İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

THE USE OF PARTIALLY DOUBLY STOCHASTIC PARAMETER ESTIMATION METHOD FOR UNDERDETERMINED AND OVERDETERMINED PROBLEMS OF RESERVOIR CHARACTERIZATION

M.Sc. Thesis by
Melek DENİZ

Department : **Petroleum And Natural Gas Engineering**

Program : **Petroleum And Natural Gas Engineering**

Thesis Supervisor: Prof. Dr. Mustafa ONUR

JUNE, 2011

İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

THE USE OF PARTIALLY DOUBLY STOCHASTIC PARAMETER ESTIMATION METHOD FOR UNDERDETERMINED AND OVERDETERMINED PROBLEMS OF RESERVOIR CHARACTERIZATION

M.Sc. Thesis by
Melek DENİZ
(0505081509)

Date of submission : **06 May 2011**
Date of defence examination: **10 June 2011**

Supervisor (Chairman) : **Prof. Dr. Mustafa ONUR (ITU)**

Members of the Examining Committee : **Assist. Prof. Ömer İnanç TÜREYEN (ITU)**
Assoc. Prof. Ayşe KAŞLILAR (ITU)

JUNE, 2011

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

KISMİ ÇİFTE OLASILIKLI PARAMETRE TAHMİN YÖNTEMİNİN AZ TANIMLI VE AŞIRI TANIMLI REZERVUAR KARAKTERİZASYONU PROBLEMLERİNDE KULLANIMI

YÜKSEK LİSANS TEZİ

Melek DENİZ
(0505081509)

Tezin Enstitüye Verildiği Tarih : 06 Mayıs 2011
Tezin Savunulduğu Tarih : 10 Haziran 2011

Tez Danışmanı : Prof. Dr. Mustafa ONUR (İTÜ)
Diğer Jüri Üyeleri : Yrd. Doç. Dr. Ömer İnanç TÜREYEN (İTÜ)
Doç. Dr. Ayşe KAŞLILAR (İTÜ)

HAZİRAN, 2011

To my family and my Fiancé,

FOREWORD

In this thesis, the objective is to determine the reservoir variables by using doubly stochastic model and reduce the uncertainty on samples and prior information. Thus, it is shown that by using this stochastic model, the performance prediction in reservoirs can be made with less error. I would like to express my deep appreciation and thanks for my advisor Prof. Dr. Mustafa Onur who suggested me the topic. I sincerely thank him for helping and supporting me in every phase of this work. This thesis cannot be completed without him. Besides, I would like to thank to Assist. Prof. Dr. Ö. İnanç Türeyen for useful discussions to assist me in understanding my problems and for his help to improve my programming abilities in FORTRAN.

I extend my special thanks to the members of the Examining Committee; Assist. Prof. Dr. Ö. İnanç Türeyen and Assoc. Prof. Ayşe KAŞLILAR for their attention.

During this research, I would like to give my endless thanks to my mother, father, and sisters for their moral support. My special thanks are also extended to my future husband Emrah Paker for giving me support at every stage of this work and for his material aid and spiritual support.

May 2011

Melek Deniz
Petroleum and Natural Gas Engineer

TABLE OF CONTENTS

Page

ABBREVIATIONS

Abs	:Absolute Value
BU	: Buildup Periods
C	: Covariance Matrix
$\mathbf{C_w}$	: Wellbore Storage Coefficient, B/psi
$\mathbf{c_t}$	: Total Compressibility, 1/psi
DD	: Drawdown Periods
DS	: Doubly Stochastic
E	: Unit Vector
E	: Identity Matrix
F	: Function of Model
Ft	: Feet
G	: Sensitivity Matrix
G	: Elements of Sensitivity Matrix
GSlib	: Geostatistical Software Library
h	: Thickness, ft
H	: Hessian Matrix
I	: Identity Matrix
k	: Permeability, md
K	: Number of Data Sets
L	: Length, ft
Lnk	: Logarithm of Permeability
L-M	: Levenberg – Marquardt
L(m)	: Likelihood Function
LUBKS	: Back Substitution of LU
LUDCMP	: LU Decomposition
m	: Model Parameter
M	: Dimension of Parameter
MAP	: Maximum a Posteriori Estimate
Md	: MiliDarcy
ML	: Maximum Likelihood
$\mathbf{m_{pr}}$	: Prior Mean
N	: Dimension of Given Parameter
$\mathbf{N_d}$	: Number of Observations
O, Obj	: Objective Function
OK	: Ordinary Kriging
pdf	: Probability Density Function
PDS	: Partially Doubly Stochastic
Psi	: Pounds per Square Inch
$P_{\hat{m}}$	: Joint pdf
P_{Θ}	: Correction Probability Density Function
RMS	: Root-Mean-Square, psi
S	: Skin Factor, dimensionless
T	: Independent Variable, Time
UWLS	: Unweighted Least-Squares

y	**:** Observed Data
WLS	**:** Weighted Least-Squares
vart	**:** Variance of Correction
1D	**:** One Dimensional
∞	**:** Infinity
~	**:** Posterior Estimate

Greek Symbols

Φ	**:** Porosity, dimensionless
$\boldsymbol{\sigma}^2$	**:** Variance of Given Parameter
$\hat{\sigma}^2$	**:** Unbiased Variance
$\tilde{\sigma}^2$	**:** Posterior Variance
θ	**:** Correction Vector
$\boldsymbol{\theta}_0$	**:** Mean of Correction Vector
δ	**:** Search Direction
λ	**:** Constant
∇	**:** Gradient
Π	**:** Posterior Probability Density Function

Subscripts

i, j	**:** Subscript
pr	**:** Prior

Superscripts

-1	**:** Inverse of a Square Matrix
T	**:** Transpose of a Vector/Matrix

LIST OF TABLES

Page

LIST OF FIGURES

Page

THE USE OF PARTIALLY DOUBLY STOCHASTIC PARAMETER ESTIMATION METHOD FOR UNDERDETERMINED AND OVERDETERMINED PROBLEMS OF RESERVOIR CHARACTERIZATION

SUMMARY

To ascertain reservoir characterization and make a production plan, it is a necessity to have a reservoir model representative of the actual reservoir system under consideration. Such a reservoir model is typically constructed from and calibrated by static (such core and log) data and dynamic (such as production and well test) data. Reservoir characterization based on observed static data normally poses a linear estimation problem, whereas reservoir characterization based on observed dynamic data poses a nonlinear estimation problem. Furthermore, the linear or nonlinear estimation problems can be classified as the underdetermined and overdetermined problems. In the case of the underdetermined problems, the number of unknown parameters (of the reservoir model) to be estimated is far more than the number of observed data available, whereas the number of unknown parameters to be estimated is far less than the number of observed data in the case of overdetermined problems.

As the observed (static and/or dynamic) data alone are usually not sufficient to determine a well-defined reservoir model, it is always useful to incorporate a prior model for the parameters to be estimated from observed data by linear or nonlinear estimation methods. The prior model represents one's prior knowledge of the prior mean and uncertainties of the reservoir parameters (such as permeability, porosity, distance to the fault, etc.). However, the use of prior model in estimation biases the estimates of the model parameters. Hence, if the prior means of the reservoir parameters given are incorrect or uncertain, then the estimates of the parameters could be grossly in error.

In this work, we investigate the effect of errors in the means of the prior model parameters on both underdetermined and overdetermined problems of reservoir characterization by the use of partially doubly stochastic estimation methods with the Bayesian framework, which has shown to be effective if the prior means of the parameters are uncertain. For the case of underdetermined linear problems we consider the use of static data with a prior geostatistical model, and for the case of overdetermined nonlinear problems we consider the use of pressure transient data with a given analytical reservoir model in our investigation. The appropriate objective functions for these cases are derived from probability density functions (pdf) for both linear and nonlinear parameter estimation cases. The results obtained from the partially doubly stochastic parameter estimation methods within the theme of this thesis are compared with those from the conventional methods such as the least-squares (LS) and maximum likelihood methods which do not consider uncertainty in prior means of the model parameters. The results show that if prior

means are incorrect, then the doubly stochastic parameter estimation methods provide more accurate reservoir characterization than these conventional methods that fail to account for uncertainty in prior means of the parameters.

KISMİ ÇİFTE OLASILIKLI PARAMETRE TAHMİN YÖNTEMİNİN AZ TANIMLI VE AŞIRI TANIMLI REZERVUAR KARAKTERİZASYONU PROBLEMLERİNDE KULLANIMI

ÖZET

Yeraltı rezerv tespitleri ve üretim planlamaları yapmak için, incelenmekte olan gerçek rezervuar sistemini yansıtan bir rezervuar modeli yaratmak gerekmektedir. Bu rezervuar modeli genellikle statik (log ve karot gibi) ve dinamik (üretim ve kuyu testleri gibi) veriler ile oluşturulup, düzenlenmektedir. Gözlenen statik dataya bağlı rezervuar karakterizasyonu genellikle lineer tahmin problemi iken, gözlenen dinamik dataya bağlı rezervuar karakterizasyonu lineer olmayan tahmin problemidir. Lineer ve lineer olmayan tahmin yöntemleri ileride az tanımlı ve aşırı tanımlı olarak sınıflandırılacaktır. Az tanımlı problemlerde, rezervuardaki tahmin edilen bilinmeyen parametre sayısı gözlenen uygun data sayısından fazladır. Halbuki aşırı tanımlı problemlerde, gözlenen uygun data sayısı tahmin edilen bilinmeyen parametre sayısından fazladır.

Ancak, iyi bir rezervuar modeli oluşturmada ölçülen datalar (dinamik ve/yada statik) tek başına yeterli olmamaktadır. Bu nedenle parametrelerin tahmin edilmesinde, bu datalardan yararlanarak lineer ve lineer olmayan metodlara önsel model eklemek yararlı olmaktadır. Önsel model, kişinin önsel ortalama ve rezervuar parametrelerinin (geçirgenlik, gözeneklilik, faya olan uzaklık gibi) üzerindeki belirsizlik hakkındaki bilgisini göstermektedir. Ama tahminlerde önsel model kullanmak model parametrelerinin yanlı bulunmasına sebebiyet vermektedir. Eğer verilen önsel ortalama yanlış ya da belirsiz ise parametrelerin değerleri de oldukça yanlış tahmin edilecektir.

Bu çalışmada, Bayes' teoremi kapsamında parçalı çifte olasılık parametre tahmin yöntemi ile önsel model parametrelerinin üzerindeki hatanın az tanımlı ve aşırı tanımlı rezervuar karakterizasyonu problemlerindeki etkisi araştırılmıştır. Bu yöntemin, parametrelerin önsel ortalamalarının belirsiz olması durumunda etkin olduğu görülmüştür. Araştırmamızda, az tanımlı linear problemler için statik datalardan yararlanarak önsel bir jeoistatistik model, aşırı tanımlı linear olmayan problemler için ise kararsız basınç testi verilerinden yararlanarak varolan analitik bir rezervuar modeli göz önünde bulundurulmuştur. Lineer ve lineer olmayan tahmin problemleri için olasılıklı yoğunluk fonksiyonlarından (oyf) kullanılan uygun hedef fonksiyonları türetilmiştir. Bu tez kapsamında kullanılan parçalı çifte olasılık parametre tahmin yönteminden elde edilen sonuçlar, yaygın olarak bilinen en küçük kareler ve maksimum olasılık yöntemleri ile karşılaştırılmıştır. Sonuçlar önsel bilginin yanlış olması durumunda, parametrelerin önsel ortalamaları üzerindeki hatanın ne kadar olduğunu gösteremeyen yaygın metodlara nazaran, parçalı çifte olasılık parametre tahmin yönteminin daha doğru bir rezervuar karakterizasyonu sağladığı görülmüştür.

1. INTRODUCTION

1.1 Overview of Reservoir Characterization Problem

Most of the reservoir engineers and scientists in worldwide oil and gas industry deal with solving inverse problems. The most common problem that they dealt with is to figure out the reservoir characteristics from indirect measurements of reservoir geometry and property. In the inverse problem, one attempts to define properties of the system from available measurements/observations. Geoscientists mainly focus on the reservoir shape, structure, porosity etc. Reservoir engineers are interested in reservoir type and shape, well condition, production data, and both rock and fluid properties. Both geoscientists and engineers have the same overall goal, which is to reduce the uncertainty in reservoir parameters and achieve a correct, representative reservoir model of the unknown actual reservoir system. Depending on this model, they try to make reservoir performance predictions and most importantly assess the uncertainty in future performance predictions. However, to construct such a representative model all available data need to be integrated, which may consist of geological, petro physical, geophysical and production data. Unfortunately, integration of all these data is required in a multidisciplinary background and it is not an easy task for an engineer or a scientist. The two main objectives of this study are to generate realizations while minimizing the uncertainty that represent a correct sampling of the a posteriori probability density function of reservoir descriptions (rock property fields) and by using static permeability or pressure data, determining the unknown reservoir model parameters. To generate realizations correctly, it is essential to use all convenient data and information for generating a posteriori probability density function. The approach for doing this is by the estimation of a most probable mode. Once having formulated that, the realizations are appropriate from the maximum a posteriori estimate by using information (He, 1997). To estimate the properties of reservoir, pressure data are needful because it reflects in situ dynamic properties of reservoir which are measured by gauges in active or/and observation well during well testing process. Although the gauge technology is improved, yet the measurement errors, particularly in rate data still exist.

1.2 Forward and Inverse Problems

Two main problems that engineers are faced to solve are forward and inverse problems. Generally, making or calculating the model response from given input values of all model parameters in a known mathematical model is called a forward (or direct) problem. The solution of this kind of problem is unique. In an inverse problem, the model and the model parameters are inferred from observed response data alone. On the contrary, to the forward problem, the solution of the inverse problem is normally not unique (Tarantola, 2005).

One of the examples for inverse problem is well test interpretation. If we use an analytical model that can be defined by a few model parameters for the interpretation and in addition, we have a number of observed pressure data larger than the model parameters, then this type of well test interpretation is an example of an over-determined inverse problem. Consequently, in this problem, the number of observed data is more than number of unknown model parameters. As mentioned before, the non-unique results are inherent in inverse problems, and there are several reasons for non-uniqueness, such as noise in measurements, uncertainty in real system, nonlinear relation between measured data and model parameters and no considerable effect of some parameters on observed during the time ranges (Onur, 2010). To solve this kind of an over-determined nonlinear problem, the well-known Least-Squares (LS) method is used.

In many circumstances, unfortunately, the number of the unknown model parameters is greater than observed data. This kind of inverse problem is called “ill–posed” or generally underdetermined inverse problem. In which case, there is no unique solution, but there exists multiple solutions. To solve such an inverse problem, we typically formulate an objective function containing a regularization term that incorporates our prior information on the model and the model parameters. The prior information is usually obtained from our prior knowledge of the system from geosciences data available, e.g., geology, geophysics, and geostatistics. This is achieved within the framework of Bayesian estimation (Tarantola, 2005; He, 1997).

1.3 Reservoir Characterization

Reservoir characterization is an important technique to gain the knowledge of the reservoir characteristics by use of all available data. The purpose is to estimate true production features that influence the amount, position, and accessibility of reservoir flowing fluids with minimum uncertainty. Reservoir characteristics include all information about reservoir as porosity, permeability, the structure of reservoir, etc. The available information generally comes from geological, geophysical and petro physical knowledge (core and log) and specific observation of reservoir (well test, production and tracer data) (Hegstad et al., 1998; Kelkar et al., 2002; Damsleth, E., 1994). However, the basic problem is how to integrate effectively all available data obtained from different sources and especially to quantify the existing major uncertainty (Holden et al., 1992).

Recently, improvements of reservoir characterization methods, especially geostatistical methods provide a realistic reservoir description. Geostatistical methods are a decisive fact to estimate the distribution of reservoir parameters in the reservoir. Although the relation between parameters is random, somehow, they are appropriately related through a spatial correlation with each other (such as permeability and porosity). Therefore, geostatistical methods help reservoir engineers for the computation of hydrocarbon reserves, properly selecting production or injection wells locations and for more accurate performance estimation (Ceyhan, 1997).

To quantify spatial relationships of reservoir characteristics miscellaneous modeling techniques are used to estimate parameters' values at unobserved locations. Kriging is one of the conventional geostatistical technique to interpolate the value of a random field and is more common to utilize (Journel and Huijbregts, 1978; Kelkar et al., 2002; Url-1). It is based on the linear model-data theory and the goal in the linear estimation procedure is to calculate a set of weights minimize the estimation variances in individual neighboring points according to the geometry of the field. These weights depend on the spatial relationship between the unsampled location and the neighboring values, nearer samples are assigned higher weights than distant samples (Brummbert et al., 1991). Kriging provides unbiased estimates with minimum variance. It makes use of semivariogram models. Semivariogram is a measure of dissimilarity for features that alter in space. Analysis of semivariogram is useful in comparing such features and in designing their sufficient sampling (Olea, 1994).

2. PARTIALLY DOUBLY STOCHASTIC PARAMETER ESTIMATION MODEL

In both over and under-determined problems, the prior information may be used to reduce uncertainty on reservoir parameters. As mentioned before, the prior information is given by engineers/scientists or obtained from log data and the core data that may contain errors. The errors could be related to the mean of the reservoir parameters that will be used as the prior mean of the reservoir parameters. To account for errors in the prior mean of the reservoir parameters, one may use a doubly stochastic estimation method within the framework of Bayesian statistics (Tjelmeland et al., 1994). Tjelmeland and his team made a model, which both mean and variance are allowed to be unknowns to account for uncertainty in the prior mean and prior covariance (or variance) to be used in parameter estimation (Oliver et al., 2008). In this work, it is assumed that only the prior mean can be in error, but the covariance of prior model is known. This method is referred to as the partially doubly stochastic model (Li et al., 2009; Reynolds et al., 1999; He et al., 2000; Oliver et al., 2008). In this thesis, considering both under and over-determined parameter estimation problems of reservoir characterization with prior information and prior model, the use of partially doubly stochastic model is applied to both least-squares and maximum likelihood estimation methods. As the mathematical expression of fluid flow or future production prediction of a petroleum reservoir is a nonlinear problem, it is usually unrealizable to calculate straightforwardly the probability distribution (Oliver et al., 1997). Even if the problem is solved straightforward, the results are not likely the original ones quite a bit. The main problem is related with the numbers of parameters and the number of the observations and generally, even so the number of data is higher than the unknowns the solution can be non-unique (Gavalas et al. 1976; Shah et al.1978). Although the measurements are not sensitive and generally the number of them is not enough, a variant interpretation method which uses a prior probability density function (pdf) in combination of all prior information or priori knowledge is developed by a 18th century British mathematician and statistician Thomas Bayes and improved by French mathematician Pierre-Simon Laplace (Url-1, Shah et al., 1978). The prior distribution may be characterized by means of the prior density function P_o (m). The new composed function is called a posteriori probability density function (posteriori pdf). Though the number of unknown variables or the number of measurements can not be changed, the Bayesian estimation theory can reduce the roughness on data by utilization of prior information and the problem transforms into a better – determined problem (Gavalas et al., 1976). Besides, the theorem gives a chance for updating the posteriori pdf according to new available information or data (Zhang et al., 2005).

Nowadays, it is conventional to make reservoir models by using geological and geophysical data interpretation or statistical model for the outcomes and with reference to these models, long term production performance predictions are performed (Li et al., 2009). Due to this critical decision, the unknown reservoir parameters must be estimated with minimum error. However, in many samples, the values of all parameters of reservoir include errors on measurements such as log data, cores, seismic, etc. Prior information is assisted to diminish the haziness of the observation data and besides, if there is any faulty data comes from correlations, the priori knowledge provides to reject these implausible values (Kuchuk et al., 2010). This knowledge is represented by prior means in the sampling posteriori pdf and even though the prior means reduce the roughness on unknown variables, yet, the method does not account for the uncertainty on these prior means.

A new approach has been developed by Norwegian mathematicians Håkon Tjelmeland and his advisor, Henning Omre. According to them, the uncertainty of data is reducible by posteriori pdf and the uncertainty of prior means are quantifiable by this new approached is called doubly stochastic model. In their model, both the mean and the variance of the prior is allowed to be unknowns (Tjelmeland et al., 1994; He, N., 1997; Reynolds et al., 1999; He et al., 2000; Oliver et al., 2008; Li et al., 2009). Nevertheless, within this thesis framework, it is considered that only the prior means is unknown and the uncertainty of the prior is allowed to be known. Because it is simpler case of doubly stochastic model (DS), the name of the model is called 'partially doubly stochastic' (PDS) model. The model is based on an additional new term called correction vector, $\boldsymbol{\theta}$, of Bayesian framework and on the authority of Oliver et al. (2008), the correction vector is used to adjust values of the real model parameters up or down whereas it controls the prior means. The correction vector have own prior mean, $\boldsymbol{\theta}_0$, and covariance matrix, $\mathbf{C}_\theta$. This is a new approach whereby wrong prior means cannot dominate the minimization procedure of objective function anymore and so the determined model parameters are more reliable and more accurate. The point to take into consideration is the dimension of correction vector $\boldsymbol{\theta}$. As $\boldsymbol{\theta}$ is added in posterior pdf, it adjusts the model parameters by roughing down the prior means; the dimension of it must be equal to unknown model parameters and prior means.

2.1 Derivation of Partially Doubly Stochastic Parameter Estimation Method

This approach is used in generally accepted parameters estimation methods which are least-squares (LS) and maximum likelihood (ML) estimation based on Bayes' theory for the evaluation of reservoir model variables. Because LS estimation can be directly applied in ad hoc manner to the deterministic model within any importance of probability distribution of the observations, the usage of that is most widely for curve fitting procedure without constructing confidence intervals. Contrary to this, ML estimation is more appropriate to the matching statistically as it analyses the observations such random variables with certain probability distributions and it can be said that the LSE is a special implementation of ML estimation (Kuchuk et al., 2010). The objective functions minimized of both linear and nonlinear problems are derived for model parameters vector $\mathbf{m}$ and $\boldsymbol{\theta}$ from posterior probability density functions. These estimation methods are applied to find the true values of reservoir parameters. Here and throughout, the small letter bold faces denote vectors, whereas the capital letter bold faces denote matrices.

From a statistical point of view, the unknown model parameter $\mathbf{m}$ is M – dimensional vector to be estimated, $\mathbf{m} = \left[m_1, m_2, ..., m_M\right]^{\mathrm{T}}$ and the superscript T denotes transpose of a vector or a matrix and lower case letters in bold type refer to vectors, generally column vectors, while capital case letters in bold type refer to matrices. The model is given as,

$$\mathbf{y} = \mathbf{f}(\mathbf{m}, \mathbf{t}) + \mathbf{e}\,. \qquad \textbf{(2.1)}$$

In Eq. 2.1, $\mathbf{y}$ represents N_d-dimensional observed-data vector with each entry representing observation of the dependent variable as shown below,

$$\mathbf{y} = \left[y_1, y_2, ..., y_{N_d} \right]^{\mathrm{T}} \qquad \textbf{(2.2)}$$

It is considered that the population is unknown and these observations are random sample of this distribution. $\mathbf{f(m,t)}$ is the function that represents the relationship between model parameter $\mathbf{m}$ and independent variable N_d-dimensional vector $\mathbf{t} = \left[t_1, t_2, ..., t_{N_d} \right]^{\mathrm{T}}$. The relationship between these vectors can be linear or nonlinear according to the problem and due to this, $\mathbf{f}(\mathbf{m},\mathbf{t})$ is called deterministic part of the equation. $\mathbf{e}$, $\mathbf{e} = \left(e_1, e_2, ..., e_{N_d} \right)^{\mathrm{T}}$, is the stochastic part of this equation which is again N_d-dimensional vector and it is also called error vector of unknown measurement. It is assumed that the error vector is obtained from normal distribution with zero mean and $N_d \times N_d$ covariance matrix (variogram) $\mathbf{C}_D$. The term provides to quantify the noise in observed data, which includes both measurements and stochastic errors.

The objective of data analysis is to generate the most likely distribution from the observation **y** and all distributions have a unique model parameter **m**, which indicates probability of measuring data y. Because of this, the probability density function of observation is defined as $P(\mathbf{y}|\mathbf{m})$. The most common distribution function is Gaussian (Kuchuk et al., 2010). According to Bard (1974), there are many reasons to use this distribution and the reasons are, respectively,

- The behavior of this distribution is seen often in the environment and it can be shown in analytically.
- It is needed least information in order to form the distribution function.
- Besides if the number of sample is increased through infinity in any distribution, it approaches to the Gaussian distribution. It is called central limit theorem (Feller, 1966).

If we assume that $\mathbf{y}$ can be described by a Gaussian pdf with mean $\mathbf{f}$ (vector) and covariance matrix $\mathbf{C}_D$ than the conditional pdf is given by,

$$P_{\mathbf{m}}(\boldsymbol{y}|\boldsymbol{m}, \boldsymbol{C}_D) = \frac{1}{(2\pi)^{N_d/2} \left(\det \boldsymbol{C}_D \right)^{1/2}} \exp\left[-\frac{1}{2} \left[\mathbf{y} - \mathbf{f(m,t)} \right]^T \boldsymbol{C}_D^{-1} \left[\mathbf{y} - \mathbf{f(m,t)} \right] \right] \qquad \textbf{(2.3)}$$

In Eq. 2.3, $\mathbf{C}_D$ is an $N_d \times N_d$ positive definite and symmetric error covariance (nonsingular) matrix that represents the correlation between errors in observed data and det($\mathbf{C}_D$) represents the determinant of $\mathbf{C}_D$. The superscript "-1" shows the inverse of a square matrix. The diagonal elements of $\mathbf{C}_D$ are just the variances distributed independently, σ_j^2, j=1,2,…,N_d, that is described the variance of error at each observed y_j. In this thesis, it is assumed that all errors are distributed identically so that $\sigma_j^2 = \sigma^2$ for all j, then $\mathbf{C}_D = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is N_d x N_d identity matrix.

