0987590924	337.3545532227
29	2	4	0.1963707358	6023.30859375
30	2	4	0.0149905542	48.110824585

Observation Number	Well Location	Oberservation Location	Drawdown	Time is s
31	2	5	0.2039298564	771.9548339844
32	2	5	0.4047627747	6940.6591796875
33	2	5	0.0233647935	69.182258606
34	2	6	0.1988374144	1646.0910644531
35	2	6	0.3799045682	7995.6127929687
36	2	6	0.0251412131	337.3545532227
37	2	7	0.0950667113	1050.6979980469
38	2	7	0.1752719134	7995.6127929687
39	2	7	0.0115322173	121.2813720703
40	2	8	0.1813874543	1646.0910644531
41	2	8	0.3262759745	7995.6127929687
42	2	8	0.0237032399	232.5646209717
43	2	9	0.1763631552	1646.0910644531
44	2	9	0.3519567549	9208.8095703125
45	2	9	0.0276763309	401.8125305176
46	3	4	0.0804158002	1906.8594970703
47	3	4	0.1434198916	7995.6127929687
48	3	4	0.0110676698	401.8125305176
49	3	5	0.2919389904	1646.0910644531
50	3	5	0.5449278951	7995.6127929687
51	3	5	0.0389122739	337.3545532227
52	3	6	0.4680910707	1646.0910644531
53	3	6	0.8356221318	7995.6127929687
54	3	6	0.0667017102	281.304107666
55	3	7	0.0986957774	1906.8594970703
56	3	7	0.1947930306	9208.8095703125
57	3	7	0.0141806379	475.9392089844
58	3	8	0.3191044331	1646.0910644531
59	3	8	0.610876739	7995.6127929687
60	3	8	0.0535416529	401.8125305176
61	3	9	0.3867999911	1646.0910644531
62	3	9	0.7333145738	7995.6127929687
63	3	9	0.0655375198	401.8125305176
64	4	5	0.2149297744	659.2174072266
65	4	5	0.3151738048	6023.30859375
66	4	5	0.0957087949	69.182258606
67	4	6	0.1025620922	1419.3359375
68	4	6	0.1941308826	7995.6127929687
69	4	6	0.0217628181	281.304107666
70	4	7	0.6456546187	2948.2055664062
71	4	7	0.5629690289	771.9548339844
72	4	7	0.2252462357	69.182258606
73	4	8	0.3498367369	5225.6118164063
74	4	8	0.2302719355	659.2174072266
75	4	8	0.0822696462	48.110824585
76	4	9	0.1987262368	7995.6127929687
77	4	9	0.1113987267	1419.3359375
78	4	9	0.025864318	232.5646209717
79	5	6	0.5468754768	1222.1575927735
80	5	6	0.0931979865	153.3283843994

Observation Number	Well Location	Observation Location	Drawdown	Time is s
81	5	6	0.9006897807	6023.30859375
82	5	7	0.4439965487	6023.30859375
83	5	7	0.2529279888	561.1848754883
84	5	7	0.0535371639	48.110824585
85	5	8	0.5593643785	659.2174072266
86	5	8	0.9901368022	6023.30859375
87	5	8	0.1875112802	69.182258606
88	5	9	0.5403273106	1050.6979980469
89	5	9	0.9319388866	6940.6591796875
90	5	9	0.0945863649	121.2813720703
91	6	7	0.1853881925	1646.0910644531
92	6	7	0.0368327387	281.304107666
93	6	7	0.3118919134	7995.6127929687
94	6	8	0.6393870115	1050.6979980469
95	6	8	1.1204125881	6940.6591796875
96	6	8	0.1363338232	153.3283843994
97	6	9	0.7347593307	901.6028442383
98	6	9	1.3480013609	6023.30859375
99	6	9	0.0952525362	93.4144058228
100	7	8	0.8364143968	3404.2912597656
101	7	8	0.5054364204	337.3545532227
102	7	8	0.1877137274	48.110824585
103	7	9	0.1856811047	1222.1575927735
104	7	9	0.3373119533	6940.6591796875
105	7	9	0.0404153317	190.1824493408
106	8	9	0.7648185492	659.2174072266
107	8	9	1.3659850359	6023.30859375
108	8	9	0.2038827389	69.182258606