Eq. 2.3 identifies the distribution function of observations and similarly, the distribution of prior model is assumed to have a multivariate Gaussian probability density function with mean, $\mathbf{m}_{pr}$, and $M \times M$ dimension prior covariance matrix, $\mathbf{C}_M$. Both these vectors are known as a priori for a Gaussian random field (Oliver et al., 1997). The pdf is given by

$$P_0(\mathbf{m}) = \frac{1}{(2\pi)^{M/2} \left(\det \mathbf{C}_M\right)^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{m} - \mathbf{m}_{pr}\right)^T \mathbf{C}_M^{-1} \left(\mathbf{m} - \mathbf{m}_{pr}\right)\right] \tag{2.4}$$

In Eq. 2.4, the dimension of model parameter and its dimension is M. If $\mathbf{m}$ is modeled as stationary random functions, then $\mathbf{m}_{pr}$ is treated as a constant vector (Reynolds et al., 1999). However, in doubly stochastic model, $\mathbf{m}_{pr}$ is corrected and so, a new conditional pdf must be derived as:

$$P_0(\mathbf{m} \mid \boldsymbol{\theta}) = \frac{1}{(2\pi)^{M/2} \left(\det \mathbf{C}_M\right)^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1} \left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)\right] \tag{2.5}$$

The random vector $\boldsymbol{\theta}$ represents the correction to $\mathbf{m}_{pr}$ with $\boldsymbol{\theta}$ indicating specific realizations of $\boldsymbol{\theta}$. The vectors $\mathbf{m}_{pr}$ and $\boldsymbol{\theta}$ have M-dimension just like model parameter $\mathbf{m}$. The new pdf now, includes the uncertainty in the prior means vector $\mathbf{m}_{pr}$, contrary Eq. 2.4. The conditional pdfs of data and prior means are defined, respectively, Eq. 2.3 and Eq. 2.5 and if it is assumed that the correction vector must be sampling, the correction probability density function, Θ is given by

$$P_\Theta(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{M/2} \left(\det \mathbf{C}_\theta\right)^{1/2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)\right] \tag{2.6}$$

In Eq. 2.6, the pdf for Θ is also assumed Gaussian with mean $\boldsymbol{\theta}_0$ and covariance matrix $\mathbf{C}_\theta$. $\boldsymbol{\theta}_0$ is the mean of the random vector $\boldsymbol{\theta}$ and $\mathbf{C}_\theta$ is symmetric and positive definite matrix. It is assumed that $\mathbf{C}_\theta$ is a diagonal matrix, because of independent error in the prior means (Reynolds, 1999). So, the joint pdf for $\mathbf{m}$ and $\boldsymbol{\theta}$ is shown below,

$$P_{\hat{m}}(\mathbf{m},\boldsymbol{\theta}) = P_0(\mathbf{m}\,|\,\boldsymbol{\theta})P_\Theta(\boldsymbol{\theta}) = \mathrm{a}\exp\begin{bmatrix} -\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) \\ -\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \end{bmatrix} \tag{2.7}$$

where $\hat{\mathbf{m}} = \left[\mathbf{m}\ \boldsymbol{\theta}\right]^{\mathbf{T}}$ and is a constant that normalize the pdf. The probability density function is also called the likelihood function. The expectation of this equation is not equal just prior mean $\mathbf{m}_{pr}$, it is equal the sum of $\mathbf{m}_{pr}$ and $\boldsymbol{\theta}_0$. As it is known by standard applications of Bayes theorem, the combination of priori probability density function and likelihood function gives a posteriori pdf (Url-1) and the posteriori pdf of this model is defined as

$$\Pi(\mathbf{m},\boldsymbol{\theta}) = P_m(\mathbf{y}\,|\,\mathbf{m},\mathbf{C}_D)P_{\hat{m}}(\mathbf{m},\boldsymbol{\theta}) = b\exp\begin{bmatrix} -\frac{1}{2}\left[\mathbf{y}-\mathbf{f}(\mathbf{m},\mathbf{t})\right]^T \mathbf{C}_D^{-1}\left[\mathbf{y}-\mathbf{f}(\mathbf{m},\mathbf{t})\right] \\ -\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) \\ -\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \end{bmatrix} \tag{2.8}$$

where b is a constant.

The extension of the likelihood function given by Eq. 2.8 for a K set of observed data each having a N_d data and data error covariance matrix $\mathbf{C}_{Dj}$, for $j = 1,2,\ldots,K$, can be given as (Kuchuk et al., 2010):

$$\Pi\left(\mathbf{m},\boldsymbol{\theta},\mathbf{C}_{\mathrm{Dj}}\right) = \hat{b}\exp\begin{bmatrix} -\frac{1}{2}\sum_{j=1}^{K}\left(\mathbf{y}_j-\mathbf{f}_j\left(\mathbf{m},\mathbf{t}\right)\right)^T \mathbf{C}_{Dj}^{-1}\left(\mathbf{y}-\mathbf{f}\left(\mathbf{m},\mathbf{t}\right)\right) \\ -\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) \\ -\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \end{bmatrix} \tag{2.9}$$

where $\hat{b}$ is a constant.

It should be noted that depending on the treatment of the data error covariance matrix $\mathbf{C}_D$ (or $\mathbf{C}_{Dj}$) in Eq. 2.8 (in Eq. 2.9), we have two different parameter estimation problems. If $\mathbf{C}_D$ (or $\mathbf{C}_{Dj}$) is treated as unknown together with $\mathbf{m}$ and $\boldsymbol{\theta}$, then this is called maximum likelihood estimation. On the other hand, if we assume that the data error covariance matrix $\mathbf{C}_D$ (or $\mathbf{C}_{Dj}$) is known, i.e., the weights for observed data are known, then to generate the most probable model, and then this is called least-squares estimation. In the following sections, we treat each problem separately by assuming that $\mathbf{C}_D$ (or each $\mathbf{C}_{Dj}$ in Eq. 2.9) is a diagonal matrix, with the diagonal elements representing the variance of error in observed data.

2.2 Likelihood Function and Maximum Likelihood Estimation

A likelihood function is a function of the parameters of a statistical model, defined as follows: the *likelihood* of a set of parameter values given some observed data is equal to the *probability* of those observed data given those parameter values (Url-1). Since the probability density function depends on a model parameter and the distribution parameter in given observed data, if the value of model parameter is changed, the observed data of the distribution are no longer the same outcomes. In general, the likelihood function of parameter indicates how likely a value of the parameter is, in given observed data.

If we treat the error covariance matrix $\mathbf{C}_{Dj}$ as unknown in Eq. 2.9, then the posterior distribution function, also called the likelihood function, can be written as:

$$\Pi\left(\mathbf{m},\boldsymbol{\theta},\mathbf{C}_{Dj}\right)=\hat{b}\exp\begin{bmatrix}-\frac{1}{2}\sum_{j=1}^{K}\left(\mathbf{y}_j-\mathbf{f}_j\left(\mathbf{m},\mathbf{t}\right)\right)^T\mathbf{C}_D^{-1}\left(\mathbf{y}-\mathbf{f}\left(\mathbf{m},\mathbf{t}\right)\right)\\-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T\mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)\\-\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T\mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)\end{bmatrix} \quad \textbf{(2.10)}$$

or can be written as:

$$\Pi\left(\mathbf{m},\boldsymbol{\theta},\mathbf{C}_{Dj}\right)=\hat{b}L\left(\mathbf{m},\mathbf{C}_{Dj}\right)P_0(\mathbf{m})P_\theta(\boldsymbol{\theta}) \quad \textbf{(2.11)}$$

Eq. 2.11 is called the posteriori pdf of maximum likelihood method for the partially doubly stochastic model. In maximum likelihood function, we work with the natural logarithm of the posterior pdf given by Eq. 2.11:

$$\ln\left[\Pi\left(\mathbf{m},\boldsymbol{\theta},\mathbf{C}_{Dj}\right)\right]=\ln\left(\hat{b}\right)+\ln\left[L\left(\mathbf{m},\mathbf{C}_{Dj}\right)\right]+\ln\left[P_0(\mathbf{m})\right]+\ln\left[P_\theta(\boldsymbol{\theta})\right] \quad \textbf{(2.12)}$$

Eq. 2.12 can be written more explicitly as:

$$\begin{aligned}\ln\left[\Pi\left(\mathbf{m},\boldsymbol{\theta},\mathbf{C}_{Dj}\right)\right]=&-\frac{1}{2}\left(\sum_{j=1}^{K}N_{dj}\right)\ln(2\pi)-\frac{1}{2}\sum_{j=1}^{K}\ln\left(\det\mathbf{C}_{Dj}\right)\\&-\frac{1}{2}\sum_{j=1}^{K}\left(\mathbf{y}_j-\mathbf{f}_j(\mathbf{m},\mathbf{t})\right)^T\mathbf{C}_{Dj}^{-1}\left(\mathbf{y}-\mathbf{f}(\mathbf{m},\mathbf{t})\right)\\&-\frac{M}{2}\ln(2\pi)-\frac{1}{2}\ln\left(\det\mathbf{C}_M\right)-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T\mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)\\&-\frac{M}{2}\ln(2\pi)-\frac{1}{2}\ln\left(\det\mathbf{C}_\theta\right)-\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T\mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)\end{aligned}$$

(2.13)

For simplicity, we assune that each data error covariance matrix $\mathbf{C}_D$ is diagonal and all diagonal entries for the same data set is identical (but could be different for the other data sets) and in addition, $\mathbf{C}_M$ and $\mathbf{C}_\theta$ are also diagonal and their variances. If the matrices $\mathbf{C}_M$ and $\mathbf{C}_\theta$ are diagonal, then it is easy to show that their inverses are also diagonal; i.e.,

$$\mathbf{C}_M^{-1}=\begin{bmatrix}\frac{1}{\sigma_{m_1}^2} & 0 & \cdots & 0\\ 0 & \frac{1}{\sigma_{m_2}^2} & \cdots & 0\\ \vdots & 0 & \ddots & 0\\ 0 & 0 & \cdots & \frac{1}{\sigma_{m_M}^2}\end{bmatrix} \quad \textbf{(2.14)}$$

and

$$\mathbf{C}_\theta^{-1}=\begin{bmatrix}\frac{1}{\sigma_{\theta_1}^2} & 0 & \cdots & 0\\ 0 & \frac{1}{\sigma_{\theta_2}^2} & \cdots & 0\\ \vdots & 0 & \ddots & 0\\ 0 & 0 & \cdots & \frac{1}{\sigma_{\theta_M}^2}\end{bmatrix} \quad \textbf{(2.15)}$$

Then, for this simplistic case, we can show that Eq. 2.13 can be written as:

$$\ln(\Pi) = -\frac{1}{2}\left(\sum_{j=1}^{K} N_{dj}\right)\ln(2\pi) - \frac{1}{2}\sum_{j=1}^{K} N_{dj}\ln\left(\sigma_j^2\right)$$
$$-\frac{1}{2}\sum_{j=1}^{K}\frac{1}{\sigma_j^2}\sum_{i=1}^{N_{dj}}\left[\mathbf{y}_{i,j} - \mathbf{f}_{i,j}(\mathbf{m})\right]^2$$
$$-\frac{M}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{M}\ln\left(\sigma_{m,i}^2\right) - \frac{1}{2}\sum_{i=1}^{M}\left(\frac{\mathbf{m}_i - \mathbf{m}_{pr,i} - \boldsymbol{\theta}_i}{\sigma_{m,i}}\right)^2 \tag{2.16}$$
$$-\frac{M}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{M}\ln\left(\sigma_{\theta,i}^2\right) - \frac{1}{2}\sum_{i=1}^{M}\left(\frac{\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0,i}}{\sigma_{\theta,i}}\right)^2$$

To maximize Eq. 2.16, we can proceed by the *stage-wise maximization* method (Bard, 1974) and this method includes uncovering the values of σ_j^2 that maximize Eq. 2.16 for any value of model parameter, $\mathbf{m}$. The values of σ_j^2 will be some function of $\mathbf{m}$ like $\tilde{\sigma}_j^2(m)$. Substitution of $\tilde{\sigma}_j^2(m)$ for σ_j^2 in Eq. 2.16 decreases O to a function $\tilde{O}$ of $\tilde{\mathbf{m}}$ alone (Kuchuk et al., 2010). By this way, it is searched for $\tilde{\mathbf{m}}$ which maximizes Eq. 2.16.

Procedure of stage-wise maximization method :

i. Substitution of $\tilde{\sigma}_j^2(m)$ for σ_j^2 in Eq. 2.16 and differentiation respect to σ_j^2 gives following equations

$$\frac{\partial O}{\partial \sigma_j^2} = -\frac{1}{2}\sum_{j=1}^{K}\frac{N_{dj}}{\left(\tilde{\sigma}_j^2\right)} + \frac{1}{2}\sum_{j=1}^{K}\frac{1}{\left(\tilde{\sigma}_j^2\right)^2}\sum_{i=1}^{N_{dj}}\left[\mathbf{y}_{i,j} - \mathbf{f}_{i,j}(\mathbf{m})\right]^2 = 0 \tag{2.17}$$

$$\sum_{j=1}^{K}\frac{1}{\tilde{\sigma}_j^2}\left(-N_{dj} + \frac{1}{\tilde{\sigma}_j^2}\sum_{i=1}^{N_{dj}}\left[\mathbf{y}_{i,j} - \mathbf{f}_{i,j}(\mathbf{m})\right]^2\right) = 0$$

(2.18)

Rearranging Eq. 2.18 in following equation

$$\tilde{\sigma}_j^2 = \frac{1}{N_d}\sum_{i=1}^{N_{dj}}\left[\mathbf{y}_{i,j} - \mathbf{f}_{i,j}(\mathbf{m})\right]^2 \tag{2.19}$$

The new variance obtained by using this method is biased, but also proper to use.

ii. Re-substitution of variances defined in Eq. 2.19 in Eq. 2.16, a new equation is obtained which is called *concentrated likelihood* (Kuchuk et al., 2010).

$$\tilde{L}(\mathbf{m},\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=1}^{K} N_{dj}\left[\ln\left(\frac{N_{dj}}{2\pi}\right)-1\right] - \frac{1}{2}\sum_{j=1}^{K} N_{dj}\ln\left[\sum_{i=1}^{N_{dj}}\left[\mathbf{y}_{i,j}-\mathbf{f}_{i,j}(\mathbf{m})\right]^2\right]$$
$$+M\ln(2\pi) + \frac{1}{2}\sum_{i=1}^{M}\frac{(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta})_i^2}{\sigma_{m,i}^2} + \frac{1}{2}\sum_{i=1}^{M}\frac{(\boldsymbol{\theta}-\boldsymbol{\theta}_o)_i^2}{\sigma_{\theta,i}^2} \quad \textbf{(2.20)}$$

iii. Maximizing Eq. 2.20 is equivalent to minimizing the following objective function:

$$\tilde{O} = \frac{1}{2}\sum_{j=1}^{K} N_{dj}\ln\left[\sum_{i=1}^{N_{dj}}\left[\mathbf{y}_{i,j}-\mathbf{f}_{i,j}(\mathbf{m})\right]^2\right] + \frac{1}{2}\sum_{i=1}^{M}\frac{(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta})_i^2}{\sigma_{m,i}^2} + \frac{1}{2}\sum_{i=1}^{M}\frac{(\boldsymbol{\theta}-\boldsymbol{\theta}_o)_i^2}{\sigma_{\theta,i}^2} \quad \textbf{(2.21)}$$

In Eq. 2.21 can be generalized for the more general case of nondiagonal $\mathbf{C}_M$ and $\mathbf{C}_\theta$ as:

$$O(\mathbf{m},\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=1}^{K} N_{dj}\ln\left\{\left[\mathbf{y}_i-\mathbf{f}_i(\mathbf{m})\right]^T\left[\mathbf{y}_i-\mathbf{f}_i(\mathbf{m})\right]\right\}$$
$$+\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T\mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)+\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T\mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \quad \textbf{(2.22)}$$

It is important to note that we have considered a general formulation that each observed data vector, $\mathbf{y}_i$ may contain a different total number of observed data, N_{dj}, $j = 1,2,\ldots,K$. It is throughout that the total number of unknown model parameter is M, and hence, $\mathbf{m}$ is an M-dimensional vector, and $\mathbf{C}_M^{-1}$ is an $M\times M$ diagonal matrix.

Furthermore, we have considered that the total number of model parameters with uncertain means is M and hence, $\boldsymbol{\theta}$ is an M-dimensional vector, and $\mathbf{C}_\theta^{-1}$ is an $M\times M$ diagonal matrix. $\mathbf{m}_{pr}$ denotes the M-dimensional prior vector with elements equal to the prior means ($\mathbf{m}_{pr,i}$, $i= 1,2,\ldots,M$) of the model parameters $\mathbf{m}_i$, $i=1,2,\ldots,M$. $\boldsymbol{\theta}_0$ denotes the mean correction vector with elements equal to the means ($\boldsymbol{\theta}_{0,i}$, i = 1,2,…,$M$) of the unknown correction parameters $\boldsymbol{\theta}_i$, i=1,2,…,M.

The objective function Eq. 2.22 can be minimized by using the Levenberg-Marquardt method. This method requires computing the gradient of the objective function and the approximate Hessian matrix.

To obtained the gradient of the objective function, it is convenient to partition the gradient as

$$\nabla O(\hat{\mathbf{m}}) = \left[\nabla_{\mathbf{m}} O(\hat{\mathbf{m}}) \quad \nabla_{\boldsymbol{\theta}} O(\hat{\mathbf{m}})\right]^T \tag{2.23}$$

Where $\nabla_{\mathbf{m}}$ represents the gradient vector with respect to $\mathbf{m}$ and $\nabla_{\boldsymbol{\theta}}$ represents the gradient vector with respect to $\boldsymbol{\theta}$. Now, taking the gradient of Eq. 2.23 with respect to $\mathbf{m}$ and $\boldsymbol{\theta}$ gives

$$\nabla_{\mathbf{m}} O(\hat{\mathbf{m}}) = \frac{1}{2}\sum_{j=1}^{K} \frac{-2N_{dj}}{\left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]^T \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]} \nabla_{\mathbf{m}} \left[\mathbf{f}_j^T(\mathbf{m})\right] \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right] + \mathbf{C}_M^{-1}\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right) \tag{2.24}$$

and

$$\nabla_{\boldsymbol{\theta}} O(\hat{\mathbf{m}}) = -\mathbf{C}_M^{-1}\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right) + \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) \tag{2.25}$$

(Please see *Appendix A.1* for the derivations of Eqs. 2.24 and 2.25.)

At this point, it is convenient to introduce the matrix $\mathbf{G}_j$ (referred to as the sensitivity coefficient matrix for the model of the data set *j*) defined by:

$$\mathbf{G}_j = \begin{bmatrix} \frac{\partial f_{j,1}(\mathbf{m})}{\partial m_1} & \frac{\partial f_{j,1}(\mathbf{m})}{\partial m_2} & \cdots & \frac{\partial f_{j,1}(\mathbf{m})}{\partial m_M} \\ \frac{\partial f_{j,2}(\mathbf{m})}{\partial m_1} & \frac{\partial f_{j,2}(\mathbf{m})}{\partial m_2} & \cdots & \frac{\partial f_{j,2}(\mathbf{m})}{\partial m_M} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial f_{j,N_{dj}}(\mathbf{m})}{\partial m_1} & \frac{\partial f_{j,N_{dj}}(\mathbf{m})}{\partial m_2} & \cdots & \frac{\partial f_{j,N_{dj}}(\mathbf{m})}{\partial m_M} \end{bmatrix} \tag{2.26}$$

where $\mathbf{G}_j$, *j*= 1,2,…,*K*, is an $N_{dj} \times M$ matrix. Hence, it can be shown that $\nabla_{\mathbf{m}}\left[\mathbf{f}_j^T(\mathbf{m})\right]$ in Eq. 2.24 can be expressed in terms of the transpose of the matrix $\mathbf{G}_j$ as

$$\nabla_{\mathbf{m}}\left[\mathbf{f}_j^T(\mathbf{m})\right] = \begin{bmatrix} \frac{\partial}{\partial m_1} \\ \frac{\partial}{\partial m_2} \\ \vdots \\ \frac{\partial}{\partial m_M} \end{bmatrix} \left[f_{j,1}(\mathbf{m}) \;\; f_{j,2}(\mathbf{m}) \;\; \cdots \;\; f_{j,N_{dj}}(\mathbf{m}) \right] \tag{2.27}$$

$$\mathbf{G}_j^T = \nabla_{\mathbf{m}}\left[\mathbf{f}_j^T(\mathbf{m})\right] = \begin{bmatrix} \frac{\partial f_{j,1}(\mathbf{m})}{\partial m_1} & \frac{\partial f_{j,2}(\mathbf{m})}{\partial m_1} & \cdots & \frac{\partial f_{j,N_{dj}}(\mathbf{m})}{\partial m_1} \\ \frac{\partial f_{j,1}(\mathbf{m})}{\partial m_2} & \frac{\partial f_{j,2}(\mathbf{m})}{\partial m_2} & \cdots & \frac{\partial f_{j,N_{dj}}(\mathbf{m})}{\partial m_2} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial f_{j,1}(\mathbf{m})}{\partial m_M} & \frac{\partial f_{j,2}(\mathbf{m})}{\partial m_M} & \cdots & \frac{\partial f_{j,N_{dj}}(\mathbf{m})}{\partial m_M} \end{bmatrix} \qquad \textbf{(2.28)}$$

where $\mathbf{G}_j^T$, $j = 1,2,\ldots,K$, is an $M \times N_{dj}$ matrix. Then, Eq. 2.24 can be expressed as

$$\nabla_{\mathbf{m}} O(\hat{\mathbf{m}}) = \sum_{j=1}^{K} \frac{-N_{dj}}{\left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]^T \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]} \mathbf{G}_j^T \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right] + \mathbf{C}_M^{-1}\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right) \qquad \textbf{(2.29)}$$

Using Eqs. 2.28 and 2.29 in Eq. 2.23 gives the total gradient of the objective function. Note that the total gradient vector of the objective function is a $2M$-dimensional vector.

Now, we derive the overall (or total) Hessian matrix $\mathbf{H}$ needed for Gauss-Newton (G-N) or L-M method. Note that the overall Hessian matrix $\mathbf{H}$ will be $2M \times 2M$ matrix. It is important to note that using the vector-matrix calculus, the overall Hessian matrix is to be obtained as

$$\mathbf{H}(\hat{\mathbf{m}}) = \nabla\left[\left(\nabla O(\hat{\mathbf{m}})\right)^T\right] \qquad \textbf{(2.30)}$$

More explicitly using Eq. 2.23, Eq. 2.29 can be expressed as

$$\mathbf{H}(\hat{\mathbf{m}}) = \begin{bmatrix} \nabla_{\mathbf{m}} \\ \nabla_{\boldsymbol{\theta}} \end{bmatrix} \left[\left(\nabla_{\mathbf{m}} O(\hat{\mathbf{m}})\right)^T \; \left(\nabla_{\boldsymbol{\theta}} O(\hat{\mathbf{m}})\right)^T\right] \qquad \textbf{(2.31)}$$

or

$$\mathbf{H}(\hat{\mathbf{m}}) = \begin{bmatrix} \nabla_{\mathbf{m}}\left[\left(\nabla_{\mathbf{m}} O(\hat{\mathbf{m}})\right)^T\right] & \nabla_{\mathbf{m}}\left[\left(\nabla_{\boldsymbol{\theta}} O(\hat{\mathbf{m}})\right)^T\right] \\ \nabla_{\boldsymbol{\theta}}\left[\left(\nabla_{\mathbf{m}} O(\hat{\mathbf{m}})\right)^T\right] & \nabla_{\boldsymbol{\theta}}\left[\left(\nabla_{\boldsymbol{\theta}} O(\hat{\mathbf{m}})\right)^T\right] \end{bmatrix} \qquad \textbf{(2.32)}$$

Each “element” (actually a block matrix) of the Hessian matrix given by Eq. 2.32 is derived as follows:

First, it is noted the following equations which can be simply obtained by transpose of the gradient vectors given by Eqs. 2.24 and 2.29

$$\left[\nabla_{\mathbf{m}} O(\hat{\mathbf{m}})\right]^T = \left[\sum_{j=1}^{K} \frac{-N_{dj}}{\mathbf{S}_j} \mathbf{G}_j^T \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right] + \mathbf{C}_M^{-1}\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)\right]^T$$
$$= \sum_{j=1}^{K} \frac{-N_{dj}}{\mathbf{S}_j} \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]^T \mathbf{G}_j + \left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1} \quad \textbf{(2.33)}$$

where it is defined the term

$$\mathbf{S}_j = \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]^T \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right] \quad \textbf{(2.34)}$$

for simplification, and

$$\left(\nabla_{\boldsymbol{\theta}} O(\hat{\mathbf{m}})\right)^T = -\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1} + \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1} \quad \textbf{(2.35)}$$

Next, it is taken the gradients of Eqs. 2.33 and 2.35 with respect to $\mathbf{m}$ and $\boldsymbol{\theta}$ to obtain

$$\nabla_{\mathbf{m}}\left[\nabla_{\mathbf{m}} O(\hat{\mathbf{m}})\right]^T = \nabla_{\mathbf{m}}\left[\sum_{j=1}^{K} -\frac{N_{dj}}{\mathbf{S}_j}\left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]^T \mathbf{G}_j + \left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\right]$$
$$= \mathbf{C}_M^{-1} + \sum_{j=1}^{K} \frac{N_{dj}}{\mathbf{S}_j} \mathbf{G}_j^T \mathbf{G}_j - \sum_{j=1}^{K} \frac{N_{dj}}{\mathbf{S}_j^2} \mathbf{G}_j^T \mathbf{r}_j \mathbf{r}_j^T \mathbf{G}_j + \sum_{j=1}^{K} \frac{N_{dj}}{\mathbf{S}_j} \sum_{i=1}^{N_{dj}} r_{ij} \nabla^2 r_{ij} \quad \textbf{(2.36)}$$

(Please see *Appendix A.1* for the derivation of Eq. 2.36) where the vector $\mathbf{r}_\mathbf{j}$ is defined as

$$\mathbf{r}_j = \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right] \quad \textbf{(2.37)}$$

And $\mathbf{r}_{ij}$ is the i^{th} component of the vector $\mathbf{r}_j$. $\nabla^2 r_{ij}$ is the matrix of second derivatives of the $\mathbf{r}_{ij}$.

$$\nabla_{\mathbf{m}}\left[\left(\nabla_{\boldsymbol{\theta}} O(\hat{\mathbf{m}})\right)^T\right] = \nabla_{\mathbf{m}}\left[-\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1} + \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\right] = -\mathbf{C}_M^{-1} \quad \textbf{(2.38)}$$

$$\nabla_{\boldsymbol{\theta}}\left[\left(\nabla_{\mathbf{m}} O(\hat{\mathbf{m}})\right)^T\right] = \nabla_{\boldsymbol{\theta}}\left[\sum_{j=1}^{K} -\left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]^T \mathbf{C}_{Dj}^{-1} \mathbf{G}_j + \left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\right]$$
$$= -\mathbf{C}_M^{-1} \quad \textbf{(2.39)}$$

and

$$\nabla_{\theta}\left[\left(\nabla_{\theta}O(\hat{\mathbf{m}})\right)^T\right] = \nabla_{\theta}\left[-\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T\mathbf{C}_M^{-1}+\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T\mathbf{C}_\theta^{-1}\right]=\mathbf{C}_M^{-1}+\mathbf{C}_\theta^{-1} \qquad \textbf{(2.40)}$$

Using Eqs. 2.36-2.40 in Eq. 2.32 yields the overall Hessian matrix for the Newton method. However, it should be noted in Gauss-Newton and L-M methods, Eq. 2.36 is approximated as:

$$\nabla_{\mathbf{m}}\left[\nabla_{\mathbf{m}}O(\hat{\mathbf{m}})\right]^T \approx \mathbf{C}_M^{-1}+\sum_{j=1}^{K}\frac{N_{dj}}{\mathbf{S}_j}\mathbf{G}_j^T\mathbf{G}_j \qquad \textbf{(2.41)}$$

Hence, the approximate Hessian (or G-N or L-M Hessian matrix) can be expressed as

$$\mathbf{H}(\hat{\mathbf{m}})=\begin{bmatrix}\sum_{j=1}^{K}\left[\frac{N_{dj}}{\mathbf{S}_j}\mathbf{G}_j^T\mathbf{G}_j\right]+\mathbf{C}_M^{-1} & -\mathbf{C}_M^{-1}\\ -\mathbf{C}_M^{-1} & \mathbf{C}_M^{-1}+\mathbf{C}_\theta^{-1}\end{bmatrix} \qquad \textbf{(2.42)}$$

As considered cases where the total number of observed data ($N_{dj} = N_{d1}+N_{d2}+\ldots+N_{dK}$) is much larger than the unknown model parameters $2M$, and most importantly, $\mathbf{C}_M$ is a diagonal matrix in our applications (see Eq. 2.37), it can be worked directly with the Hessian matrix given by Eq. 2.41. However, it may be tried to use a further approximated Hessian matrix in the G-N or L-M method, where the off-block matrix $\mathbf{C}_M^{-1}$ in Eq. 2.41 is set to the MxM null matrix, $\mathbf{O}$. Hence, we may consider using the following modified Hessian matrix in the L-M method:

$$\mathbf{H}(\hat{\mathbf{m}})=\begin{bmatrix}\sum_{j=1}^{K}\left[\frac{N_{dj}}{\mathbf{S}_j}\mathbf{G}_j^T\mathbf{G}_j\right]+\mathbf{C}_M^{-1} & \mathbf{O}\\ \mathbf{O} & \mathbf{C}_M^{-1}+\mathbf{C}_\theta^{-1}\end{bmatrix} \qquad \textbf{(2.43)}$$

The basic L-M algorithm for minimizing an arbitrary objective function (in this case Eq. 2.21 or 2.22) can be given as

$$\left(\lambda_l\mathbf{I}+\mathbf{H}_l\right)\delta\hat{\mathbf{m}}^{l+1}=-\nabla O(\hat{\mathbf{m}}^{l+1}) \qquad \textbf{(2.44)}$$

and

$$\hat{\mathbf{m}}_c^{l+1}=\hat{\mathbf{m}}^l+\delta\hat{\mathbf{m}}^{l+1} \qquad \textbf{(2.45)}$$

It should be noted that it can be used either the Hessian matrix given by Eq. 2.42 or 2.43. However, if Eq. 2.43 is used, then it may be obtained a simpler computational scheme which can be described as

$$\left(\lambda_l \mathbf{I} + \sum_{j=1}^{K}\left[\frac{N_{dj}}{\mathbf{S}_j}\mathbf{G}_j^T\mathbf{G}_j + \mathbf{C}_M^{-1}\right]\right)\delta\mathbf{m}^{l+1} = -\left[\sum_{j=1}^{K}\frac{N_{dj}}{\mathbf{S}_j}\mathbf{G}_j^T\left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right] + \mathbf{C}_M^{-1}\left(\mathbf{m} - \mathbf{m}_{pr} - \theta\right)\right] \quad \textbf{(2.46)}$$

and

$$\left[\lambda_l \mathbf{I} + \left(\mathbf{C}_M^{-1} + \mathbf{C}_\theta^{-1}\right)\right]\delta\theta^{l+1} = -\left[-\mathbf{C}_M^{-1}\left(\mathbf{m}^l - \mathbf{m}_{pr} - \theta^l\right) + \mathbf{C}_\theta^{-1}\left(\theta^l - \theta_0\right)\right] \quad \textbf{(2.47)}$$

$$\mathbf{m}_c^{l+1} = \mathbf{m}^l + \delta\mathbf{m}^{l+1} \quad \textbf{(2.48)}$$

and

$$\theta_c^{l+1} = \theta^l + \delta\theta^{l+1} \quad \textbf{(2.49)}$$

MLE is used for a simple case in Chapter 3.

2.3 Least-Squares Estimation Method

The basic principle of least-squares (LS) is developed by German mathematician and scientist Carl F. Gauss around 1794, however, French mathematician Adrien M. Legendre was the first to publish the method independently. Gauss did not publish the method until 1809, when it appeared in volume two of his work on celestial mechanics, *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium (Theory of Celestial Bodies in the Section of Conicarum Surrounding the Sun)*. In 1822, Gauss was able to state that the least-squares approach to regression analysis is optimal in the sense that in a linear model where the errors have a mean of zero, are uncorrelated, and have equal variances, the best linear unbiased estimator of the coefficients is the least-squares estimator. This result is known as the Gauss–Markov theorem (Url-1).

If we assume that the data error covariance matrix $\mathbf{C}_D$ is known in Eq. 2.8, i.e., the weights for observed data are known, and then to generate the most probable model is obtained by minimizing the argument of the exponential function given by Eq. 2.8:

$$\begin{aligned} O(\mathbf{m},\theta) = &\frac{1}{2}\left[\mathbf{y} - \mathbf{f}(\mathbf{m})\right]^T \mathbf{C}_D^{-1}\left[\mathbf{y} - \mathbf{f}(\mathbf{m})\right] \\ &+\frac{1}{2}\left(\mathbf{m} - \mathbf{m}_{pr} - \theta\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m} - \mathbf{m}_{pr} - \theta\right) \\ &+\frac{1}{2}\left(\theta - \theta_0\right)^T \mathbf{C}_\theta^{-1}\left(\theta - \theta_0\right) \end{aligned} \quad \textbf{(2.50)}$$

If we have K sets of observed data and assume that each $\mathbf{C}_{Dj}$ is known in Eq. 2.9, then the most probable model is obtained by minimizing the following objective function:

$$O(\mathbf{m},\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=1}^{K}\left(\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right)^T \mathbf{C}_{Dj}^{-1}\left(\mathbf{y}_{obs,j} - \mathbf{f}_j(\mathbf{m})\right)$$
$$+\frac{1}{2}\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m} - \mathbf{m}_{pr} - \boldsymbol{\theta}\right) \quad \textbf{(2.51)}$$
$$+\frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)$$

It should be noted that the least-squares is equivalent to the maximum likelihood if the observed data or experimental errors have a Gaussian (normal) distribution (for example see Kuchuk et al. 1990).

In the simple case where each $\mathbf{C}_{Dj}$ is diagonal with all diagonal elements equal to the same variance, but different for each data set, and the matrices $\mathbf{C}_M$ and $\mathbf{C}_\theta$ are diagonal, then the objective function given by Eq. 2.51 is simply expressed as

$$O(\mathbf{m},\boldsymbol{\theta}) = \frac{1}{2}\left[\sum_{j=1}^{K}\frac{1}{\sigma_j^2}\sum_{i=1}^{m}\left[y_i - f_i(\mathbf{m})\right]^2\right]$$
$$+\frac{1}{2}\sum_{i=1}^{M}\frac{(m - m_{pr} - \theta)_i^2}{\sigma_{m,i}^2} + \frac{1}{2}\sum_{i=1}^{M}\frac{(\theta - \theta_o)_i^2}{\sigma_{\theta,i}^2} \quad \textbf{(2.52)}$$

where σ_j^2 is diagonal entries of $\boldsymbol{C}_{Dj}$, $j=1,2,..,N_d$.

2.3.1 An application for a simple case

For simplicity, it will be assumed that the vector $\mathbf{m}$ contains the same single parameter, $\mathbf{m}$, and model, f is simply equal to $\mathbf{m}$, $\mathbf{f}(\mathbf{m}) = \mathbf{m}$. Further suppose that $\mathbf{y}$ contains N observed (measured) value of $\mathbf{m}$ with noise having zero mean and a standard deviation equal to σ_d. Then, assuming diagonal covariances, Eq. 2.51 can be expressed as

$$O(m,\theta) = \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\mathrm{y}_i - m}{\sigma_d}\right)^2 + \frac{1}{2}\left(\frac{m - m_{pr} - \theta}{\sigma_m}\right)^2 + \frac{1}{2}\left(\frac{\theta - \theta_0}{\sigma_\theta}\right)^2 \quad \textbf{(2.53)}$$

Taking the derivatives of the objective function given by Eq. 2.53 with respect to $\mathbf{m}$ and $\boldsymbol{\theta}$ and equating them to zero gives, respectively,

$$\frac{\partial O(m,\theta)}{\partial m} = \frac{1}{2}\sum_{i=1}^{N}2\left(\frac{\mathrm{y}_i - m}{\sigma_d}\right)\left(\frac{-1}{\sigma_d}\right) + \frac{2}{2}\left(\frac{m - m_{pr} - \theta}{\sigma_m}\right)\left(\frac{1}{\sigma_m}\right) = 0 \quad \textbf{(2.54)}$$

and

$$\frac{\partial O(m,\theta)}{\partial\theta}=\frac{2}{2}\left(\frac{m-m_{pr}-\theta}{\sigma_m}\right)\left(\frac{-1}{\sigma_m}\right)+\frac{2}{2}\left(\frac{\theta-\theta_0}{\sigma_\theta}\right)\left(\frac{1}{\sigma_\theta}\right)=0 \qquad \textbf{(2.55)}$$

Eqs. 2.54 and 2.55 can be simplified and rearranged, respectively, as

$$\left(\frac{N}{\sigma_d^2}+\frac{1}{\sigma_m^2}\right)m-\frac{1}{\sigma_m^2}\theta=\frac{1}{\sigma_d^2}\sum_{i=1}^{N}\mathrm{y}_i+\frac{m_{pr}}{\sigma_m^2} \qquad \textbf{(2.56)}$$

and

$$\frac{-1}{\sigma_m^2}m+\left(\frac{1}{\sigma_m^2}+\frac{1}{\sigma_\theta^2}\right)\theta=-\frac{m_{pr}}{\sigma_m^2}+\frac{\theta_0}{\sigma_\theta^2} \qquad \textbf{(2.57)}$$

Solving Eqs. 56 and 57 for m and θ gives, respectively,

$$\tilde{m}=\frac{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\frac{1}{\sigma_d^2}\sum_{i=1}^{N}\mathrm{y}_i+\left(\frac{m_{pr}+\theta_0}{\sigma_\theta^2}\right)}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\left[\frac{N}{\sigma_d^2}+\frac{1}{\sigma_m^2}-\frac{1}{\sigma_m^2\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)}\right]} \qquad \textbf{(2.58)}$$

or

$$\tilde{m}=\frac{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\frac{1}{\sigma_d^2}\sum_{i=1}^{N}\mathrm{y}_i+\left(\frac{m_{pr}+\theta_0}{\sigma_\theta^2}\right)}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\left[\frac{N}{\sigma_d^2}+\frac{1}{\sigma_\theta^2\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)}\right]} \qquad \textbf{(2.59)}$$

or

$$\tilde{\mathrm{m}}=\frac{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\frac{1}{\sigma_d^2}\sum_{i=1}^{N}\mathrm{y}_i+\left(\frac{m_{pr}+\theta_0}{\sigma_\theta^2}\right)}{\frac{N}{\sigma_d^2}+\frac{N}{\sigma_d^2}\frac{\sigma_m^2}{\sigma_\theta^2}+\frac{1}{\sigma_\theta^2}} \qquad \textbf{(2.60)}$$

and

$$\tilde{\theta} = \frac{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\frac{1}{\sigma_d^2}\sum_{i=1}^{N}\mathrm{y}_i + \left(\frac{m_{pr}+\theta_0}{\sigma_\theta^2}\right)}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)^2\left[\frac{N}{\sigma_d^2}+\frac{1}{\sigma_m^2}-\frac{1}{\sigma_m^2\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)}\right]} - \frac{\left(m_{pr}-\frac{\sigma_m^2}{\sigma_\theta^2}\theta_0\right)}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)} = \frac{\tilde{m}-\left(m_{pr}-\frac{\sigma_m^2}{\sigma_\theta^2}\theta_0\right)}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)} \quad \textbf{(2.61)}$$

Note that if we take $\theta_0 = 0$, then Eqs. 2.57 and 2.61 become

$$\tilde{m} = \frac{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\frac{1}{\sigma_d^2}\sum_{i=1}^{N}\mathrm{y}_i + \frac{m_{pr}}{\sigma_\theta^2}}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\left[\frac{N}{\sigma_d^2}+\frac{1}{\sigma_m^2}-\frac{1}{\sigma_m^2\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)}\right]} \quad \textbf{(2.62)}$$

and

$$\tilde{\theta} = \frac{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\frac{1}{\sigma_d^2}\sum_{i=1}^{N}\mathrm{y}_i + \frac{m_{pr}}{\sigma_\theta^2}}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)^2\left[\frac{N}{\sigma_d^2}+\frac{1}{\sigma_m^2}-\frac{1}{\sigma_m^2\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)}\right]} - \frac{m_p}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)} = \frac{\tilde{\mathrm{m}}-m_{pr}}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)} \quad \textbf{(2.63)}$$

In Eqs. 2.57-2.63, $\tilde{m}$ and $\tilde{\theta}$ represent the posterior estimates (after conditioning the data) of m and θ.

Next, we derive the estimates for some limiting cases:

Case (1) Suppose the variance of θ approaches infinity, i.e., $\sigma_\theta^2 \to \infty$, then it can be shown that Eqs. 2.57 and 2.58(and also Eqs. 2.62 and 2.63) reduce to

$$\tilde{m} = \frac{\frac{1}{\sigma_d^2}\sum_{i=1}^{N}\mathrm{y}_i}{\frac{N}{\sigma_d^2}} = \frac{\sum_{i=1}^{N}\mathrm{y}_i}{N} \quad \textbf{(2.64)}$$

and

$$\tilde{\theta} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{y}_i - m_{pr} \quad \textbf{(2.65)}$$

Case (2) Suppose the variances of m and θ approaches infinity, simultaneously, i.e., $\sigma_\theta^2 \to \infty$ and $\sigma_m^2 \to \infty$, then it can be shown that Eqs. 2.57 and 2.58 reduce to

$$\tilde{m} = \frac{(1+1)\dfrac{1}{\sigma_d^2}\sum_{i=1}^{N} \mathrm{y}_i}{(1+1)\left[\dfrac{N}{\sigma_d^2}\right]} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{y}_i \qquad \textbf{(2.66)}$$

and

$$\tilde{\theta} = \frac{\tilde{m} - (m_{pr} - \theta_0)}{(1+1)} = \frac{1}{2}\left(\frac{1}{N}\sum_{i=1}^{N} \mathrm{y}_i - m_{pr} + \theta_0\right) \qquad \textbf{(2.67)}$$

Case (3) Suppose the variance of m approaches infinity, i.e., $\sigma_m^2 \to \infty$, then it can be shown that Eqs. 2.57 and 2.58 reduce to

$$\tilde{m} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{y}_i \qquad \textbf{(2.68)}$$

and

$$\tilde{\theta} = \theta_0 \qquad \textbf{(2.69)}$$

Case (4) Suppose the variance of m approaches zero, i.e., $\sigma_m^2 \to 0$, then it can be shown that Eqs. 2.57 and 2.58 reduce to

$$\tilde{m} = \frac{\dfrac{1}{\sigma_d^2}\sum_{i=1}^{N} \mathrm{y}_i + \left(\dfrac{m_{pr} + \theta_0}{\sigma_\theta^2}\right)}{\dfrac{N}{\sigma_d^2} + \dfrac{1}{\sigma_\theta^2}} \qquad \textbf{(2.70)}$$

and

$$\tilde{\theta} = \frac{\dfrac{1}{\sigma_d^2}\sum_{i=1}^{N} \mathrm{y}_i + \left(\dfrac{m_{pr} + \theta_0}{\sigma_\theta^2}\right)}{\dfrac{N}{\sigma_d^2} + \dfrac{1}{\sigma_\theta^2}} - m_{pr} \qquad \textbf{(2.71)}$$

Case (5) Suppose the variance of θ approaches zero, i.e., $\sigma_\theta^2 \to 0$, then it can be shown that Eqs. 2.57 and 2.58 reduce to

$$\tilde{m} = \frac{\frac{1}{\sigma_d^2}\sum_{i=1}^{N} y_i}{\left(\frac{N}{\sigma_d^2} + \frac{1}{\sigma_m^2}\right)} + \frac{m_{pr} + \theta_0}{\sigma_m^2\left(\frac{N}{\sigma_d^2} + \frac{1}{\sigma_m^2}\right)} \quad \textbf{(2.72)}$$

and

$$\tilde{\theta} = \theta_0 \quad \textbf{(2.73)}$$

Case (6) Suppose the variances of m and θ approach zero, simultaneously, i.e., $\sigma_m^2 \to 0$ and $\sigma_\theta^2 \to 0$, then it can be shown that Eqs. 2.57 and 2.58 reduce to

$$\tilde{m} = m_{pr} + \theta_0 \quad \textbf{(2.74)}$$

and

$$\tilde{\theta} = \theta_0 \quad \textbf{(2.75)}$$

Now, it will be inspected some statistical properties of $\tilde{m}$ and $\tilde{\theta}$. Specifically, it is wanted to see what the expectations and variances of the estimates $\tilde{m}$ and $\tilde{\theta}$ given by Eqs. 2.57 and 2.58. First, their expectations are derived. For this purpose, E will be defined to be the expectation operator and hence

$$E(\tilde{m}) = E\left[\frac{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\frac{1}{\sigma_d^2}\sum_{i=1}^{N} y_i + \left(\frac{m_{pr}+\theta_0}{\sigma_\theta^2}\right)}{\frac{N}{\sigma_d^2} + \frac{N}{\sigma_d^2}\frac{\sigma_m^2}{\sigma_\theta^2} + \frac{1}{\sigma_\theta^2}}\right] = \frac{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)\frac{Nm_{true}}{\sigma_d^2} + \left(\frac{m_{pr}+\theta_0}{\sigma_\theta^2}\right)}{\frac{N}{\sigma_d^2} + \frac{N}{\sigma_d^2}\frac{\sigma_m^2}{\sigma_\theta^2} + \frac{1}{\sigma_\theta^2}} \quad \textbf{(2.76)}$$

where m_{true} represents the true value of m. Note that the expectation of $\tilde{m}$ will be identical to m_{true} if and only if $m_{pr} + \theta_0 = m_{true}$.

Similarly, the expectation of θ is given by

$$E(\tilde{\theta}) = \frac{E(\tilde{m}) - \left(m_{pr} - \frac{\sigma_m^2}{\sigma_\theta^2}\theta_0\right)}{\left(1+\frac{\sigma_m^2}{\sigma_\theta^2}\right)} \quad \textbf{(2.77)}$$

The inverse of the variance of $\tilde{m}$ can be found by taking derivative of Eq. 2.53 with respect to m and it will be obtained the variance of $\tilde{m}$ as:

$$\sigma_{\tilde{m}}^2 = \frac{1}{\left(\frac{N}{\sigma_d^2} + \frac{1}{\sigma_m^2}\right)} \tag{2.78}$$

which represents the posterior variance of the model parameter m, given the data and the value of the θ.

The inverse of posterior variance of θ is not quite as simple of the posterior variance of m, because it is wanted the posterior variance given the data, not the model and the data. So, it is needed to calculate $\partial^2 O / \partial^2 \theta$ at the posterior estimate of m. Note that the first derivative $\partial O / \partial \theta$ is given by

$$\frac{\partial O}{\partial \theta} = \frac{-1}{\sigma_m^2} m + \left(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_\theta^2}\right)\theta + \frac{m_{pr}}{\sigma_m^2} - \frac{\theta_0}{\sigma_\theta^2} \tag{2.79}$$

Now solving Eq. 2.56 for m in terms of θ to obtain:

$$m = \left(\frac{N}{\sigma_d^2} + \frac{1}{\sigma_m^2}\right)^{-1} \left(\frac{1}{\sigma_m^2}\theta + \frac{1}{\sigma_d^2}\sum_{i=1}^{N} \mathrm{y}_i + \frac{m_{pr}}{\sigma_m^2}\right) \tag{2.80}$$

Using Eq. 2.80 in Eq. 2.79 gives

$$\frac{\partial O}{\partial \theta} = \frac{-1}{\sigma_m^2}\left(\frac{N}{\sigma_d^2} + \frac{1}{\sigma_m^2}\right)^{-1} \left(\frac{1}{\sigma_m^2}\theta + \frac{1}{\sigma_d^2}\sum_{i=1}^{N} \mathrm{y}_i + \frac{m_{pr}}{\sigma_m^2}\right) + \left(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_\theta^2}\right)\theta + \frac{m_{pr}}{\sigma_m^2} - \frac{\theta_0}{\sigma_\theta^2} \tag{2.81}$$

Differentiating Eq. 2.81 with respect to θ gives the inverse of the variance of posterior variance

$$\left(\sigma_{\tilde{\theta}}^2\right)^{-1} = \frac{-1}{\sigma_m^2}\left(\frac{N}{\sigma_d^2} + \frac{1}{\sigma_m^2}\right)^{-1}\left(\frac{1}{\sigma_m^2}\right) + \left(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_\theta^2}\right) \tag{2.82}$$

or

$$\left(\sigma_{\tilde{\theta}}^2\right) = \frac{1}{\left(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_\theta^2}\right) - \left(\frac{1}{\sigma_m^4}\right)\left(\frac{N}{\sigma_d^2} + \frac{1}{\sigma_m^2}\right)^{-1}} \tag{2.83}$$

An Example Application:

Let's consider 10 values of observed m, with Gaussian error with mean 0 and data variance 6.25. The true value of m is 10. Tabulated in Table 2.1 are values of m with error.

Table 2.1: Noisy m values

Data values of m
7.885
6.259
13.143
12.187
10.057
9.995
9.139
6.551
7.858
12.839

It will be studied the following cases, and in all cases it will be considered WLS minimization with known error variance, σ_d^2.

Case 1: No prior term in the objective function, i.e.,

$$O(m,\theta) = \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\mathrm{y}_i - m}{\sigma_d}\right)^2 \quad \textbf{(2.84)}$$

Case 2: A prior term in the objective function, i.e.,

$$O(m,\theta) = \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\mathrm{y}_i - m}{\sigma_d}\right)^2 + \frac{1}{2}\left(\frac{m - m_{pr}}{\sigma_m}\right)^2 \quad \textbf{(2.85)}$$

Case 3: An uncertain prior mean in the objective function

$$O(m,\theta) = \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\mathrm{y}_i - m}{\sigma_d}\right)^2 + \frac{1}{2}\left(\frac{m - m_{pr} - \theta}{\sigma_m}\right)^2 + \frac{1}{2}\left(\frac{\theta - \theta_0}{\sigma_\theta}\right)^2 \quad \textbf{(2.86)}$$

Results:

Case 1. The estimate of *m* minimizing Eq. 2.84 is calculated from

$$\tilde{m} = \frac{\sum_{i=1}^{N}\mathrm{y}_i}{N} = \frac{95.913}{10} = 9.5913 \quad \textbf{(2.87)}$$

The variance of this estimate is calculated from

$$\tilde{\sigma}_d^2 = \frac{\sigma_d^2}{N} = \frac{6.25}{10} = 0.625 \quad \textbf{(2.88)}$$

As it is expected, without prior the model parameter is arithmetic average of data. In general, if the prior means are not considered, the value of model parameter goes to the mean of data.

Case 2: The estimate of m minimizing Eq. 2.85 is calculated from

$$\tilde{m} = \frac{\dfrac{\sum_{i=1}^{N} y_i}{\sigma_d^2} + \dfrac{m_{pr}}{\sigma_m^2}}{\dfrac{N}{\sigma_d^2} + \dfrac{1}{\sigma_m^2}} = \frac{\dfrac{95.913}{6.25} + \dfrac{m_{pr}}{\sigma_m^2}}{\dfrac{10}{6.25} + \dfrac{1}{\sigma_m^2}} \quad \textbf{(2.89)}$$

The variance of this estimate is calculated from

$$\sigma_{\tilde{m}}^2 = \frac{1}{\dfrac{N}{\sigma_d^2} + \dfrac{1}{\sigma_m^2}} = \frac{1}{\dfrac{10}{6.25} + \dfrac{1}{\sigma_m^2}} \quad \textbf{(2.90)}$$

Table 2.2 presents twelve different prior models with their corresponding m_{pr} and σ_m^2 and the computed values of $\tilde{m}$ and $\sigma_{\tilde{m}}^2$ from Eqns. 2.89 and 2.90 for each prior model described in Table 2.2.