List of Figures

2.1	PDF of normal distribution with $\mu=1$ and $\sigma=2$	8
2.2	PDF of normal distribution with $\mu=2$ and $\sigma=4$	8
2.3	PDF of standard normal distribution	9
2.4	CDF of standard normal distribution	9
2.5	Example for midpoint discretization	12
3.1	Design point location in the standard normal space	20
4.1	Well locations at NCRS	26
4.2	Hydraulic head change due to a pumping test	26
4.3	Example for drawdown curve (7.5m distance from PW)	27
4.4	Example for drawdown curve (~ 21 m distance from PW)	27
4.5	Finite Differences Mesh of NCRS	29
4.6	Finite Differences Mesh of Area of Interest	29
4.7	Realization of (4.3) for $l=5$ (red), $l=10$ (green) and $l=15$ (blue)	31
4.8	Linear Relationship between standard normal random variables and $\log K(\log S)$	32
4.9	Numbering of locations for pumping and observation in the area of interest	33
4.10	Linear Relationship between standard normal random variables and the hydraulic heads	33
4.11	Structure of FORM implementation	34
4.12	Monte Carlo Simulation of $\log K$ inside the area of interest	35
4.13	Monte Carlo Simulation of $\log S$ inside the area of interest	35
4.14	FORM solution of the mean of $\log K$ for $\varepsilon=0$	36
4.15	FORM solution of the mean of $\log K$ for $\varepsilon=0.1$	36

4.16 FORM solution of the mean of logK for $\varepsilon=0.0075$	36
4.17 FORM solution of the mean of logS for $\varepsilon=0$	36
4.18 FORM solution of the mean of logS for $\varepsilon=0.1$	36
4.19 FORM solution of the mean of logS for $\varepsilon=0.0075$	36
4.20 FORM solution of the standard deviation of logK for $\varepsilon=0$	36
4.21 FORM solution of the standard deviation of logK for $\varepsilon=0.1$	36
4.22 FORM solution of the standard deviation of logK for $\varepsilon=0.0075$	36
4.23 FORM solution of the standard deviation of logS for $\varepsilon=0$	36
4.24 FORM solution of the standard deviation of logS for $\varepsilon=0.1$	36
4.25 FORM solution of the standard deviation of logS for $\varepsilon=0.0075$	36
4.26 Drawdowns for initial means (green), simulated radnom variables (red) and FORM estimation(blue)	37
4.27 Drawdowns for initial means (green), simulated radnom variables (red) and FORM estimation(blue)	37
4.28 Drawdowns for initial means (green), simulated radnom variables (red) and FORM estimation(blue)	37
4.29 FORM solution of the mean of logK for $n=10$	38
4.30 FORM solution of the mean of logK for $n=50$	38
4.31 FORM solution of the mean of logK for $n=108$	38
4.32 FORM solution of the mean of logS for $n=10$	38
4.33 FORM solution of the mean of logS for $n=50$	38
4.34 FORM solution of the mean of logS for $n=108$	38
4.35 FORM solution of the standard deviation of logK for $n=10$	38
4.36 FORM solution of the standard deviation of logK for $n=50$	38
4.37 FORM solution of the standard deviation of logK for $n=108$	38
4.38 FORM solution of the standard deviation of logS for $n=10$	39
4.39 FORM solution of the standard deviation of logS for $n=50$	39
4.40 FORM solution of the standard deviation of logS for $n=108$	39
4.41 FORM solution of the mean of logK for $m=25$	40
4.42 FORM solution of the mean of logK for $m=50$	40
4.43 FORM solution of the mean of logK for $m=75$	40
4.44 FORM solution of the mean of logS for $m=25$	40