Table 2.2: Different Prior Models and The Estimation of $\tilde{m}$ and $\sigma_{\tilde{m}}^2$.

Prior Models	m_{pr}	σ_m^2	$\tilde{m}$	$\sigma_{\tilde{m}}^2$
Prior Model 1	10	0.01	9.99356	0.00984
Prior Model 2	10	1	9.74849	0.38462
Prior Model 3	10	10	9.61534	0.58824
Prior Model 4	10	100	9.59384	0.62112
Prior Model 5	0	0.01	0.15104	0.00984
Prior Model 6	0	1	5.90234	0.38462
Prior Model 7	0	10	9.02711	0.58824
Prior Model 8	0	100	9.53173	0.62112
Prior Model 9	100	0.01	98.5762	0.00984
Prior Model 10	100	1	44.3639	0.38462
Prior Model 11	100	10	14.9095	0.58824
Prior Model 12	100	100	10.1528	0.62112

Case 3: The estimates of m and θ minimizing Eq. 2.86 is calculated from Eqs. 2.57 and 2.58. The associated variances for these parameters can be computed from Eqns. 2.95 and 2.100.

As the objective of using partially doubly stochastic model is to adjust the mean in cases where prior mean is incorrect, it will be considered a few incorrect prior models given in Table 2.1. For the purpose of illustration, it will be considered Prior Models 5-12. It will be set $\theta_0 = 0$ in applications to be given. We selected different values of σ_θ^2. Table 2.3 presents the results obtained for the Prior Model 5 for six different values of σ_θ^2.

Table 2.3: Estimates of $\tilde{m}$, $\tilde{\theta}$, $\sigma_{\tilde{m}}^2$ and $\sigma_{\tilde{\theta}}^2$ for Prior Model 5 given in Table 2.1.

Variances	$\tilde{m}$	$\tilde{\theta}$	$\sigma_{\tilde{m}}^2$	$\sigma_{\tilde{\theta}}^2$
$\sigma_\theta^2 = 0.01$	0.2974	0.1487	0.0098	0.0098
$\sigma_\theta^2 = 0.1$	1.4354	1.3049	0.0098	0.0864
$\sigma_\theta^2 = 1$	5.9249	5.8662	0.0098	0.3884
$\sigma_\theta^2 = 10$	9.0276	9.0186	0.0098	0.5971
$\sigma_\theta^2 = 100$	9.5317	9.5308	0.0098	0.6310
$\sigma_\theta^2 = 1000$	9.5853	9.5852	0.0098	0.6346

Table 2.4: Estimates of $\tilde{m}$, $\tilde{\theta}$, $\sigma_{\tilde{m}}^2$ and $\sigma_{\tilde{\theta}}^2$ for Prior Model 6 given in Table 2.1.

Variances	$\tilde{m}$	$\tilde{\theta}$	$\sigma_{\tilde{m}}^2$	$\sigma_{\tilde{\theta}}^2$
$\sigma_\theta^2 = 0.01$	5.9293	5.8662×10^{-2}	0.3846	0.0099
$\sigma_\theta^2 = 0.1$	6.1162	0.5560	0.3846	0.0942
$\sigma_\theta^2 = 1$	7.3076	3.6538	0.3846	0.6190
$\sigma_\theta^2 = 10$	9.0756	8.2506	0.3846	1.3978
$\sigma_\theta^2 = 100$	9.5323	9.4379	0.3846	1.5990
$\sigma_\theta^2 = 1000$	9.5853	9.5757	0.3846	1.6224

Table 2.5: Estimates of $\tilde{m}$, $\tilde{\theta}$, $\sigma_{\tilde{m}}^2$ and $\sigma_{\tilde{\theta}}^2$ for Prior Model 7 given in Table 2.1.

Variances	$\tilde{m}$	$\tilde{\theta}$	$\sigma_{\tilde{m}}^2$	$\sigma_{\tilde{\theta}}^2$
$\sigma_\theta^2 = 0.01$	9.0276	9.0186×10^{-3}	0.5882	0.0100
$\sigma_\theta^2 = 0.1$	9.0323	8.9429×10^{-2}	0.5882	0.0991
$\sigma_\theta^2 = 1$	9.0756	0.8250	0.5882	0.9140
$\sigma_\theta^2 = 10$	9.3006	4.6503	0.5882	5.1515
$\sigma_\theta^2 = 100$	9.5371	8.6701	0.5882	9.6045
$\sigma_\theta^2 = 1000$	9.5854	9.4905	0.5882	10.5133

Table 2.6: Estimates of $\tilde{m}$, $\tilde{\theta}$, $\sigma_{\tilde{m}}^2$ and $\sigma_{\tilde{\theta}}^2$ for Prior Model 8 given in Table 2.1.

Variances	$\tilde{m}$	$\tilde{\theta}$	$\sigma^2_{\tilde{m}}$	$\sigma^2_{\tilde{\theta}}$
$\sigma^2_\theta = 0.01$	9.5317	9.5308×10^{-4}	0.6211	0.0100
$\sigma^2_\theta = 0.1$	9.5318	9.5222×10^{-3}	0.6211	0.0999
$\sigma^2_\theta = 1$	9.5323	9.4379×10^{-2}	0.6211	0.9902
$\sigma^2_\theta = 10$	9.5371	0.8670	0.6211	9.0960
$\sigma^2_\theta = 100$	9.5614	4.7807	0.6211	50.1558
$\sigma^2_\theta = 1000$	9.5858	8.7144	0.6211	91.4253

As it can be seen in Tables 2.3 and 2.4, when the prior mean is incorrect and variance of prior is pretty small, the value of $\tilde{m}$ is changing in a positive way. Even the value σ^2_θ is small, the results are better than using just prior mean especially if the value of σ^2_θ is taken bigger i.e., 100 or 1000. In addition, Tables 2.5 and 2.6 show that, if it is not trusted the prior mean then it must be taken the values of prior and correction variances bigger. Thus, how so ever great the error in prior mean may be, the value of $\tilde{m}$ goes through the true value even if the value of variances are small.

Table 2.7: Estimates of $\tilde{m}$, $\tilde{\theta}$, $\sigma^2_{\tilde{m}}$ and $\sigma^2_{\tilde{\theta}}$ for Prior Model 9 given in Table 2.1.

Variances	$\tilde{m}$	$\tilde{\theta}$	$\sigma^2_{\tilde{m}}$	$\sigma^2_{\tilde{\theta}}$
$\sigma^2_\theta = 0.01$	97.1966	-1.4016	0.0098	0.0098
$\sigma^2_\theta = 0.1$	86.4694	-12.301	0.0098	0.0864
$\sigma^2_\theta = 1$	44.1512	-55.296	0.0098	0.3884
$\sigma^2_\theta = 10$	14.9044	-85.011	0.0098	0.5971
$\sigma^2_\theta = 100$	10.1528	-89.838	0.0098	0.6310
$\sigma^2_\theta = 1000$	9.6477	-90.351	0.0098	0.6346

Table 2.8: Estimates of $\tilde{m}$, $\tilde{\theta}$, $\sigma^2_{\tilde{m}}$ and $\sigma^2_{\tilde{\theta}}$ for Prior Model 10 given in Table 2.1.

Variances	$\tilde{m}$	$\tilde{\theta}$	$\sigma^2_{\tilde{m}}$	$\sigma^2_{\tilde{\theta}}$
$\sigma^2_\theta = 0.01$	44.1512	-0.5530	0.3846	0.0099
$\sigma^2_\theta = 0.1$	42.3481	-5.2411	0.3846	0.0942
$\sigma^2_\theta = 1$	31.1172	-34.4414	0.3846	0.6190
$\sigma^2_\theta = 10$	14.4520	-77.7709	0.3846	1.3978
$\sigma^2_\theta = 100$	10.1473	-88.9631	0.3846	1.5990
$\sigma^2_\theta = 1000$	9.6477	-90.2620	0.3846	1.6224

Table 2.9: Estimates of $\tilde{m}$, $\tilde{\theta}$, $\sigma^2_{\tilde{m}}$ and $\sigma^2_{\tilde{\theta}}$ for Prior Model 11 given in Table 2.1.

Variances	$\tilde{m}$	$\tilde{\theta}$	$\sigma^2_{\tilde{m}}$	$\sigma^2_{\tilde{\theta}}$
$\sigma^2_{\theta} = 0.01$	14.9045	-0.0850	0.5882	0.0100
$\sigma^2_{\theta} = 0.1$	14.8599	-0.8430	0.5882	0.0991
$\sigma^2_{\theta} = 1$	14.4520	-7.7771	0.5882	0.9140
$\sigma^2_{\theta} = 10$	12.3310	-43.8345	0.5882	5.1515
$\sigma^2_{\theta} = 100$	10.1021	-81.7254	0.5882	9.6045
$\sigma^2_{\theta} = 1000$	9.6472	-89.4582	0.5882	10.5133

As it is clearly seen in Tables 2.7 and 2.8, when the prior mean is quite incorrect and variance of prior is small, the value of $\tilde{m}$ goes to incorrect prior mean. However, Tables 2.9 and 2.10 shows that while using wrong prior mean, if the values of $\sigma^2_{\tilde{m}}$ and σ^2_{θ} are assumed to be bigger the value of model parameter approaches the real one.

Table 2.10: Estimates of $\tilde{m}$, $\tilde{\theta}$, $\sigma^2_{\tilde{m}}$ and $\sigma^2_{\tilde{\theta}}$ for Prior Model 12 given in Table 2.1.

Variances	$\tilde{m}$	$\tilde{\theta}$	$\sigma^2_{\tilde{m}}$	$\sigma^2_{\tilde{\theta}}$
$\sigma^2_{\theta} = 0.01$	10.1528	-0.0090	0.6211	0.0100
$\sigma^2_{\theta} = 0.1$	10.1523	-0.0898	0.6211	0.0999
$\sigma^2_{\theta} = 1$	10.1473	-0.8896	0.6211	0.9902
$\sigma^2_{\theta} = 10$	10.1021	-8.1725	0.6211	9.0960
$\sigma^2_{\theta} = 100$	9.8729	-45.0635	0.6211	50.1558

2.3.2 Unweighted least-squares parameter estimation method

One of the methods for minimization of the objective function is unweighted least-squares estimation method (UWLS). It is considered that each data point has the same importance. Because there is no weight on error variances or the errors in the original measurements are uncorrelated, the method does not need to covariance matrix, C_D. It is assumed that all diagonal elements (variances) is equal to each other, $\sigma^2_j = \sigma^2$, j=1,2,…,N_d, then the covariance matrix, $C_D = I\sigma^2$ and it is assumed that the variance is known. Eq. 2.52 (objective of ML) is simplified form of least-squares. It should be realized that the ML of m is identical to the unweighted least-squares (UWLS) of m which in case the diagonal variances are equal each other, i.e.,

$\sigma_j^2 = \sigma^2$ for all *j*, notwithstanding consideration if σ^2 is known or unknown (Kuchuk et al., 2010).

In this case, for a partially doubly stochastic model, it will be minimized the following objective function for *N* observed data points,

$$\begin{aligned} O(\mathbf{m},\boldsymbol{\theta}) &= \frac{1}{2}\frac{1}{\sigma^2}\left[\mathbf{y}-\mathbf{f}\left(\mathbf{m}\right)\right]^T\left[\mathbf{y}-\mathbf{f}\left(\mathbf{m}\right)\right] \\ &+\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T\mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) \\ &+\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T\mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \end{aligned} \qquad \textbf{(2.91)}$$

In Eq. 2.91, the constant in each term is not required for the minimization process, so the equation is written again as

$$\begin{aligned} O(\mathbf{m},\boldsymbol{\theta}) &= \left[\mathbf{y}-\mathbf{f}\left(\mathbf{m}\right)\right]^T\left[\mathbf{y}-\mathbf{f}\left(\mathbf{m}\right)\right] \\ &+\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T\mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) \\ &+\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T\mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \end{aligned} \qquad \textbf{(2.92)}$$

Redefined, as is a general form of objective function for least-squares estimation method. For simplicity, it is assumed that the *M*-dimensional vector $\mathbf{m}$ contains ln*k* values over a uniformly spaced block centered mesh (M= $N_x \times N_y$, where N_x represent the number of grid blocks in the *x*-direction and N_y represent the number of grid blocks in the *y*-direction). In Eq. 2.92, $\mathbf{f}$ is simply equal to $\mathbf{Gm}$, $\mathrm{f}_{\mathrm{model}}\left(\mathrm{m}\right) = \mathrm{Gm}$ and $\mathbf{G}$ is the N×M sensitivity matrix. The covariance matrix $\mathbf{C_M}$ is *M*×*M* dimensional and off-diagonal to be computed from given semivariogram. It is also assumed that the correction vector covariance matrix $\mathbf{C}_{\boldsymbol{\theta}}$ is a M×M diagonal matrix. It should be noted that Eq. 2.109 is given in a general formulation that the mean of each attribute varies from gridblock to gridblock and hence the vectors of $\mathbf{m}_{pr}$, $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0$ are *M*-dimensional.

Taking the derivatives of the objective function given by Eq. 2.92 with respect to $\mathbf{m}$ and $\boldsymbol{\theta}$, and equating them to zero gives, respectively,

$$\nabla_{\mathbf{m}}\mathbf{O}(\mathbf{m},\boldsymbol{\theta}) = \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)-\mathbf{G}^T\left[\mathbf{y}-\mathbf{Gm}\right]=0 \qquad \textbf{(2.93)}$$

and

$$\nabla_{\boldsymbol{\theta}}\mathbf{O}(\mathbf{m},\boldsymbol{\theta}) = -\mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)+\mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)=0 \quad \textbf{(2.94)}$$

It is needed to simultaneously solve for $\mathbf{m}$ and $\boldsymbol{\theta}$ to obtain the posterior estimates $\tilde{\mathbf{m}}$ and $\tilde{\boldsymbol{\theta}}$ and it can be shown that the solution of Eqs. 2.93 and 2.94 is given by

$$\begin{bmatrix} \mathbf{G}^T\mathbf{G} + \mathbf{C}_M^{-1} & -\mathbf{C}_M^{-1} \\ -\mathbf{C}_M^{-1} & \mathbf{C}_M^{-1}+\mathbf{C}_\theta^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{m}} \\ \tilde{\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_M^{-1}\mathbf{m}_{pr} + \mathbf{G}^T\mathbf{y} \\ -\mathbf{C}_M^{-1}\mathbf{m}_{\mathrm{pr}} + \mathbf{C}_\theta^{-1}\boldsymbol{\theta}_0 \end{bmatrix} \quad \textbf{(2.95)}$$

It is obviously seen, solving Eq. 2.95 is a matrix problem. In this thesis, for solving matrix equations such as equation above, LU decomposition method is used. It may be noted that Eq. 2.95 can be solved by LU decomposition for small sized problems, but will require that we need to use iterative or sparse matrix techniques for efficient storage and solution for large size problems (Onur, 2009).

2.4 Minimization of the Objective Functions

In general, to minimize the objective functions derived previously for the maximum likelihood and least-squares estimations, we can use a gradient-based method such as Newton, Gauss Newton or Levenberg Marquardt. In this work, Levenberg–Marquardt algorithm is used to minimize the objective functions.

The usage of gradient algorithms is efficient however; it is required to calculate sensitivity matrix and depending on the cases, gradient of the objective function to compute the sensitivities. In general, the capability of automatic history matching leans on the parameterization of the model and the efficiency of the optimization algorithm. The computational efficiency of the optimization process depends on the number of iterations and the cost per iteration and most commonly, the cost per iteration depends of computing the required sensitivity coefficients when gradient-based algorithms are used (Onur, 2009). In this work, it is used to Levenberg-Marquardt algorithm in order to minimize objective function. In addition, LU decomposition is chosen to solve the matrix equations.

2.4.1 Levenberg-Marquardt method

The Levenberg–Marquardt method is a hybrid method based on the gradient descent (steepest-descent) method and the Gauss–Newton method. In Steepest–descent method, the parameters are updated with the direction of the enormous reduction of the objective function and so, the sum of squared errors is diminished by this updating procedure. In the Gauss–Newton method, the least - squares function is assumed locally quadratic and at the minimum of the quadratic is aimed in order to diminish the sum of the squared errors. The Levenberg–Marquardt (L-M) method behaves like a steepest–descent method when the model variables are far from their finest value and behaves like the Gauss–Newton method when the parameters approach to their best value (Marquardt, 1963; Gavin, 2010).

The idea of L-M method is to modify the direction and the steps length simultaneously. Thus, the method involves solving at each iteration the following system (Marquardt, 1963; Onur, 2009):

$$\left[\mathbf{H}(\hat{\mathbf{m}}_k)+\lambda_k\mathbf{I}\right]\delta\hat{\mathbf{m}}_{k+1}=-\nabla O(\hat{\mathbf{m}}_k) \tag{2.96}$$

where λ_k is a positive constant, $\mathbf{I}$ is $2M\times 2M$ identity matrix, $\delta\hat{\mathbf{m}}_{k+1}$ is the search direction for updating model parameter and it is stated as

$$\delta\hat{\mathbf{m}}_{k+1}=\left(\hat{\mathbf{m}}_{k+1}-\hat{\mathbf{m}}_k\right) \tag{2.97}$$

and $\nabla O(\hat{\mathbf{m}}_k)$ is M-dimensional gradient vector, elements are the first derivation of $O(\mathbf{m},\boldsymbol{\theta})$ with respect to $\mathbf{m}$ and $\boldsymbol{\theta}$. It is defined as

$$\nabla O(\hat{\mathbf{m}}_k)=\begin{bmatrix}\nabla O(\mathbf{m}_k)\\ \nabla O(\boldsymbol{\theta}_k)\end{bmatrix} \tag{2.98}$$

and the transpose of gradient is as

$$\nabla O(\hat{\mathbf{m}}_k)=\begin{bmatrix}\dfrac{\delta O}{\delta m_1} & \dfrac{\delta O}{\delta m_2} & \cdots & \dfrac{\delta O}{\delta m_M} & \dfrac{\delta O}{\delta \theta_1} & \dfrac{\delta O}{\delta \theta_2} & \cdots & \dfrac{\delta O}{\delta \theta_M}\end{bmatrix}^T \tag{2.99}$$

In Eq. 2.96, $\mathbf{H}(\hat{\mathbf{m}}_k)$ represents Hessian matrix which is a $2M\times 2M$ symmetric matrix and includes the second derivatives of $O(\mathbf{m},\boldsymbol{\theta})$ with respect to $\mathbf{m}$ and $\boldsymbol{\theta}$. The general form is defined by

$$\mathbf{H}(\hat{\mathbf{m}})=\begin{bmatrix}\nabla_{\mathbf{m}}\left[\left(\nabla_{\mathbf{m}}O(\hat{\mathbf{m}})\right)^T\right] & \nabla_{\mathbf{m}}\left[\left(\nabla_{\boldsymbol{\theta}}O(\hat{\mathbf{m}})\right)^T\right]\\ \nabla_{\boldsymbol{\theta}}\left[\left(\nabla_{\mathbf{m}}O(\hat{\mathbf{m}})\right)^T\right] & \nabla_{\boldsymbol{\theta}}\left[\left(\nabla_{\boldsymbol{\theta}}O(\hat{\mathbf{m}})\right)^T\right]\end{bmatrix} \tag{2.100}$$

and expansion of the Hessian, according to the problem is stated by

$$\mathbf{H}(\mathbf{m},\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 O}{\partial m_1^{\ 2}} & \cdots & \frac{\partial^2 O}{\partial m_1 \partial m_M} & \frac{\partial^2 O}{\partial m_1 \partial \theta_1} & \cdots & \frac{\partial^2 O}{\partial m_1 \partial \theta_M} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 O}{\partial m_M \partial m_1} & \cdots & \frac{\partial^2 O}{\partial m_M^{\ 2}} & \frac{\partial^2 O}{\partial m_M \partial \theta_1} & \cdots & \frac{\partial^2 O}{\partial m_M \partial \theta_M} \\ \frac{\partial^2 O}{\partial \theta_1 \partial m_1} & \cdots & \frac{\partial^2 O}{\partial \theta_1 \partial m_M} & \frac{\partial^2 O}{\partial \theta_1^{\ 2}} & \cdots & \frac{\partial^2 O}{\partial \theta_1 \partial \theta_M} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 O}{\partial \theta_M \partial m_1} & \cdots & \frac{\partial^2 O}{\partial \theta_M \partial m_M} & \frac{\partial^2 O}{\partial \theta_M \partial \theta_1} & \cdots & \frac{\partial^2 O}{\partial \theta_M^{\ 2}} \end{bmatrix} \quad \textbf{(2.101)}$$

The basic L-M algorithm for minimizing an arbitrary objective function is given in Eqs. 2.96 and 2.97. The iterative procedure is given by Fletcher (1991) and Oliver et al. (2008) as

- **Step 1:** Set k = 0, $\lambda_0 = 0.001$ (it is referred to give a small value to λ), give an initial value to the model parameter and the correction $\hat{m}_0$.
- **Step 2:** Set k = k+1, compute $O(\hat{\mathbf{m}}_k)$, $\nabla O(\hat{\mathbf{m}}_k)$ and $\mathbf{H}(\hat{\mathbf{m}}_k)$
- **Step 3:** Compute the correction vector $\delta \hat{\mathrm{m}}_{\mathrm{k+1}}$ given in Eq. 2.10 by using LU decomposition method. Calculate Eq. 2.11 and $O(\hat{\mathbf{m}}_{k+1})$.
- **Step 4:** If $O(\hat{\mathbf{m}}_c^{k+1}) < O(\hat{\mathbf{m}}^k)$, if it is satisfied, then set $\hat{\mathbf{m}}^k = \hat{\mathbf{m}}_c^{k+1}, \lambda_{k+1} = \lambda_k / a_1$, where $a_1 > 1$, and go to the next iteration. Otherwise, increase λ_k by some factor a_2, where $a_2 > 1, \lambda_{k+1} = \lambda_k a_2$ and repeat the iteration. Choosing $a_1 = a_2 \approx 10$ is common, but optimal choice for these parameters is difficult to determine *a priori*.
- **Step 5:** Check your stopping criteria (In this thesis, the stopping criteria is defined as $abs(O(\hat{\mathbf{m}}_c^{k+1}) - O(\hat{\mathbf{m}}^k)) < 10^{-10}$), if the criteria is satisfied, then stop iterating. Otherwise, go to **Step 2** to start a new iteration level.

In L-M iterative method, it is needed to solve Eq. 2.10 by LU decomposition (factorization) to find the values of model parameters $\hat{\mathbf{m}}$.

LU decomposition: To solve a general non–singular (have an inverse) linear system like $\mathbf{Ax} = \mathbf{b}$, a direct technique is to devise an equivalent linear system in which the matrix is triangular (Gill et al., 1981). As it is stated by Press et al. (1992), Suppose that matrix $\mathbf{A}$ is able to be written as a product of two triangular matrices,

$$\mathbf{A} = \mathbf{LU} \qquad \textbf{(2.102)}$$

where $\mathbf{L}$ is lower triangular and $\mathbf{U}$ is upper triangular matrix. The decomposition of given linear equation can be made as

$$\mathbf{Ax} = (\mathbf{LU})\mathbf{x} = \mathbf{L}(\mathbf{Ux}) = \mathbf{b} \qquad \textbf{(2.103)}$$

To solve this linear set, first the vector $\mathbf{y}$ must be solved such that

$$\mathbf{Ly} = \mathbf{b} \qquad \textbf{(2.104)}$$

and when vector $\mathbf{y}$ is solved, then vector $\mathbf{x}$ can be solved by

$$\mathbf{Ux} = \mathbf{y} \qquad \textbf{(2.105)}$$

The advantage of separating the original equation in two parts is that the solution of a triangular set of equation is very simple. Thus, Eq. 2.18 can be solved by forward substitution and Eq. 2.19 can be solved by back substitution (Press et al., 1981).

In following sub-chapters, the objective functions are derivated for various minimization methods such as weighted least-squares (WLS), unweighted least-squares (UWLS) and maximum likelihood (ML) estimation methods considering partially doubly stochastic model. Typically, this will be the case when it is considered geostatistical models for rock properties such as permeability and porosity. For this problem, it is convenient to write the objective function in terms vectors and matrices. Suppose that it is tried to estimate log-permeability (ln*k*) assuming that ln*k* is a Gaussian multivariate with a given mean and spatial covariance matrix over a discretizated reservoir mesh either for a 1D or 2D problem.

2.4.2 A Synthetic example for maximum likelihood and least-squares estimation methods

In this chapter, we demonstrate the application of the partially doubly stochastic model for both ML and LS estimation methods for a linear model. A two-parameter linear problem is considered and it is assumed that observed data consists of three observed data sets as a function of time. The function of the linear model given by

$$f(m,t) = m_1 + m_2 \ln t \qquad \textbf{(2.106)}$$

According to this model, each data set have $N_{d_1} = 15$, $N_{d_2} = 10$ and $N_{d_3} = 15$ observations and for each data set, the true values of variances are specified as $\sigma_1^2 = 4$, $\sigma_2^2 = 0.25$ and $\sigma_3^2 = 0.0001$. So, this is an overdetermined problem.

It will be assumed that all matrices $\mathbf{C}_M{}^{-1}$, and $\mathbf{C}_\theta{}^{-1}$ are diagonal matrices; i.e.,

$$\mathbf{C}_M^{-1} = \begin{bmatrix} \frac{1}{\sigma_{m_1}^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_{m_2}^2} & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{\sigma_{m_M}^2} \end{bmatrix} \tag{2.107}$$

and

$$\mathbf{C}_\theta^{-1} = \begin{bmatrix} \frac{1}{\sigma_{\theta_1}^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_{\theta_2}^2} & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{\sigma_{\theta_M}^2} \end{bmatrix} \tag{2.108}$$

The observed data and observation time are given in Table 3.1. The true values of model parameters are $m_1 = 10$ and $m_2 = 5$. A code in Fortran90 programming language has been developed for solving the objective functions of these models. As mentioned before, to minimizing these objective functions, Levenberg-Marquardt algorithm is used. The value of stopping criteria or tolerance used is equal to 10^{-10}. Besides, for LU decomposition method, the subroutines are provided from Numerical Recipes (Press et al., 1992). The observed data and observation time are given Table 2.11.

Table 2.11: $\mathbf{y}$ and **t** values obtained from measurements.

$\mathbf{t}_1$	$\mathbf{y}_1$	$\mathbf{t}_2$	$\mathbf{y}_2$	$\mathbf{t}_3$	$\mathbf{y}_3$
0.2	2.84903	3.2	15.26713	5.2	18.24777
0.4	4.84396	3.4	16.23261	5.4	18.42912
0.6	8.20325	3.6	15.86329	5.6	18.61761
0.8	8.14101	3.8	16.64221	5.8	18.78557
1	8.12388	4	17.04152	6	18.94941
1.2	8.94669	4.2	16.95869	6.2	19.11292
1.4	11.91136	4.4	18.02271	6.4	19.28263
1.6	13.95758	4.6	17.67031	6.6	19.44338
1.8	9.93653	4.8	17.7389	6.8	19.5696
2	11.47588	5	19.18386	7	19.7196
2.2	15.30103			7.2	19.87719
2.4	17.17171			7.4	20.02137
2.6	14.38451			7.6	20.13877
2.8	17.73598			7.8	20.28355
3	19.45149			8	20.41699

2.5 Maximum Likelihood Estimation with Prior Information

The generalized objective function for MLE is in Eq. 2.22 as

$$O(\mathbf{m},\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=1}^{K} N_{dj} \ln\left\{\left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]^T \left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]\right\} + \frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) + \frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \quad \textbf{(2.109)}$$

As it is seen in Eq. 2.109, although the problem is linear, the objective function to be minimized is a nonlinear function. It is important to note it is considered a general formulation that each observed data vector, $y_{obs,j}$ may contain a different total number of observed data, N_{dj}, $j = 1, 2, \ldots, K$ and for this example K=3. As the model parameter $\mathbf{m}$ is a *M-dimensional* vector and hence, $\mathbf{m}_{pr}$, $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0$ are also *M-dimensional* vectors. In addition, the diagonal matrix of prior $\mathbf{C}_\mathbf{M}^{-1}$ and the diagonal matrix of correction $\mathbf{C}_{\boldsymbol{\theta}}^{-1}$ are both $M\times M$ *dimension* matrices and the number of model parameters are defined as M = 2.

The variances of data is calculated by given equation

$$\tilde{\sigma}_j^2(m) = \frac{1}{N_{dj}}\sum_{i=1}^{N_{dj}}\left[y_{i,j} - f(m,t_{i,j})\right]^2 \quad \textbf{(2.110)}$$

However, the variance calculated from Eq. 2.110 is biased, it must be used Eq. 2.111 to perform unbiased variance as

$$\hat{\sigma}_j^2 = \frac{N_{dj}\tilde{\sigma}_j^2(m)}{\left(N_{dj} - \frac{M}{K}\right)} \quad \textbf{(2.111)}$$

The program is based on observed data and observation time given in Table 2.11. On the other hand, the program converges the results iteratively. Based on this, the value of model parameters and the corrections are initially defined in columns 2-5 according to Prior Models in column 1. The values of $\mathbf{m}_{pr}$, $\boldsymbol{\theta}_0$ and the value of variances $\sigma_{m,i}^2$ and $\sigma_{\theta,i}^2$ are also given for each case. The inputs and the outputs are given Tables 2.12-2.13

Table 2.12: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 1	10	5	0	0	10	5	0	0	10000	10000	10000	10000
Prior Model 2	10	5	0	0	0	0	0	0	10000	10000	10000	10000
Prior Model 3	10	5	0	0	8	3	0	0	10000	10000	10000	10000
Prior Model 4	10	5	0	0	1	15	0	0	10000	10000	10000	10000

Table 2.13: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Prior Models	m_1	m_2	θ_1	θ_2	σ_1^2	σ_2^2	σ_3^2	Obj_i	Obj_f
True values	10	5	-	-	4	0.25	0.0001	-	-
Prior Model 1	9.9453	5.0298	-0.027	0.015	4.0108	0.2507	0.00008	-14.45	-15.93
Prior Model 2	9.9453	5.0298	4.973	2.515	4.0108	0.2507	0.00008	-14.44	-15.93
Prior Model 3	9.9453	5.0298	0.973	1.015	4.0108	0.2507	0.00008	-14.45	-15.93
Prior Model 4	9.9453	5.0298	4.473	-4.985	4.0108	0.2507	0.00008	-14.44	-15.93

According to Table 2.13, the variances of each data sets are computed almost same as the true values. Besides, the model parameters are very close to real ones even if the prior is very wrong (Prior Model 4). Thus, the table shows that the correction is working when variances of prior and correction are assumed very high. In addition, objective function is minimized.

Table 2.14: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta1}^2$	$\sigma_{\theta2}^2$
Prior Model 5	10	5	0	0	1	15	0	0	0.1	0.1	0.001	0.001
Prior Model 6	10	5	0	0	1	15	0	0	10	10	0.001	0.001
Prior Model 7	10	5	0	0	1	15	0	0	0.1	0.1	1	1
Prior Model 8	10	5	0	0	1	15	0	0	1	1	1	1
Prior Model 9	10	5	0	0	1	15	0	0	100	100	1	1
Prior Model 10	10	5	0	0	1	15	0	0	0.1	0.1	100	100
Prior Model 11	10	5	0	0	1	15	0	0	1	1	100	100
Prior Model 12	10	5	0	0	1	15	0	0	100	100	100	100

It is obviously seen in Tables 2.14 and 2.15 that despite the prior mean is wrong, if the prior variance is assumed to be small, the values of model parameters goes to wrong means. However, if variances of prior and correction is chosen bigger the values of model parameters goes to true values as it is expected, hence the number of iteration is higher (4-5). In addition, the value of objective function is minimized while this method is performing and the corrections are useful to find the true values. Besides, the data variances are very close to real ones.

Table 2.15: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Prior Models	m_1	m_2	θ_1	θ_2	σ_1^2	σ_2^2	σ_3^2	Obj_i	Obj_f
Prior Model 5	0.664	14.222	-0.003	-0.008	103.04	14.5434	68.290	890.5	134.4
Prior Model 6	9.943	5.030	0.001	-0.001	4.0108	0.2507	0.00018	-5.4	-6.96
Prior Model 7	9.931	5.037	8.12	-9.057	4.0108	0.2507	0.00008	890.5	65.54
Prior Model 8	9.937	5.033	4.469	-4.983	4.0108	0.2507	0.00008	76.05	28.9
Prior Model 9	9.945	5.029	0.089	-0.099	4.0108	0.2507	0.00008	-13.54	-15.05
Prior Model 10	9.945	5.029	8.936	-9.96	4.010	0.2507	0.00008	890.5	-15.04
Prior Model 11	9.945	5.029	8.857	-9.871	4.010	0.2507	0.00008	76.05	-15.05
Prior Model 12	9.945	5.029	4.473	-4.985	4.0108	0.2507	0.00008	-13.54	-15.48

For these cases, the initial values of model parameters are also chosen wrong same as prior model. As the results of program is not corresponding any different values, the outputs are not given for this problem.

In the case of the value of prior mean contains less error, then doubly stochastic model is effectively work, even if the variances of mean and correction are smaller. Still, if the value of mean and correction variances is chosen bigger, again, the results are close the real ones for each model parameter and variances. In addition, if the variance of mean is chosen incorrect but small enough, then the iteration number decreases and objective function is minimized.

Table 2.16: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 13	10	5	0	0	8	3	0	0	0.1	0.1	0.001	0.001
Prior Model 14	10	5	0	0	8	3	0	0	10	10	0.001	0.001
Prior Model 15	10	5	0	0	8	3	0	0	0.1	0.1	1	1
Prior Model 16	10	5	0	0	8	3	0	0	10	10	1	1
Prior Model 17	10	5	0	0	8	3	0	0	0.1	0.1	100	100

Table 2.17: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Prior Models	m_1	m_2	θ_1	θ_2	σ_1^2	σ_2^2	σ_3^2	Obj_i	Obj_f
Prior Model 13	9.9364	5.0345	0.019	0.02	4.011	0.251	8E-05	25.55	23.16
Prior Model 14	9.9452	5.0298	0	0	4.011	0.251	8E-05	-14.05	-15.54
Prior Model 15	9.9445	5.0302	1.768	1.846	4.011	0.251	8E-05	25.55	-12.34
Prior Model 16	9.9453	5.0298	0.177	0.185	4.011	0.251	8E-05	-14.05	-15.57
Prior Model 17	9.9453	5.0298	1.943	2.028	4.011	0.251	8E-05	25.55	-15.89

In the case of the value of prior mean contains less error, then doubly stochastic model is effectively work, even if the variances of mean and correction are smaller. Still, if the value of mean and correction variances is chosen bigger, again, the results are close the real ones for each model parameter and variances. In addition, if the variance of mean is chosen incorrect but small enough, then the iteration number decreases and objective function is minimized.

2.6 Weighted Least-Squares Parameter Estimation With Prior Information

The generalized objective function for WLSE is in Eq. 2.52 as

$$O(\mathbf{m},\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=1}^{K}\left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right]^T \mathbf{C}_{Dj}^{-1}\left[\mathbf{y}_j - \mathbf{f}_j(\mathbf{m})\right] + \frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) + \frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \quad \textbf{(2.112)}$$

It is important to note it is considered a general formulation that each observed data vector, $\mathbf{y_j}$ may contain a different total number of observed data, N_{dj}, $j = 1,2,\ldots,K$ and for this example K=3. The variances of each data are assumed to be true and their values are, respectively, $\sigma_1^2 = 4$, $\sigma_2^2 = 0.25$ and $\sigma_3^2 = 0.0001$. As the model parameter $\mathbf{m}$ is a *M-dimensional* vector and hence, $\mathbf{m}_{pr}$, $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0$ are also *M-dimensional* vectors. In addition, the diagonal matrix of prior $\mathbf{C_M^{-1}}$ and the diagonal matrix of correction $\mathbf{C}_{\boldsymbol{\theta}}^{-1}$ are both $M \times M$ *dimension* matrices and the number of model parameters are defined as $M = 2$.

The program is based on observed data and observation time given in Table 2.11. On the other hand, the program converges the results iteratively. Based on this, the value of model parameters and the corrections are initially defined in columns 2-5 according to Prior Models in column 1. The values of $\mathbf{m}_{pr}$, $\boldsymbol{\theta}_0$ and the value of variances $\sigma_{m,i}^2$ and $\sigma_{\theta,i}^2$ are also given for each case. The inputs and the outputs are given Tables 2.18-2.23.

Table 2.18: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 1	10	5	0	0	10	5	0	0	10000	10000	10000	10000
Prior Model 2	10	5	0	0	0	0	0	0	10000	10000	10000	10000
Prior Model 3	10	5	0	0	8	3	0	0	10000	10000	10000	10000
Prior Model 4	10	5	0	0	1	15	0	0	10000	10000	10000	10000

Table 2.19: The results are obtained from program, respectively, model parameters, corrections, and minimized objective functions.

Prior Models	m_1	m_2	θ_1	θ_2	Obj_f	Obj_f
True values	10	5	-	-	-	-
Prior Model 1	9.9454	5.0298	-0.027	0.015	19.052	17.763
Prior Model 2	9.9454	5.0298	4.973	2.515	19.058	17.766
Prior Model 3	9.9454	5.0298	0.973	1.015	19.052	17.763
Prior Model 4	9.9454	5.0298	4.473	-4.985	19.061	17.767

As can be seen in Tables 2.18 and 2.19 that the results obtained from program are close to the real values of parameters though the prior means are assumed to be incorrect. The values of model parameters are computed with approximately $\pm 10^{-2}$ correction.

Table 2.20 and 2.21 show that with very incorrect prior mean, if the engineer trusts his/her data and takes the prior variances smaller, the value of the model parameters goes to mean of data. Yet, if the prior variances are assumed higher than the true ones, then the estimates of the model parameters go to the true values.

Table 2.20: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 5	10	5	0	0	1	15	0	0	0.1	0.1	0.001	0.001
Prior Model 6	10	5	0	0	1	15	0	0	10	10	0.001	0.001
Prior Model 7	10	5	0	0	1	15	0	0	0.1	0.1	1	1
Prior Model 8	10	5	0	0	1	15	0	0	1	1	1	1
Prior Model 9	10	5	0	0	1	15	0	0	100	100	1	1
Prior Model 10	10	5	0	0	1	15	0	0	0.1	0.1	100	100
Prior Model 11	10	5	0	0	1	15	0	0	1	1	100	100
Prior Model 12	10	5	0	0	1	15	0	0	100	100	100	100

Table 2.21: The results are obtained from program, respectively, model parameters, corrections, and minimized objective functions.

Prior Models	m_1	m_2	θ_1	θ_2	Obj_i	Obj_f
Prior Model 5	9.7605	5.1279	0.087	-0.098	924.052	892.976
Prior Model 6	9.9435	5.0308	0.001	-0.001	28.102	26.732
Prior Model 7	9.9281	5.0389	8.116	-9.056	924.052	99.208
Prior Model 8	9.9359	5.0348	4.468	-4.983	109.552	62.585
Prior Model 9	9.9452	5.0299	0.089	-0.099	19.957	18.651
Prior Model 10	9.9452	5.0299	8.936	-9.96	924.052	18.659
Prior Model 11	9.9452	5.0299	8.857	-9.871	109.552	18.651
Prior Model 12	9.9453	5.0298	4.473	-4.985	19.957	18.211

Table 2.22: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 13	10	5	0	0	8	3	0	0	0.1	0.1	0.001	0.001
Prior Model 14	10	5	0	0	8	3	0	0	10	10	0.001	0.001
Prior Model 16	10	5	0	0	8	3	0	0	0.1	0.1	1	1
Prior Model 17	10	5	0	0	8	3	0	0	10	10	1	1
Prior Model 19	10	5	0	0	8	3	0	0	0.1	0.1	100	100

Table 2.23: The results are obtained from program, respectively, model parameters, corrections, and minimized objective functions.

Prior Models	m_1	m_2	θ_1	θ_2	Obj_i	Obj_f
Prior Model 13	9.934	5.0357	0.019	0.02	59.052	56.844
Prior Model 14	9.9452	5.0298	0	0	19.452	18.158
Prior Model 16	9.9443	5.0303	1.768	1.846	59.052	21.355
Prior Model 17	9.9452	5.0298	0.177	0.185	19.452	18.122
Prior Model 19	9.9453	5.0298	1.943	2.028	59.052	17.802

Tables 2.22 - 23 show that the error in the prior means can be taken very wrong if the variances of priors and corrections are assumed high enough. In results of these, it can be said that the correction of prior means are useful to reduce the error on prior means.

2.7 Unweighted Least-Squares Parameter Estimation With Prior Information

The generalized objective function for UWLS is in Eq. 2.113 as

$$O(\mathbf{m},\boldsymbol{\theta}) = \left[\mathbf{y}-\mathbf{f}(\mathbf{m})\right]^T \left[\mathbf{y}-\mathbf{f}(\mathbf{m})\right] + \left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) + \left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \quad \textbf{(2.113)}$$

It is important to note it is considered a general formulation that each observed data vector, $\mathbf{y}$ may contain a different total number of observed data, $N_{dj},\ j=1,2,\ldots,K$ and for this example K=3. As the model parameter $\mathbf{m}$ is a *M-dimensional* vector and hence, $\mathbf{m}_{pr}$, $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0$ are also *M-dimensional* vectors. In addition, the diagonal matrix of prior $\mathbf{C}_\mathbf{M}^{\mathbf{-1}}$ and the diagonal matrix of correction $\mathbf{C}_{\boldsymbol{\theta}}^{\mathbf{-1}}$ are both $M\times M$ *dimension* matrices and the number of model parameters are defined as M = 2.

The program is based on observed data and observation time given in Table 2.11. On the other hand, the program converges the results iteratively. Based on this, the value of model parameters and the corrections are initially defined in columns 2-5 according to prior models in column 1. In addition of these, we assumed the initial values of $\mathbf{m}_{pr}$ and $\boldsymbol{\theta}_0$, and the variances of corrections and the variances of prior means. The values of $\mathbf{m}_{pr}$, $\boldsymbol{\theta}_0$ and the value of variances $\sigma^2_{m,i}$ and $\sigma^2_{\theta,i}$ are also given for each case. The inputs and the outputs are given in Tables 2.24-2.29.

Table 2.24: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta1}^2$	$\sigma_{\theta2}^2$
Prior Model 1	10	5	0	0	10	5	0	0	10000	10000	10000	10000
Prior Model 2	10	5	0	0	0	0	0	0	10000	10000	10000	10000
Prior Model 3	10	5	0	0	8	3	0	0	10000	10000	10000	10000
Prior Model 4	10	5	0	0	1	15	0	0	10000	10000	10000	10000

Table 2.25: The results are obtained from program, respectively, model parameters, corrections, and minimized objective functions.

Prior Models	m_1	m_2	θ_1	θ_2	Obj_i	Obj_f
True values	10	5	-	-		
Prior Model 1	9.9989	5.093	-0.001	0.047	29.889	29.535
Prior Model 2	9.9989	5.093	4.999	2.547	29.895	29.539
Prior Model 3	9.9989	5.093	0.999	1.047	29.889	29.536
Prior Model 4	9.9989	5.093	4.499	-4.953	29.898	29.54

In this case, the effect of correction is not seen when the values of prior means are equal or close to the values of real model parameters. Just, when the values of prior means are taken as incorrect, the effects of corrections are seen more clearly. The outputs of model parameters are same with true values because of uncorrelated data. The outputs are similar to results without using prior (Kuchuk et al., 2010).

Table 2.26: The results are obtained from program, respectively, model parameters, corrections, and minimized objective functions.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 5	10	5	0	0	1	15	0	0	0.1	0.1	0.001	0.001
Prior Model 6	10	5	0	0	1	15	0	0	10	10	0.001	0.001
Prior Model 7	10	5	0	0	1	15	0	0	0.1	0.1	1	1
Prior Model 8	10	5	0	0	1	15	0	0	1	1	1	1
Prior Model 9	10	5	0	0	1	15	0	0	100	100	1	1
Prior Model 10	10	5	0	0	1	15	0	0	0.1	0.1	100	100
Prior Model 11	10	5	0	0	1	15	0	0	1	1	100	100
Prior Model 12	10	5	0	0	1	15	0	0	100	100	100	100

Table 2.27: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_1	m_2	θ_1	θ_2	Obj_i	Obj_f
Prior Model 5	4.9048	8.6868	0.039	-0.063	934.889	513.114
Prior Model 6	9.8993	5.1605	0.001	-0.001	38.939	38.413
Prior Model 7	9.1571	5.6647	7.416	-8.487	934.889	104.94
Prior Model 8	9.5186	5.4186	4.259	-4.791	120.389	72.431
Prior Model 9	9.989	5.0998	0.089	-0.098	30.794	30.421
Prior Model 10	9.9889	5.0999	8.98	-9.89	934.889	30.429
Prior Model 11	9.989	5.0998	8.9	-9.802	120.389	30.421
Prior Model 12	9.9939	5.0965	4.497	-4.952	30.794	29.983

Again, the outputs shows that if prior means are so wrong, then it must be used bigger variances of priors and corrections. Thus, the result are not affected and as the values of model parameters are almost true, the decisions of prediction performance can be made with very little errors.

Table 2.28: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Prior Models	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta1}^2$	$\sigma_{\theta2}^2$
Prior Model 13	10	5	0	0	8	3	0	0	0.1	0.1	0.001	0.001
Prior Model 14	10	5	0	0	8	3	0	0	10	10	0.001	0.001
Prior Model 15	10	5	0	0	8	3	0	0	0.1	0.1	1	1
Prior Model 16	10	5	0	0	8	3	0	0	10	10	1	1
Prior Model 17	10	5	0	0	8	3	0	0	0.1	0.1	100	100

Table 2.29: The results are obtained from program, respectively, model parameters, corrections, and minimized objective functions.

Prior Models	m_1	m_2	θ_1	θ_2	Obj_i	Obj_f
Prior Model 13	9.6483	5.0432	0.016	0.02	69.889	67.018
Prior Model 14	9.9932	5.0938	0	0	30.289	29.954
Prior Model 15	9.9497	5.0974	1.772	1.907	69.889	33.302
Prior Model 16	9.9937	5.0937	0.181	0.19	30.289	29.916
Prior Model 17	9.9984	5.0932	1.996	2.091	69.889	29.577

In this case, because error on prior means are smaller than Table 2.26 - 27, the results are included less errors. The model parameters are obtained almost ture when the variances of prior and corrections are high by minimizing objective functions. Hence, the final values of objective functions are half of initial values.

3. APPLICATIONS OF UNDERDETERMINED PROBLEMS

Here, we explore how the estimates of the unknown parameters to be obtained by minimizing the WLS objective model considering partially doubly stochastic model where the prior mean is uncertain for an underdetermined problem. Here, the underdetermined problem refers to a case where the number of model parameter exceeds the number of observed data. Typically, this will be the case when a geostatistical model for rock properties such as permeability and porosity are considered.

For this problem, it is convenient to write the objective function in terms vectors and matrices. Suppose that we wish to estimate log-permeability (ln*k*) assuming that ln*k* is a Gaussian multivariate with a given mean and spatial covariance matrix over a discretized reservoir mesh either for a 1D, 2D, or 3D problem. For simplicity, we will consider a 1D problem, though the same methodology and results will apply for more general cases of 2D and 3D problems.

In this case, for a partially doubly stochastic model, it will be minimized the following objective function for N observed data points

$$O(\mathbf{m},\boldsymbol{\theta}) = \frac{1}{2}\left[\mathbf{y}-\mathbf{f}\left(\mathbf{m}\right)\right]^T \mathbf{C}_D^{-1}\left[\mathbf{y}-\mathbf{f}\left(\mathbf{m}\right)\right] + \frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right) + \frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \qquad \textbf{(3.1)}$$

For simplicity, it will be assume that the M-dimensional vector **m** contains ln*k* values over a uniformly spaced block centered mesh (M = $N_x \times N_y$, where N_x represent the number of grid blocks in the *x*-direction and N_y represent the number of grid blocks in the *y*-direction). In Eq. 3.1, **f** is simply equal to **Gm**, $\mathbf{f}\left(\mathbf{m}\right)=\mathbf{Gm}$ and **G** is the $N\times M$ sensitivity matrix to be defined later. Further suppose that $\boldsymbol{y}_{\text{obs}}$ contains N observed (measured) value of lnk at some spatial points with noise having zero mean and a standard deviation equal to σ_d. So, the data covariance matrix $\mathbf{C_D}$ in Eq. 3.1 is an N×N diagonal matrix, whereas the $M \times M$ covariance matrix $\mathbf{C_M}$ is nondiagonal covariance matrix to be computed from given semi-variogram. We will also assume that the covariance matrix of correction vector $\mathbf{C}_\theta$ is an $M \times M$ diagonal matrix.

It should be noted that Eq. 3.1 is given in a general formulation that the mean of each attribute [ln*k* and porosity (ϕ) for example] varies from gridblock to gridblock and hence the vectors of $\mathbf{m}_{pr}$, $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0$ are M-dimensional. However, if the objective is to assume that the mean of each attribute is identical, but could assume different values for each attribute, for all gridblocks, then it will be needed to consider the same scalars for m_{pr}, θ and θ_0 for all gridblocks for a given attribute. This is equivalent to making correction on the global mean of the given attribute. To handle such cases, we define a N_e-dimensional column vector denoted by **e** with all components equal to unity, i.e.,

$$\mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \textbf{(3.2)}$$

where $N_a N_e = N_p$, where N_a is the number of attributes and parameters, and N_p is the total number of model parameters. For example, if it is considered both ln*k* and porosity are as unknowns in each gridblock, then $N_a = 2$, $N_p = 2M$, and $N_e = M$.

In this work, it will be considered only a single attribute, which is ln*k*, then $N_e = N_p = M$. Then, the prior mean $\mathbf{m}_{pr}$ in Eq. 3.1 can be written in a more general form

$$\mathbf{m}_{pr} = m_{pr}\mathbf{e} \quad \textbf{(3.3)}$$

Where m_p is a scalar representing the prior mean of ln*k* for all gridblocks. In this case, it must be required that the correction vector $\boldsymbol{\theta}$ and its mean $\boldsymbol{\theta}_0$ have the same structure as $\mathbf{m}_{pr}$, i.e., it is required

$$\boldsymbol{\theta} = \theta\mathbf{e} \quad \textbf{(3.4)}$$

and

$$\boldsymbol{\theta}_0 = \theta_0\mathbf{e} \quad \textbf{(3.5)}$$

where θ and θ_0 are scalars having the same values for all gridblocks. Although it is appropriate to choose $\theta_0 = 0$, the derivation is done for any value of θ_0.

Next, it will be defined the $(N_aN_e) \times N_a = N_p \times N_a$ matrix **E** defined by

$$\mathbf{E} = \begin{bmatrix} \mathbf{e} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{e} & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{e} \end{bmatrix} \quad \textbf{(3.6)}$$

Where **O**s are null matrices of appropriate sizes. Note that for our case, i.e., ln$k$ is only the unknown in all gridblocks, **E** will be actually an M-dimensional vector identical to the vector **e**. Hence, the vector $\boldsymbol{\theta}$ can be expressed as

$$\boldsymbol{\theta} = \theta\mathbf{E} = \theta\mathbf{e} \quad \textbf{(3.7)}$$

Note that the second equality of Eq. 3.7 follows from assumption of this work that lnk is only the attribute and its values at each gridblock are unknown.