4.45 FORM solution of the mean of $\log S$ for $m=50$	40
4.46 FORM solution of the mean of $\log S$ for $m=75$	40
4.47 FORM solution of the standard deviation of $\log K$ for $m=25$	40
4.48 FORM solution of the standard deviation of $\log K$ for $m=50$	40
4.49 FORM solution of the standard deviation of $\log K$ for $m=75$	40
4.50 FORM solution of the standard deviation of $\log S$ for $m=250$	40
4.51 FORM solution of the standard deviation of $\log S$ for $m=50$	40
4.52 FORM solution of the standard deviation of $\log S$ for $m=75$	40

List of Tables

4.1	Assumed mean and standard deviation of $\log K$	30
4.2	Assumed mean and standard deviation of $\log S$	30
4.3	Amounts of Observations	37
4.4	Cases with different amounts of observations	37
4.5	Computational effort for cases with different amounts of observations	39
4.6	Computational effort for cases with different amounts of random variables	41

Bibliography

- Nathabandu T. Kottegoda and Renzo Rosso. *Applied Statistics for Civil and Environmental Engineers*. Blackwell Publishing, 2008.
- O. Le Maitre and O.M. Knio *Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics*. Springer, 2010.
- Robert J. Adler *The Geometry of Random Fields*. John Wiley Sons, 1981.
- Papaioannou I. *Non-intrusive Finite Element Reliability Analysis Methods*. PhD thesis, TU München, 2011.
- Matthew Alexander, Steven J. Berg , and Walter A. Illman *Field Study of Hydrogeologic Characterization Methods in a Heterogeneous Aquifer*. National Groundwater Assotiation, 2011.
- Steven J. Berg *Hydraulic Tomography: Field and Laboratory Studies*. PhD thesis, University of Waterloo, 2011.
- Der Kiureghian A., Nikolaidis E., Ghiocel D. M. , and Singhal S., eds., *First- and second-order reliability methods. Engineering Design Reliability Handbook*. CRC Press, 2005.
- Junfeng Zhu Tian-Chyi J.Yeh *Characterization of aquifer heterogeneity using transient hydraulic tomography*. Water Resources Research, 26 July 2005.
- Karrow, P.F. *Quaternary Geology, Stratford-Conestogo Area*. Ontario, Canada: Geological Survey Report, 1997. .
- Karrow, P.F. *Geology of the University of Waterloo Campus*. Department of Earth Sciences, 1979.
- Sebol, L.A. *Determination of groundwater age using CFCs in three shallow aquifers in Southern Ontario*. PhD thesis, University of Waterloo, 2000.
- Hasofer A. M., Lind N. C. *An Exact and Invariant First-order Reliability Format*. Solid Mechanics Division, University of Waterloo, 1973.
- Rackwitz R., Fiessler B. *Non-normal vectors, quadratic limit state criteria, systems and extreme loads in first order reliability theory: Zuverlässigkeitstheorie I. [erster] Ordnung und nicht-normale Vektoren, quadratische Grenzzustandsflächen, Tragsysteme bzw. aussergewöhnliche Einwirkungen*. Sonderforschungsbereich 96, 1978.
- Debra L. Hughson, T.-C. Jim Yeh *A geostatistically based inverse model for three-dimensional variably saturated flow*. Stochastic Hydrology and Hydraulics 12, 1998.

Tian-Chyi J.Yeh, Jinqi Zhang *A geostatistical inverse method for variably saturated flow in the vadose zone*. Water Resources Research, 1996.

Jinqi Zhang, Tian-Chyi J.Yeh *An iterative geostatistical inverse method for steady flow in the vadose zone*. Water Resources Research, 1997.

J. Zaanen *Linear Analysis*. North Holland Publishing Co., 1960.

B. Sudret, A. Der Kiureghian *Stochastic Finite Element Methods and Reliability: A State-of-the-Art Report*. Department of Civil Environmental Engineering University of California, Berkeley, 2000.

C. M. Mok *Conditional Reliability Analysis of Groundwater Flow and Subsurface Contaminant Transport*. PhD thesis, University of California at Berkeley, 1999.