Eq. 3.1 can be expressed as

$$\begin{aligned} O(\mathbf{m},\boldsymbol{\theta}) &= \frac{1}{2}\left[\mathbf{y}-\mathbf{Gm}\right]^T \mathbf{C}_D^{-1}\left[\mathbf{y}-\mathbf{Gm}\right] \\ &+\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-\mathbf{m}_{pr}-\boldsymbol{\theta}\right)+\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)^T \mathbf{C}_\theta^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right) \end{aligned} \quad \textbf{(3.8)}$$

Using Eqs. 3.1-3.7, it can be rewritten Eq. 3.9 for the case only ln$k$ is only unknown as

$$\begin{aligned} O(\mathbf{m},\theta) &= \frac{1}{2}\left[\mathbf{y}-\mathbf{Gm}\right]^T \mathbf{C}_D^{-1}\left[\mathbf{y}-\mathbf{Gm}\right] \\ &+\frac{1}{2}\left(\mathbf{m}-m_{pr}\mathbf{e}-\theta\mathbf{e}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-m_{pr}\mathbf{e}-\theta\mathbf{e}\right) \\ &+\frac{1}{2}\left(\theta\mathbf{e}-\theta_0\mathbf{e}\right)^T \mathbf{C}_\theta^{-1}\left(\theta\mathbf{e}-\theta_0\mathbf{e}\right) \end{aligned} \quad \textbf{(3.9)}$$

or

$$\begin{aligned} O(\mathbf{m},\theta) &= \frac{1}{2}\left[\mathbf{y}-\mathbf{Gm}\right]^T \mathbf{C}_D^{-1}\left[\mathbf{y}-\mathbf{Gm}\right] \\ &+\frac{1}{2}\left(\mathbf{m}-m_{pr}\mathbf{e}-\theta\mathbf{e}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m}-m_{pr}\mathbf{e}-\theta\mathbf{e}\right)+\frac{1}{2}\left(\theta-\theta_0\right)^2 \mathbf{e}^T\mathbf{C}_\theta^{-1}\mathbf{e} \end{aligned} \quad \textbf{(3.10)}$$

Taking the derivatives of the objective function given by Eq. 3.10 with respect to the unknown model vector $\mathbf{m}$ and the correction scalar's θ and equating them to zero gives, respectively,

$$\nabla_{\mathbf{m}} O(\mathbf{m},\theta) = \mathbf{C}_M^{-1}\left(\mathbf{m} - m_{pr}\mathbf{e} - \theta\mathbf{e}\right) - \mathbf{G}^T\mathbf{C}_D^{-1}\left[\mathbf{y} - \mathbf{Gm}\right] = 0 \qquad \textbf{(3.11)}$$

and

$$\frac{\partial O(\mathbf{m},\theta)}{\partial\theta} = -\mathbf{e}^T\mathbf{C}_M^{-1}\left(\mathbf{m} - m_{pr}\mathbf{e} - \theta\mathbf{e}\right) + \left(\theta - \theta_0\right)\mathbf{e}^T\mathbf{C}_\theta^{-1}\mathbf{e} = 0 \qquad \textbf{(3.12)}$$

It is needed to simultaneously solve Eqs. 3.11 and 3.12 for $\mathbf{m}$ and θ to obtain the posterior estimates $\tilde{\mathbf{m}}$ and $\tilde{\theta}$ and it can be shown that the solution is given by:

$$\begin{bmatrix} \mathbf{G}^T\mathbf{C}_D^{-1}\mathbf{G} + \mathbf{C}_M^{-1} & -\mathbf{C}_M^{-1}\mathbf{e} \\ -\mathbf{e}^T\mathbf{C}_M^{-1} & \mathbf{e}^T\mathbf{C}_M^{-1}\mathbf{e} + \mathbf{e}^T\mathbf{C}_\theta^{-1}\mathbf{e} \end{bmatrix} \begin{bmatrix} \mathbf{m} \\ \tilde{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_M^{-1} m_{pr}\mathbf{e} + \mathbf{G}^T\mathbf{C}_D^{-1}\mathbf{y} \\ -\mathbf{e}^T\mathbf{C}_M^{-1} m_{\mathrm{pr}}\mathbf{e} + \theta_0\mathbf{e}^T\mathbf{C}_\theta^{-1}\mathbf{e} \end{bmatrix} \qquad \textbf{(3.13)}$$

So, above matrix problem must be solved. It should be noted that Eq. 3.13 can be solved by LU decomposition for small sized problems, but it needed to use iterative or sparse matrix techniques for efficient storage and solution for large size problems.

In this applications it will be limited to this investigation to small size problems (such as M not exciding 200), and hence it will be used LU decomposition (Press et al., 1992).

It should be noted that it is considered that the observed data at some gridblocks defined by the user. Suppose, that it is considered a 1D block centered grid and have 20 uniformly spaced gridblocks (you can take the reservoir length L = 200 ft), and from left to right ordering, like N_x = 1,2,…,20. Suppose that 5 observed data for lnk at grid blocks, 1, 5, 10, 15, and 20 with Gaussian noise having zero mean and variance σ_d^2. So, the dimension of $\mathbf{y}_{obs}$ is N and will contain these 5 observed data.

Let the index l_j, j = 1,2,…,N, denote the grid blocks where observed lnk data are available (Note that for the above specific problem, l_1 = 1, l_2 = 5, l_3 = 10, l_4 = 15, l_5 = 20), then

$$\mathbf{y} = \begin{bmatrix} y_{l_1} \\ y_{l_2} \\ \vdots \\ y_{l_{N-1}} \\ y_{l_N} \end{bmatrix} \qquad \textbf{(3.14)}$$

Gm in Eq. 3.8 represents the linear relationship between the lnk model and “true” lnk data. Hence, the sensitivty matrix $\mathbf{G} = [g_{i,j}]$ is $N \times M$ matrix with all entries equal to zero or unity. Specifically, $g_{i,j} = 1$ if $j = l_i$ and $g_{i,j} = 0$ if $j \neq l_i$. So, for the specific 1D problem we considered, the matrix **G** will be 5×20 and will be given as

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \textbf{(3.15)}$$

The prior mean vector $\mathbf{m}_{pr}$ will be *M-dimensional* vector containing the prior mean of lnk, and the correction vector $\boldsymbol{\theta}_0$ will also be *M-dimensional* vector. For this simple example, these vectors will be 20-dimensional vectors, given by

$$\mathbf{m}_{pr} = \begin{bmatrix} m_{pr,1} \\ m_{pr,2} \\ \vdots \\ m_{pr,M-1} \\ m_{pr,M} \end{bmatrix} \quad \textbf{(3.16)}$$

and

$$\mathbf{m}_{pr} = \begin{bmatrix} \theta_{0,1} \\ \theta_{0,2} \\ \vdots \\ \theta_{0,M-1} \\ \theta_{0,M} \end{bmatrix} \quad \textbf{(3.17)}$$

It should be noted that in our applications, each elements of $\mathbf{m}_{pr}$ and $\boldsymbol{\theta}_0$ vectors will be taken as identical, i.e., $m_{pr,i} = m_{pr}$ and $\theta_{0,i} = \theta_0$, forr all $i = 1,2,\ldots,M$.

Regarding the matrices $\mathbf{C}_D$ and $\mathbf{C}_\theta$, the matrix $\mathbf{C}_D$ will be $N \times N$, whereas the matrix $\mathbf{C}_\theta$ will be $M \times M$. In our applications, we will assume that these matrices are diagonal with diagonal elements are identical. For example, the diagonal elements of $\mathbf{C}_D$ will be equal to σ_d^2, whereas the diagonal elements of $\mathbf{C}_\theta$ will be equal to σ_θ^2. So the inverses of these matrices are also diagonal, but the diagonal elements will be equal to the reciprocals of these variances. Due to this, it is not actually stored these matrices and their inverses.

As to the prior covariance matrix $\mathbf{C}_M$, this matrix is $M \times M$ but typically is not diagonal and will be generated from a given semi-variogram model (e.g., spherical, exponential, or Gaussian) with specified values of range and sill values. So, once $\mathbf{C}_M$ matrix is generated, LUBKSK and LUDCMP codes should be used to obtain the inverse of $\mathbf{C}_M$ matrix. It will be stored as the $2M \times 2M$ matrix which is denoted by $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} \mathbf{G}^T \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_M^{-1} & -\mathbf{C}_M^{-1} \\ -\mathbf{C}_M^{-1} & \mathbf{C}_M^{-1} + \mathbf{C}_\theta^{-1} \end{bmatrix} \quad \textbf{(3.18)}$$

and it will be stored the vector which is denoted by $\mathbf{r}$

$$\mathbf{r} = \begin{bmatrix} \mathbf{C}_M^{-1} m_{pr} \mathbf{e} + \mathbf{G}^T \mathbf{C}_D^{-1} \mathbf{y} \\ -\mathbf{e}^T \mathbf{C}_M^{-1} m_{\mathrm{pr}} \mathbf{e} + \theta_0 \mathbf{e}^T \mathbf{C}_\theta^{-1} \mathbf{e} \end{bmatrix} \quad \textbf{(3.19)}$$

Then, Eq. 3.13 can be written as

$$\mathbf{H}\hat{\mathbf{m}} = \mathbf{r} \quad \textbf{(3.20)}$$

Where it is defined the $2M$-dimensional vector $\hat{\mathbf{m}}$ as

$$\hat{\mathbf{m}} = \begin{bmatrix} \tilde{\mathbf{m}} \\ \tilde{\theta} \end{bmatrix} \quad \textbf{(3.21)}$$

Eq. 3.9 can be solved by LU decomposition method based on the use of LUBKSK and LUDCMP codes.

3.1 An Example Application

Here, we consider an example application to demonstrate the applicability of the partially doubly stochastic model for a linear underdetermined problem of reservoir characterization. We consider a 1D reservoir problem having a multi-Gaussian distribution of ln$k$ with prior mean m_{pr} and covariance matrix $\mathbf{C}_M$ based on spherical variogram. It is assumed that the reservoir is 1000 ft long, and discretize the reservoir into 100 uniformly spaced grid blocks each having a length of 10 ft. The true mean of ln$k$ is taken ln$k$ = 3, and variance of ln$k$ = 1. It will be used a spherical variogram with a range of 200 ft. One realization (to be considered as the true) of ln$k$ is shown in Figure 3.1.

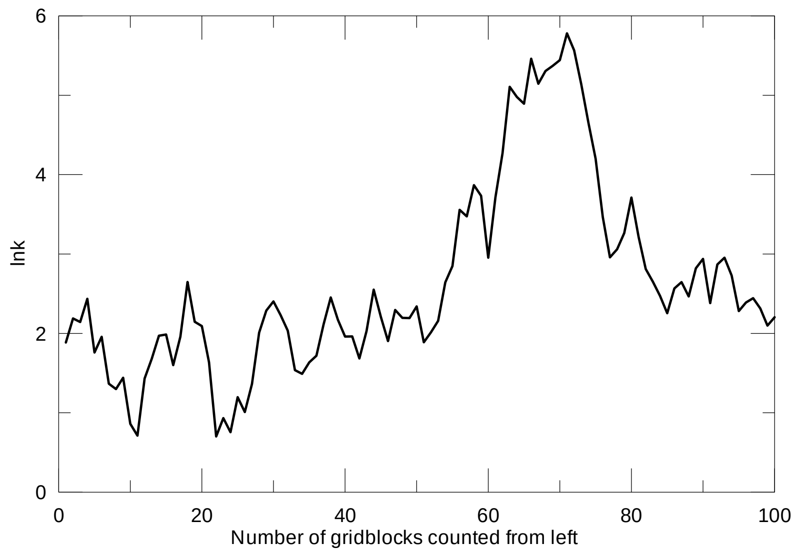

Figure 3.1: True ln*k* field, generated by using the Cholesky decomposition model.

Now, it will be assumed that 11 values of ln*k* are collected from grid blocks, 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 with Gaussian noise having zero mean and variance 0.1. These 11 sampled values of ln*k* along with the true field of ln*k* is shown in Fig. 3.2. Table 3.1 lists the noisy sampled values of ln*k* with respect to grid-block numbers.

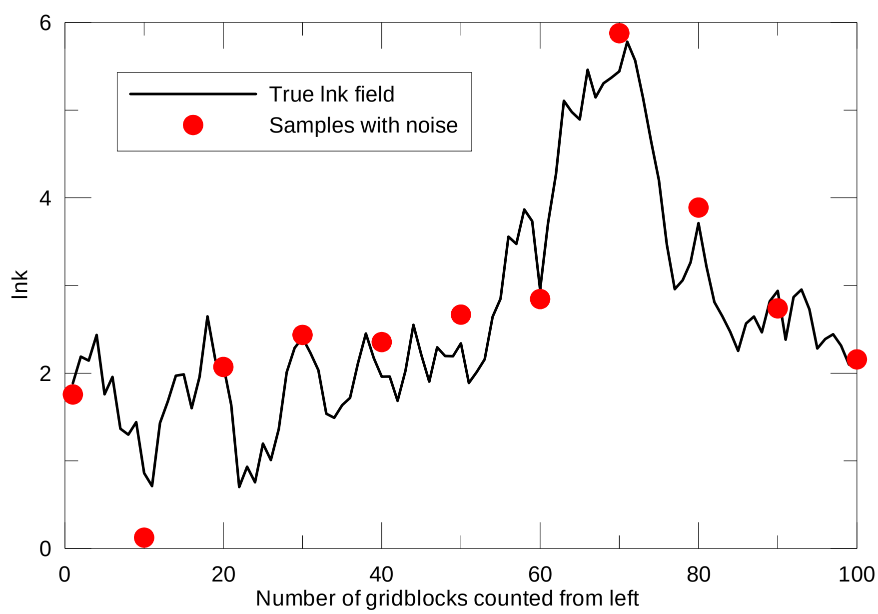

Figure 3.2: 11 samples (with Gaussian noise) of ln*k* and true ln*k* field.

It will be tried to estimate the entire field using 11 noisy samples and the given prior mean and the covariance matrix.

Table 3.1: 11 Samples of ln*k* with noise with their associated grid-block numbers

Its gridblock number	Value of Sample lnk
1	1.7568
10	0.1243
20	2.0706
30	2.4361
40	2.3547
50	2.6675
60	2.8451
70	5.8771
80	3.8883
90	2.7391
100	2.1570

The respond of model is searched in circumstances the model with correction and without correction. For this case, the true value of the prior mean is equal to 3. The variance of correction is 10 and the output value of correction is computed as $\theta = -0.3671$. This result is understandable to given true value of prior mean.

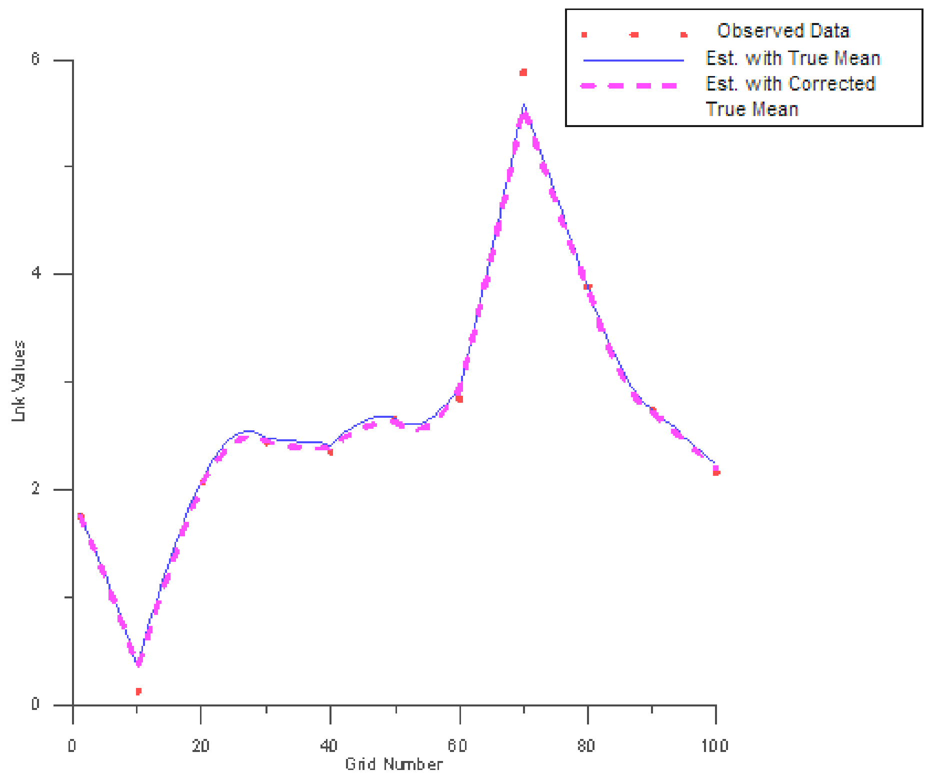

Figure 3.3: Comparison of observed data, the model with correction and without correction.

Figure 3.3 shows that the model with or without correction is perfectly match the observed data in given grids.

Case 1: The prior mean is taken as m_{pr} = 0, and θ_0 = 0, and taken 7 values of $\sigma_\theta^2 = 0.01, 0.1, 1, 10, 50, 100,$ and 1000 estimate the posterior estimates of m_{pr} and θ solving Eq. 3.13.

The models constituted wrong prior mean (m_{pr} = 0) is investigated with or without correction and by assigning small and high values to correction variances, the differences of plots are analyzed. The output values of corrections according to assigned correction variances are given in Table 3.2.

Table 3.2: The output of correction values with wrong prior mean and different values of correction variance.

Prior Mean	Correction	Correction variance, σ_θ^2
0	0.2247	0.01
0	1.4428	0.1
0	3.1502	1
0	3.5731	10
0	3.6163	50
0	3.6217	100
0	3.6272	10000

Table 3.2 demonstrates that when the value of correction variance is higher, the effect of correction can be realized clearly on model parameter ln*k*.

In Fig. 3.4, whereas the true model try to appoint true mean to grids where we have no observed data, the estimated values of ln*k* from the models with wrong mean and small correction goes through the wrong mean. It is obvious that the model with small correction variances are not enough in order to correct to wrong prior mean of ln*k*. However, when the correction variance takes a value of 0.1, it is better fit than the model without correction. Still, in the grids which have data, the all models works good.

Fig. 3.5 indicates that the correction works quite successfully to match the true model with variance 1 and 10. Yet, the ln*k* values are still not goes to mean in the grids where we have no observed data. While wrong mean model try to assign to grids 0, the corrected models match the data perfectly.

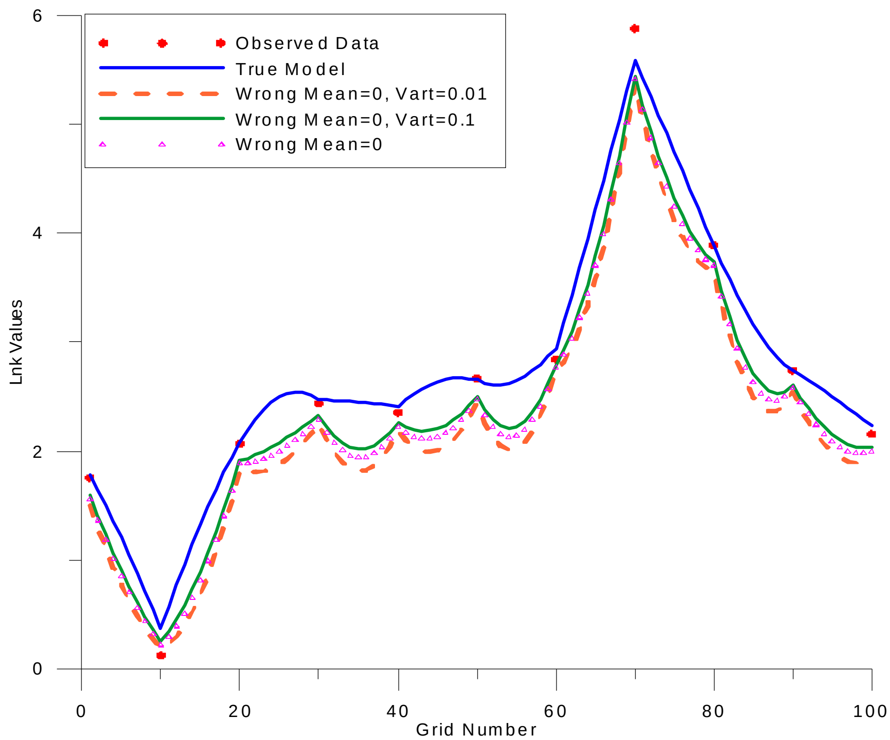

Figure 3.4: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

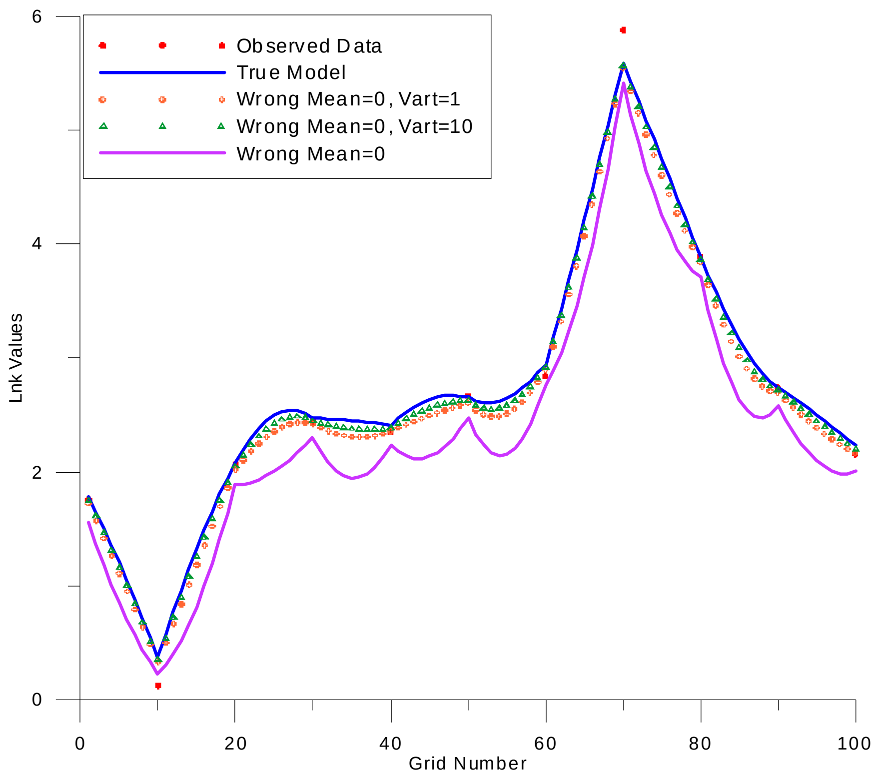

Figure 3.5: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

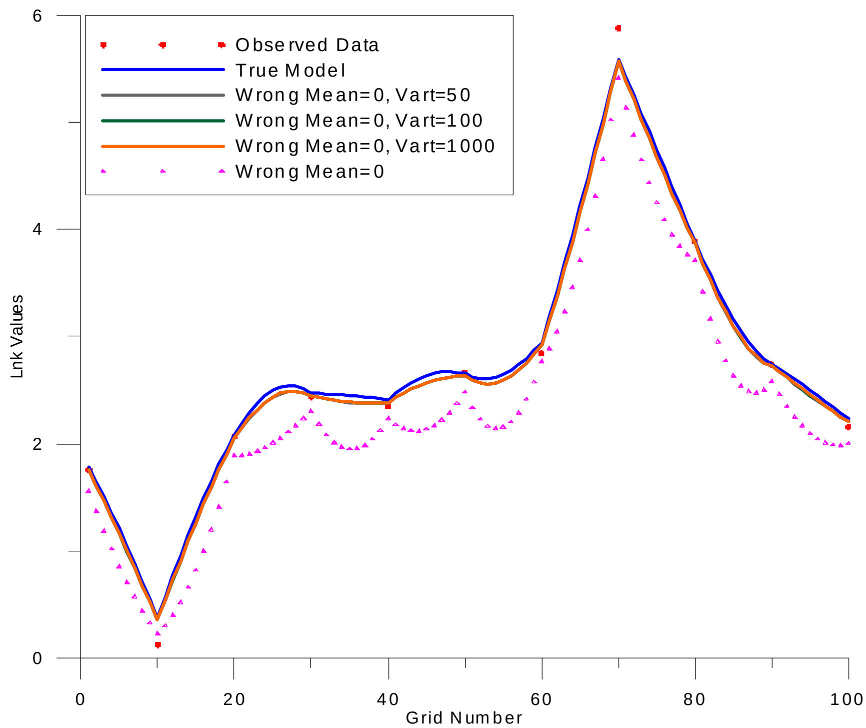

Figure 3.6: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

The results obtained from Fig. 3.6 are almost same result of Figure 4.4. With increasing the values of correction variance, the model match better to the true model.

Case 2: The prior mean is taken as m_{pr} = -1, and θ_0 = 0, and taken 7 values of $\sigma_\theta^2 = 0.01, 0.1, 1, 10, 50, 100,$ and 1000 estimate the posterior estimates of m_{pr} and θ solving Eq. 3.13.

The models constituted wrong prior mean (m_{pr} = -1) is investigated with or without correction and by assigning small and high values to correction variances, the differences of plots are analyzed. The output values of corrections according to assigned correction variances are given in Table 3.3.

In Table 3.3, when we initially assumed wrong prior mean -1, the output of the program gives the value of the corrections by changing the value of the correction variance. The correction values stablize when the correction variance gets the value of 50. After that the values of ln$k$ and correcton are almost constant.

Table 3.3: The output of correction values with wrong prior mean and different values of correction variance.

Prior Mean	Correction	Correction variance, σ_θ^2
-1	0.2247	0.01
-1	1.4428	0.1
-1	3.1502	1
-1	3.5731	10
-1	3.6163	50
-1	3.6217	100
-1	3.6272	10000

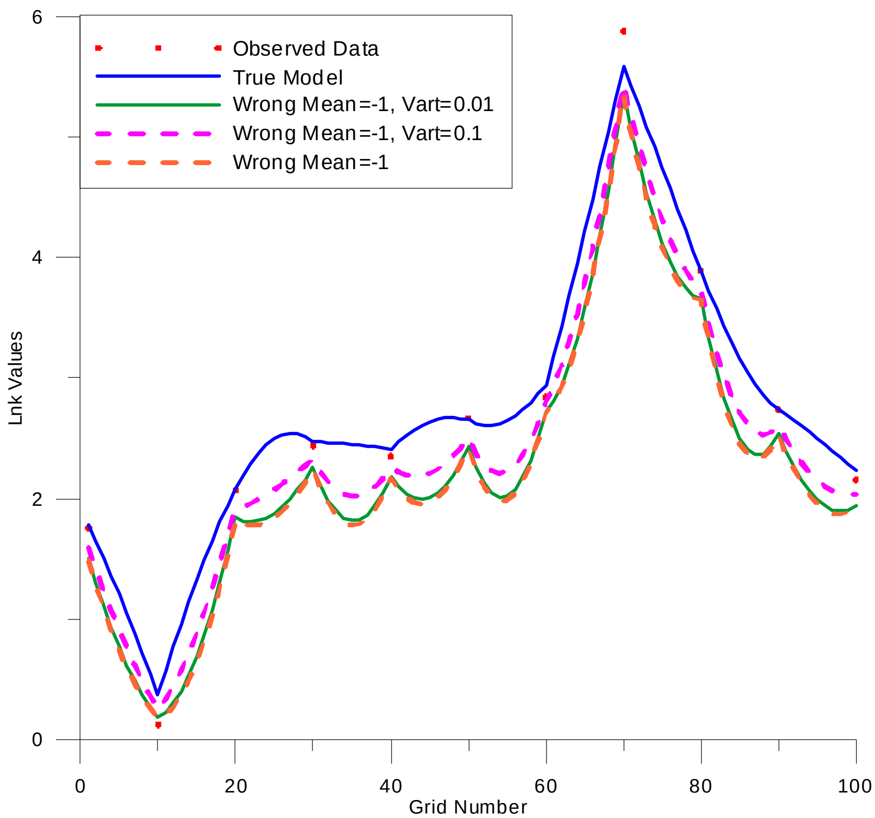

Figure 3.7: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

Figure 3.7 represents that the gridblocks with no observed data are gotten cambered values by the wrong prior model with or without correction. When we increase the value of correction variance, the model begins to better match than just using wrong prior.

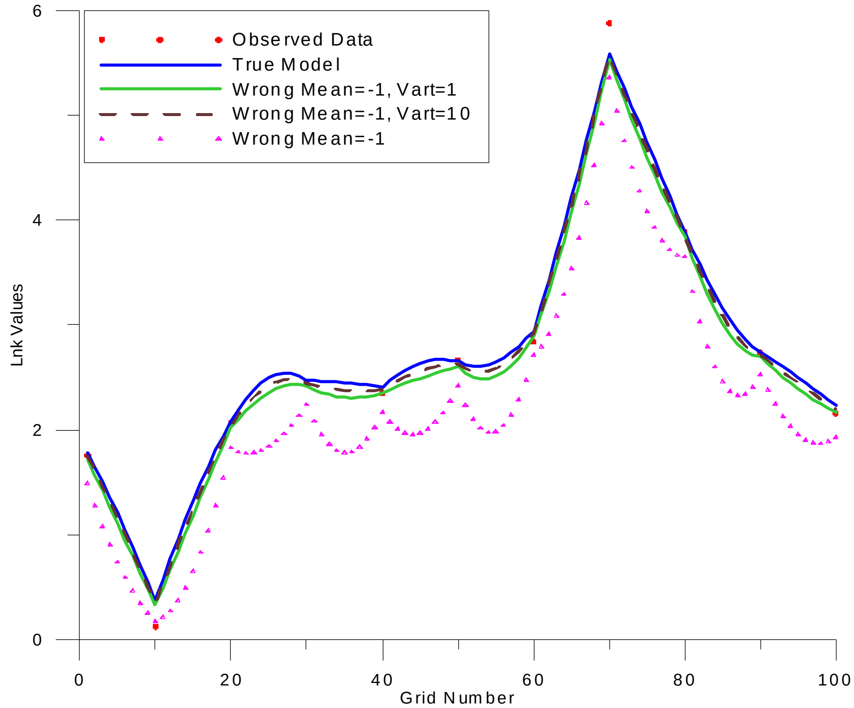

Figure 3.8: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

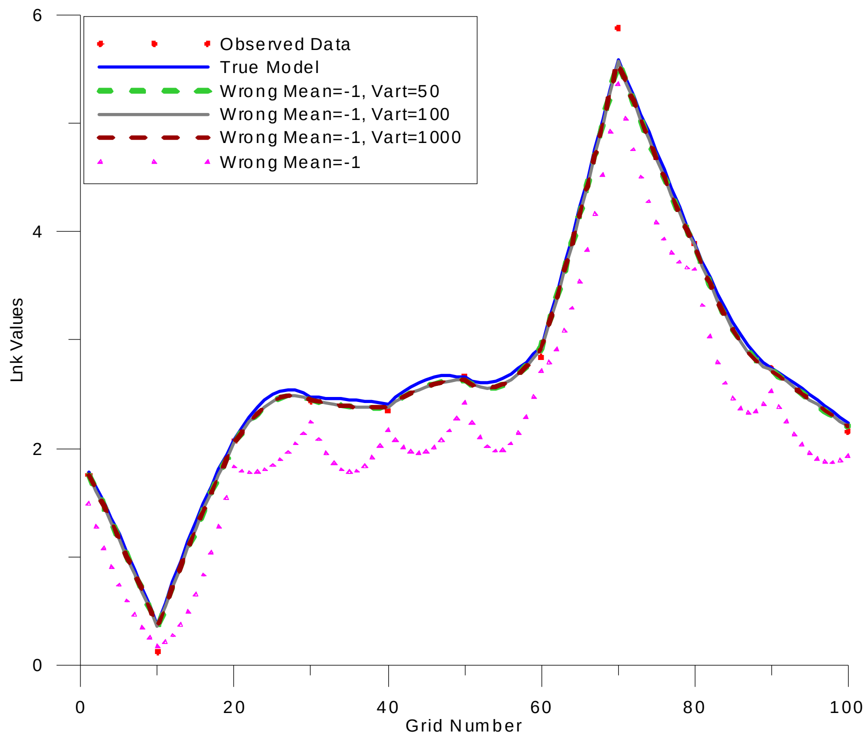

Figure 3.9: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

In Figure 3.8, the corrected ln*k* values the gridblocks having no observed data are gotten cambered values by the wrong prior model with or without correction. When the correction variances are increased, a good match is obtained with the true model.

Figure 3.9 exposes that without correction, prior mean can affect to make wrong decisions. Because of that, if the engineer is mistrust the prior mean, he/she must be use correction with higher variance, i.e., higher than 10. By this way, the rock property of field is generated well without errors.

Case 3: The prior mean is taken m_{pr} = 8, and θ_0 = 0, and taken 7 values of $\sigma_\theta^2 = 0.01, 0.1, 1, 10, 50, 100,$ and 1000 estimate the posterior estimates of m_{pr} and θ solving Eq. 3.13

The models constituted wrong prior mean (m_{pr} = 8) is investigated with or without correction and by assigning small and high values to correction variances, the differences of plots are analyzed. The output values of corrections according to assigned correction variances are given in Table 3.4.

In Table 3.4, the correction values stablize when the correction variance gets the value of 50. After that the value of ln*k* is almost constant. Besides, the correction term works even if the value of the prior mean is very wrong.

Table 3.4: The output of correction values with wrong prior mean and different values of correction variance.

Prior Mean	Correction	Correction variance, σ_θ^2
8	-0.3328	0.01
8	-2.137	0.1
8	-4.6662	1
8	-5.2926	10
8	-5.3565	50
8	-5.3646	100
8	-5.37266	10000

Figure 3.10 shows that, when we assumed prior mean incorrect, the results include big errors except sample data points. When we use partially doubly stoachastic method, as before, if the values of correction variances are smaller or the model has no correction then the ln*k* values of field tend to take wrong mean.

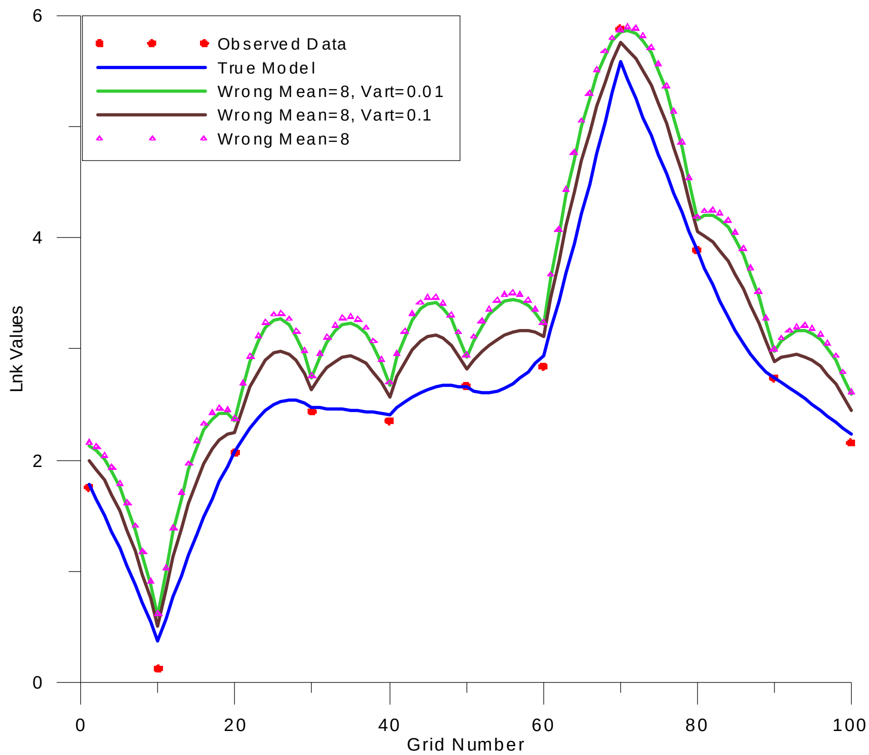

Figure 3.10: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

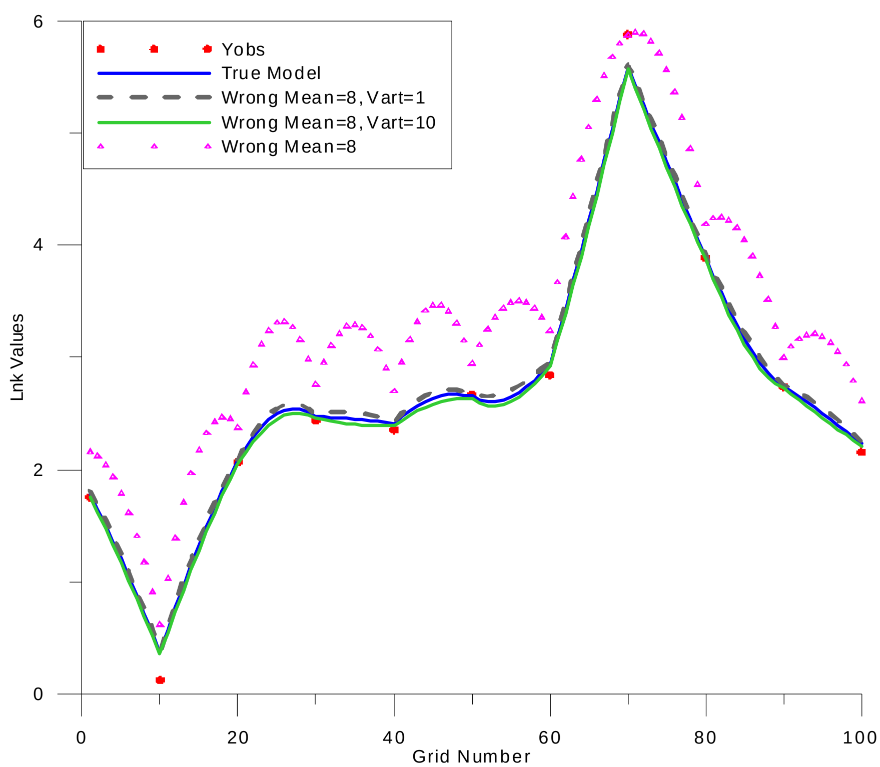

Figure 3.11: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

As it can be easily seen in Figure 3.11, even if the prior mean has very big error, the model which has correction variances equal or bigger than 1 provides almost perfect match with true model.

Generally, if the value of correction variance is chosen approximately 10, then, the model values are scattered well in grids just like true mean how so ever great the error in prior mean may be.

Case 4: The number of sample is reduced from 11 to 3 and the samples are in gridbloks 10th ,70th , and 100th . The prior mean is taken as m_{pr} = 3, and θ_0 = 0, and taken 7 values of $\sigma_\theta^2 = 0.01, 0.1, 1, 10, 50, 100,$ and 1000 estimate the posterior estimates of m_{pr} and θ solving Eq. 3.13.

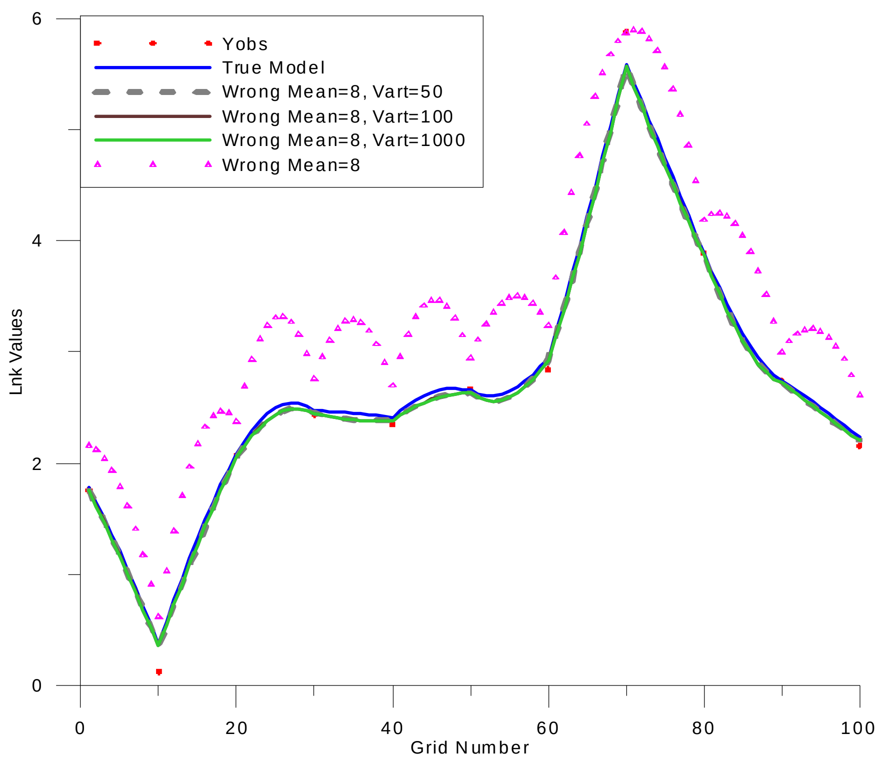

Figure 3.12: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

As can be seen in Fig. 3.13, the true model with 11 sample does not match the true model with 3 data. If the sample number decreases, the model which is wanted to represent true values is not computed. However, by using correction, the better model can be obtained even with 3 data.

The models constituted true prior mean (m_{pr} =3) is investigated with or without correction and by assigning small and high values to correction variances, the differences of plots are analyzed. The output values of corrections according to assigned correction variances are given in Table 3.5.

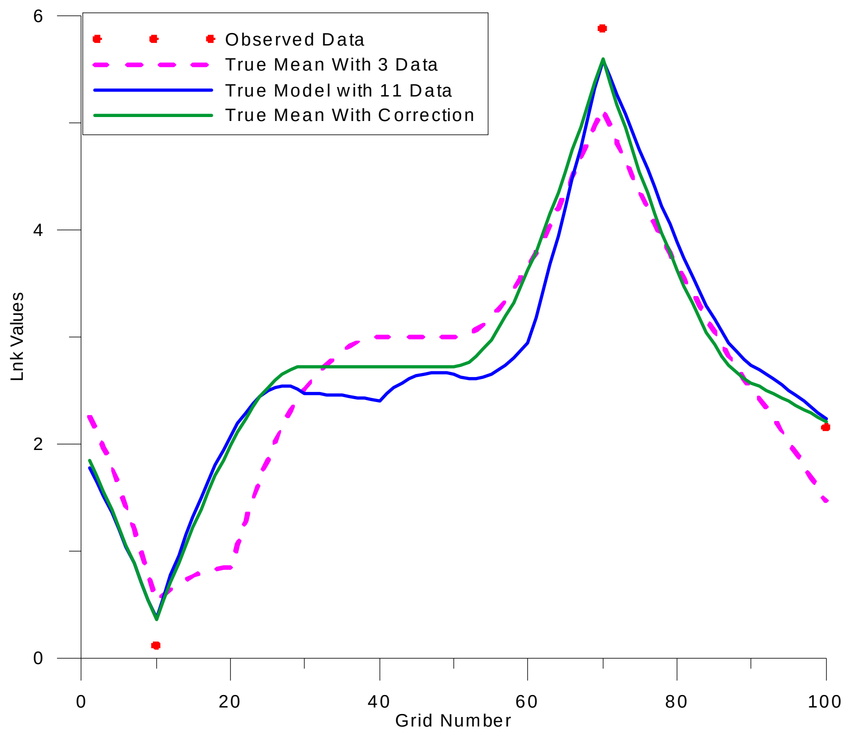

Figure 3.13: Comparison of observed data, the true model with 3 data, 11 data and correction.

Table 3.5: The output of correction values with wrong prior mean and different values of correction variance.

Prior Mean	Correction	Correction variance, σ_θ^2
-1	0.0987	0.01
-1	0.0797	0.1
-1	2.7215	1
-1	3.5879	10
-1	3.7058	100
-1	3.7193	10000

As previous case, correction with smaller variances are not good enough to generate model parameter ln*k*. Still, with larger variances the model can be represent the true rock field.

When the wrong prior has a value -1 and no correction on prior, then the computed model cannot match the true model and tends to go to the value -1. Whereas, the model can be obtained by using 1 or upper value of correction variance.

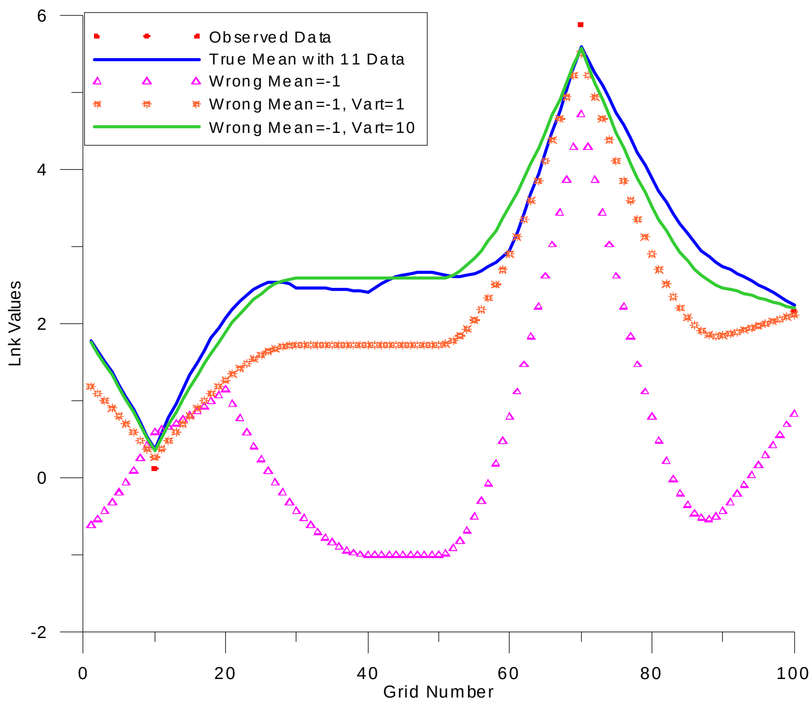

Figure 3.14: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

The model is obtained almost perfectly when the correction variance takes the value 1000. In Figure 3.15, there is a big difference between wrong prior model without correction and with correction.

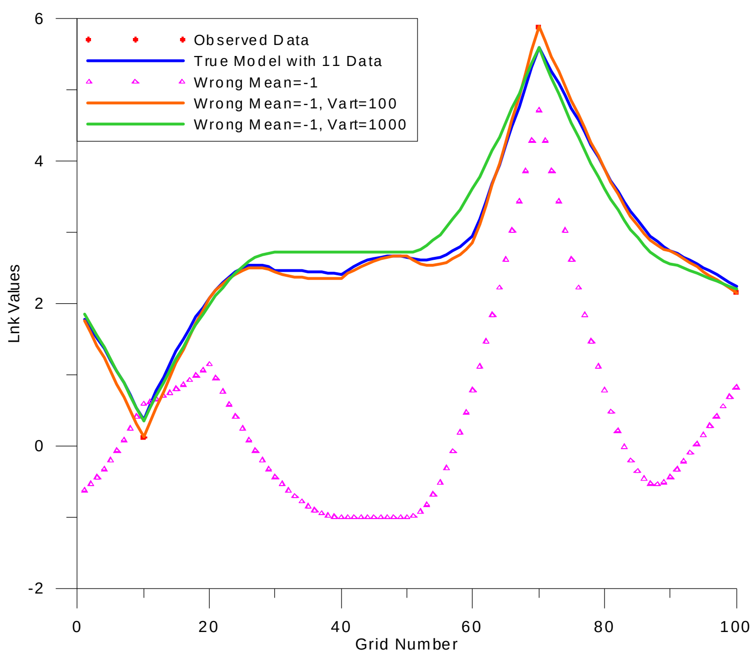

Figure 3.15: Comparison of observed data, the true and wrong model without correction and with wrong prior mean and different values of correction variance.

The models constituted true prior mean (m_{pr} =3) is investigated with or without correction and by assigning small and high values to correction variances, the differences of plots are analyzed. The output values of corrections according to assigned correction variances are given in Table 3.6.

Case 5: The uncertainty of sample is reduced from 10^{-1} to 10^{-5} . The prior mean is taken m_{pr} = 3, and θ_0 = 0, and taken 7 values of $\sigma_\theta^2 = 0.01, 0.1, 1, 10, 50, 100,$ and 1000 estimate the posterior estimates of m_{pr} and θ solving Eq. 3.13.

Figure 3.16 shows that as the variance of sample or uncertainty on sample is reduced, the model generates directly mean of the sample for the grids having no observed data.

Table 3.6: The output of correction values with wrong prior mean and different values of correction variance.

Prior Mean	Correction	Correction variance, σ_θ^2
-1	0.2385	0.01
-1	1.49889	0.1
-1	3.1776	1
-1	3.5784	10
-1	3.6241	100
-1	3.6292	10000

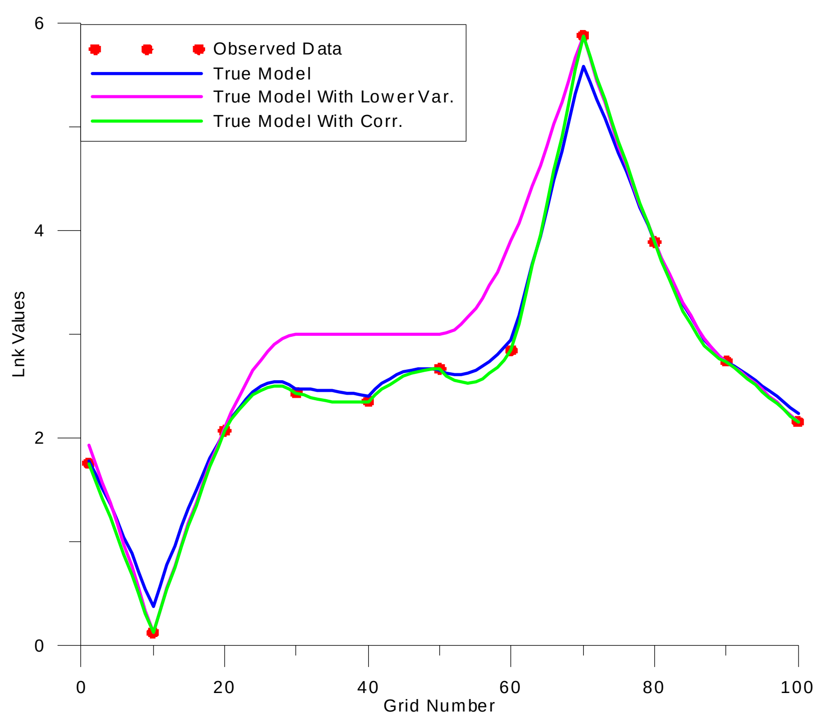

Figure 3.16: Comparison of observed data, the true model with lower variance and correction.

Although the uncertainty of the sample is reduced and with true prior the model generates data same as mean, the model can not match the true model when wrong prior is used. Still, by using partially doubly stochastic model, the true model can be obtained by using bigger correction variance.

Case 6: The same inputs are used in Case 5. This time, the model is run in GSlib commercial program for Ordinary Kriging (OK). The results of GSlib (Url-4) is compared to DS model result obtained Case 5.

It can be seen in Fig. 3.17, even if the mean is wrong, the partiall doubly stochastic model affects the model like Ordinary Kriging. Especially, if the uncertainty on sample is low, the model from ordinary kriging and the model from partially doubly stochastic model perfectly match.

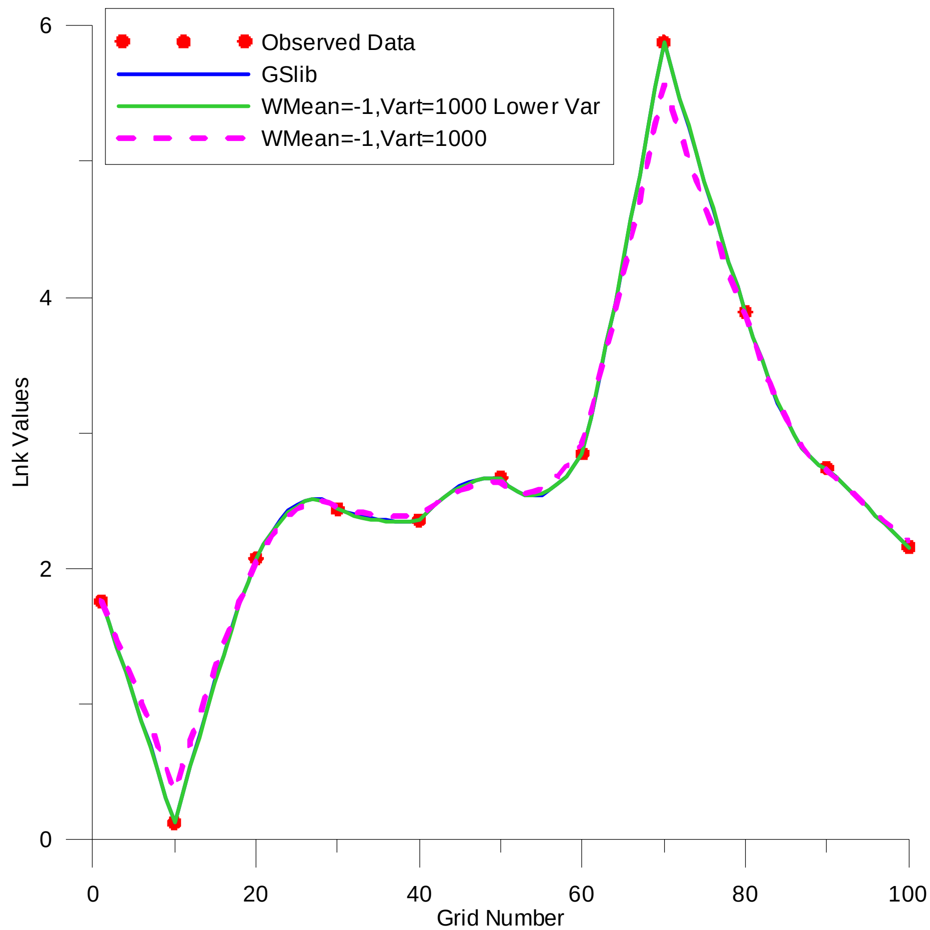

Figure 3.17: Comparison of observed data, model from GSlib and wrong corrected model with lower variance.

Case 7: Conditional realizations are generated with changing seed number of the program. The prior mean is taken -1 and it is investegated that how the model behaves with/without partially doubly stochastic model.

First of all, 10 conditional realizations are generated without partially doubly stochastic model and the prior mean is taken 3. Then, wrong prior mean is assumed -1 and without correction term, 10 conditional realizations are generated. Finally, with wrong prior -1, the partially doubly stochastic model is used and 10 conditional realizations are generated according to this model. Each figures includes both maximum posterior estimations and 10 conditional realizations.

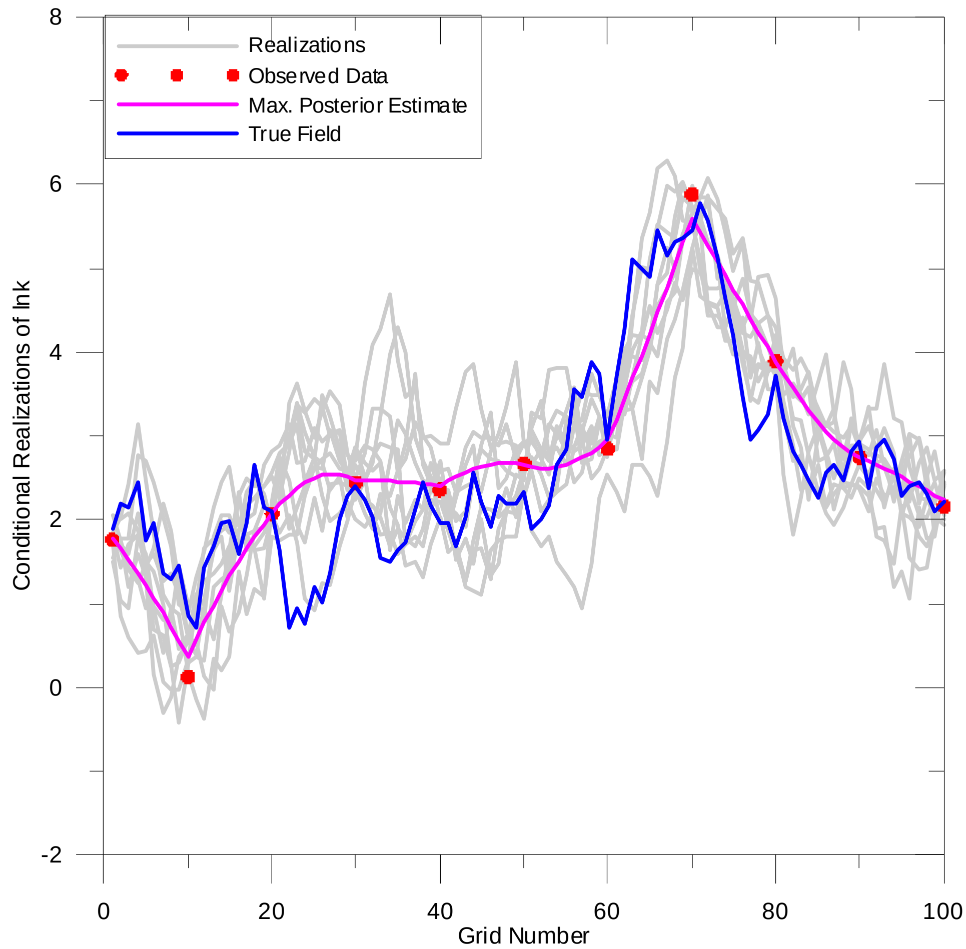

Figure 3.18: Comparison of true field, conditional realizations and posterior estimate with true prior mean, $m_{pr} = 3$.

Figure 3.18 shows that, maximum posterior estimate reflects both true field and observation data. True field seems to be rough and maximum posterior estimate of this field is smooth. While making performance predictions we need these rough realizations.

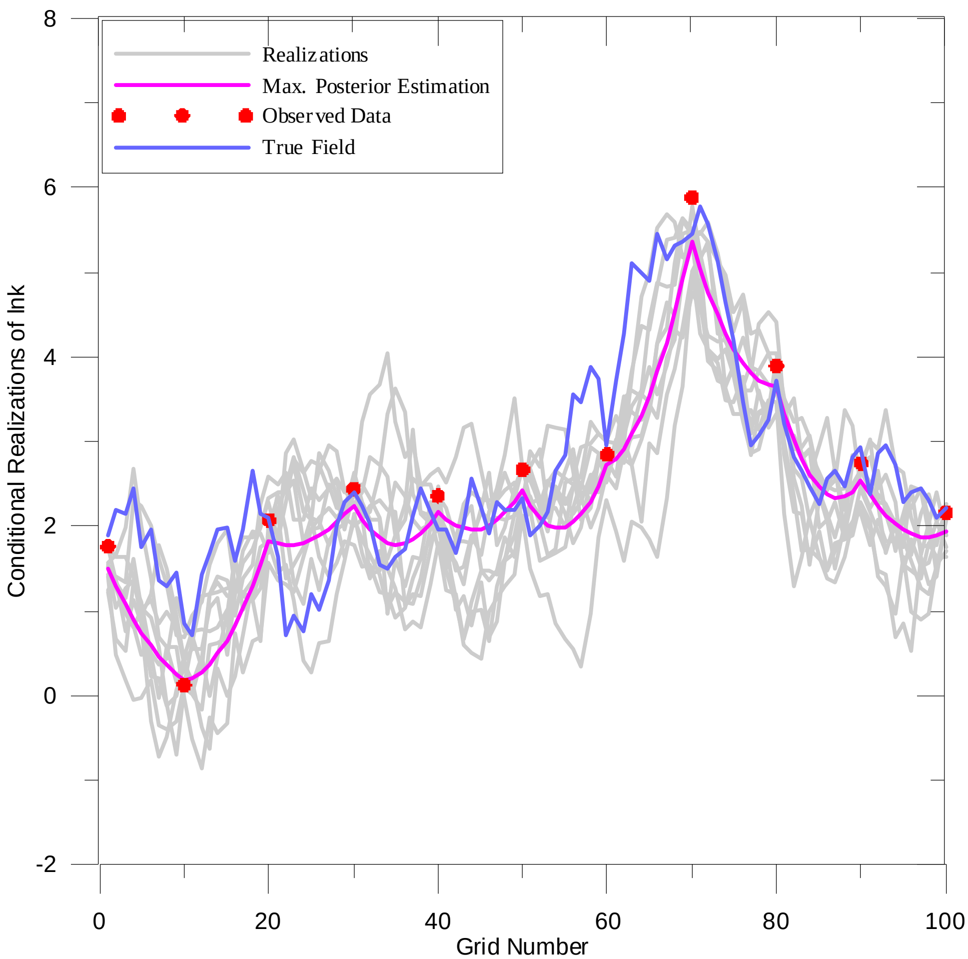

Figure 3.19: Comparison of true field, the conditional realizations and posterior estimate with wrong prior mean, m_{pr} = -1 and without correction.

It can be seen in Figure 3.19, the maximum posterior estimation of ln*k* make a convex through the wrong prior mean -1 and the conditional realizations of field is working at sample points. However, at the gridblocks without data, the conditional realizations would not represent the true field of ln*k*. The uncertainty of conditional realizations increases if the prior mean is not known.

Contrary of Figure 3.19, Figure 3.20 is more reliable and appreciable if the prior mean is assumed incorrect. The posterior estimate of lnk includes almost all data available and it matches to posterior estimation with true mean value. In addition of this, the roughness of conditional realizations decreases when we use the partially doubly stochastic model. Besides, the conditional realizations are generated partially doubly stochastic method with wrong prior mean reflect the conditional realizations generated with true prior mean. Because of these, when we estimate the performance prediction of reservoir, partially doubly stochastic method should be used for avoiding big mistakes.

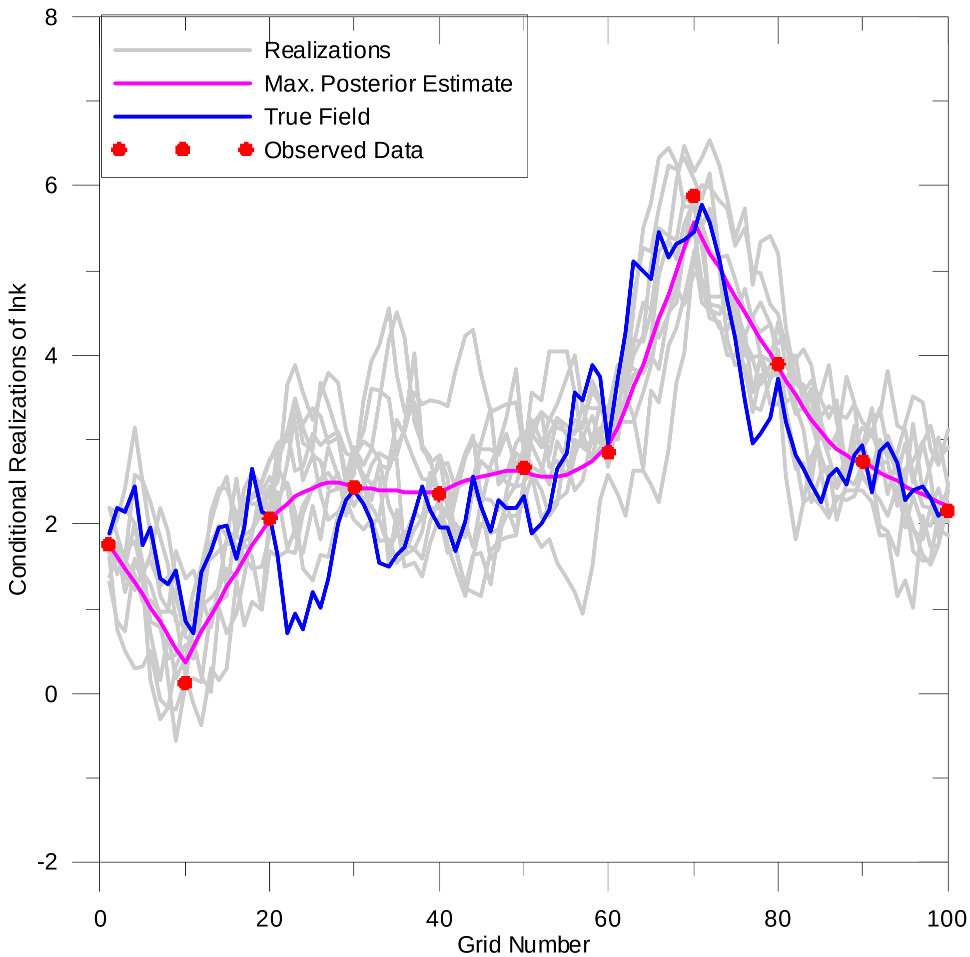

Figure 3.20: Comparison of true field, the conditional realizations and posterior estimate with wrong prior mean, m_{pr} = -1 and correction.

4. OVER-DETERMINED PROBLEM APPLICATION TO A PRESSURE TRANSIENT TEST DATA SET

In this chapter, an application to a pressure transient test data set is considered to investigate the use of prior information in history matching for the over-determined nonlinear problem. Here, the use of prior information is investigated with or without uncertainty in the prior means of the model parameters.

For this investigation, we consider a simulated multi-rate test example for which a fully penetrating vertical well is located near a single no-flow (sealing) fault in an infinite homogeneous, isotropic rectangular. The schematic of well/reservoir configuration is shown in Fig. 4.1. The distance between the well and fault is 150 ft. The input rock and fluid properties are listed in Table 4.1.

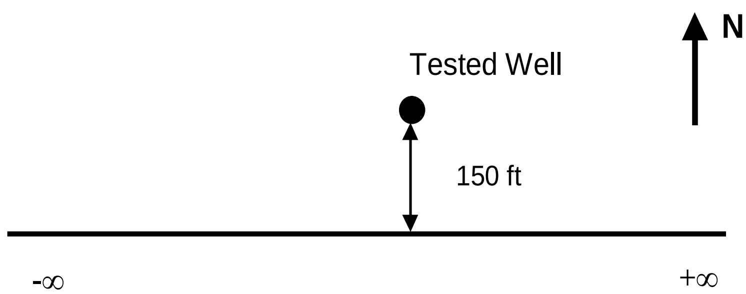

Figure 4.1: Schematic representation for the well/reservoir configuration for a vertical well located near a sealing fault in an infinite, homogeneous, and isotropic single-layer reservoir.

The simulated well pressures as well as the flow rate history at the tested well are displayed in Fig. 4.2. The test sequence contains one drawdown and one buildup periods. The total duration of the test is 30 hr as can be seen from Fig. 4.2. The duration of the drawdown (DD) period is 10 hr, the duration of the buildup (BU) period is 20 hr. Flow rate during the drawdown period is 1000 B/D, respectively.

Table 4.1: Input parameters for a synthetic test in a closed rectangle homogeneous, isotropic reservoir (Fig. 4.1).

Parameters	Values
ϕ (fraction)	0.20
h (ft)	30
c_t (psi^{-1})	1.0×10^{-5}
μ (cp)	1.0
r_w (ft)	0.354
S (dimensionless)	5
C_w (B/psi)	1.0×10^{-2}
B (RB/B)	1
k (md)	20.
p_i (psi)	5000
L (distance to the sealing fault, ft)	150

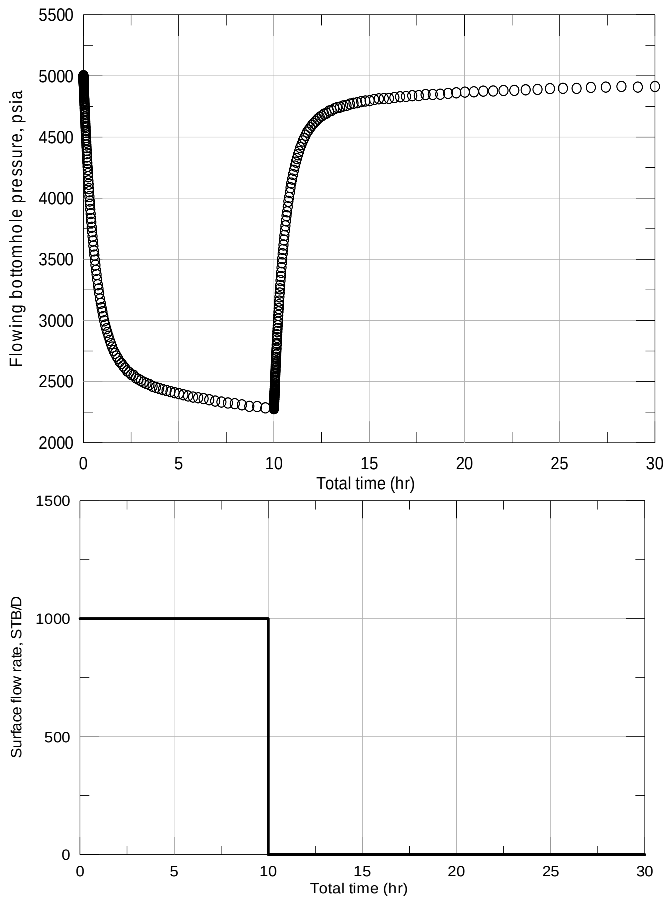

Figure 4.2: Pressure and flow rate history at the tested well.

A Gaussian noise with mean equal to zero and standard deviation equal to 2.0 psi has added to the tested well pressures.

As is well-known (Kuchuk et al. 2010), the first step in pressure transient interpretation and analysis is model identification based on the diagnostic log-log plots of pressure changes and Bourdet derivatives (based on Agarwal's equivalent time) for selected flow and/or buildup periods. Note that Bourdet derivative is the derivative of pressure change with respect to the natural logarithm of the Agarwal equivalent time. Typically, the buildup periods and good quality drawdown periods would be chosen. The log-log diagnostic plots of pressure changes and Bourdet derivatives for the drawdown and buildup periods (denoted as DD and BU) are displayed in Fig. 4.3. The hollow and stuffed symbols in Fig. 4.3 represent the pressure change and its logarithmic derivative, respectively. Buildup derivatives were taken with respect to the Agarwal equivalent time and then plotted as a function of elapsed time.

As is well known, log-log diagnostic plots such as the one shown in Fig. 4.3 are useful to identify the specific flow regimes exhibited by the test data and their time intervals. The diagnostic log-log plots of the tested well responses for the drawdown period (Fig. 4.3) indicate an infinite-acting radial flow (zero slope line in the Bourdet derivative) in the time interval from 6 to 10 hr, from which it can be estimated the values of the permeability (k) and skin factor (S) using the well known infinite-acting radial flow regime equations (Earlougher, 1977; Bourdet 2002; Kuchuk et al., 2010). On the other hand, the log-log diagnostic plot for the buildup period exhibit an increasing trend after 6 hr, which may indicate that the well is located near a sealing fault. However, the doubling zero-slope lines are not evident due to the short duration of the buildup (as well as the drawdown) period. For sure, for this test, to identify the appropriate interpretation model, additional information would be needed from other sources such as logs, cores as well as geological and geophysical data. The data from such sources may provide additional information that the reservoir under consideration is a homogeneous, single-layer system, and the well is located near a sealing fault. It should be noted that model identification based on merely the derivative response is not sufficient to identify a unique well/reservoir model for the test data considered. Hence, in practice, *a prior* information will be always needed regarding the true well/reservoir model from other independent sources mentioned above that can lead to Bourdet derivative responses responsible for the behavior observed after 6 hr. Now suppose that the model has *a priori* information from other sources supporting the hypothesis that the reservoir is homogeneous and the well is located near a sealing fault.

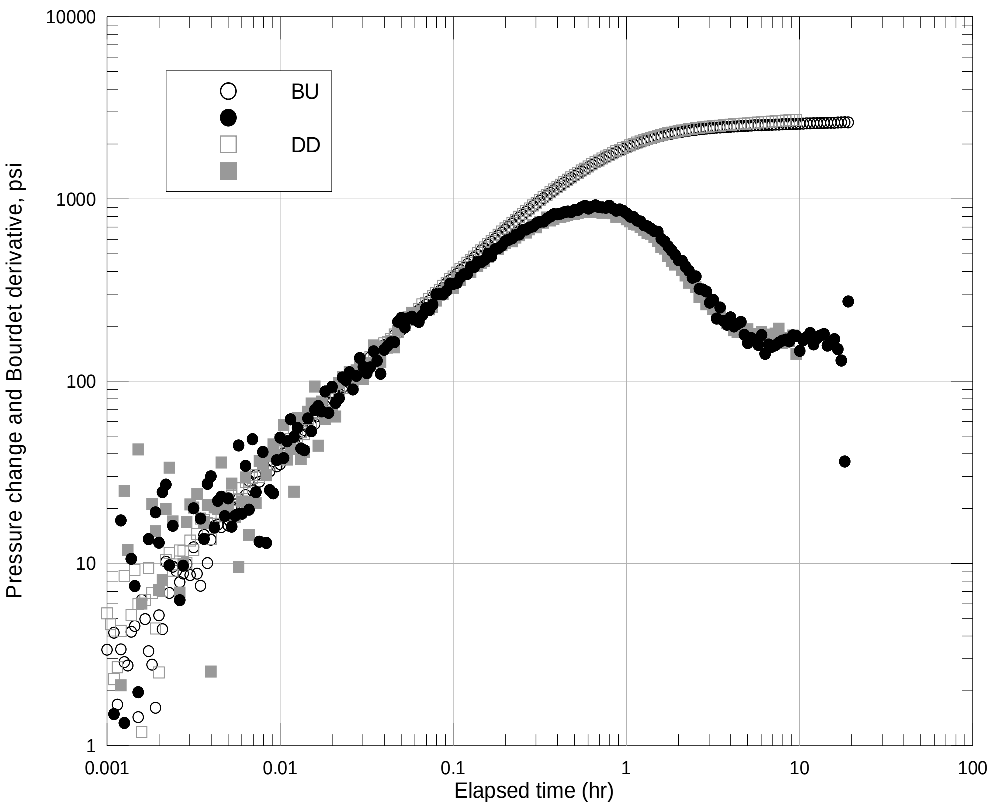

Figure 4.3: Pressure changes and Bourdet derivatives for BU and DD periods at the tested well.

The next step in modern pressure transient interpretation and analysis is to perform pressure-rate deconvolution to minimize the effects of variable rate history by converting the tested well pressure data recorded with a variable rate history into equivalent constant-rate responses (von Schroeter et al., 2004; Levitan, 2005; Onur et al., 2008; Pimonov et al., 2009; Kuchuk et al., 2010). Recall that deconvolution is data processing procedure to reconstruct the drawdown responses that would be obtained if the well were produced at a constant rate from the beginning of the test to the end of the variable flow rate history given (Kuchuk et al. 2010). For instance, for the test example case considered in Fig. 4.2, we would like to reconstruct an equivalent constant-rate drawdown pressure change and its Bourdet derivative for a total duration of 30 hrs. On the other hand, it should be noted that the model identification using conventional log-log diagnostic plots based on individual flow periods is limited only to the duration of the flow period chosen. For example, the conventional log-log diagnostic plot based on the buildup (BU) period for the test example considered in Fig. 4.2 is limited to the duration of this buildup period, which is 20 hrs, as shown in Fig. 4.3, whereas deconvolution will provide 30-hr long equivalent constant-rate drawdown responses. So, deconvolved equivalent constant-rate responses not only will provide a longer response for better identification of the reservoir/well system for this test example, but also avoid the misinterpretation of the complicated late time derivative response of the buildup portions due to the effects of producing time and variable rate if it is going to used conventional log-log diagnostic plots based on buildup portions for model identification.

The more recent robust deconvolution algorithm of Pimonov et al. (2009) is used for this example. Here, for simplicity, it will be assumed that flow rate data at the tested well and initial static pressure are accurately known so that they can be treated as known during the deconvolution process. In cases where such data are uncertain due to errors, one should consider the general deconvolution methodology given by Kuchuk et al. (2010) to get around to these problems. Fig. 4.4 shows the deconvolved responses for the tested well, for which only the BU pressure data (for a duration of 20 hrs) were used for deconvolution, in comparison with the conventional pressure change and derivative responses, based on the Agarwal equivalent time, for the BU and DD periods. All deconvolved and conventional responses shown in Fig. 4.4 were based on the same constant reference rate of 1000 B/D. In Fig. 4.4, the deconvolved responses shown by solid curves represent the deconvolved pressure change, whereas deconvolved responses shown by dashed curves represents deconvolved Bourdet derivatives.

It is clear from Fig. 4.4, deconvolved derivative responses are quite smooth and, for times greater than 6 hr, exhibit a drawdown derivative behavior of a well located near a sealing fault and, unlike the conventional derivative responses of the drawdown and buildup periods does not indicate an infinite acting radial flow. The appearance of the infinite acting radial flow regime observed in Bourdet derivatives of the buildup and drawdown periods seems to be due to the noise in pressure data. From the results shown in Fig. 4.4, it is without doubt that deconvolution provides an excellent tool for identifying the appropriate reservoir model for this synthetic example. As it will be shown, we could actually estimate most of the well/reservoir parameters of interest, *k*, *S*, C_w, and *L*, reliably from the deconvolved equivalent drawdown responses shown in Fig. 4.4 by nonlinear regression even without using a prior term in the objective function.

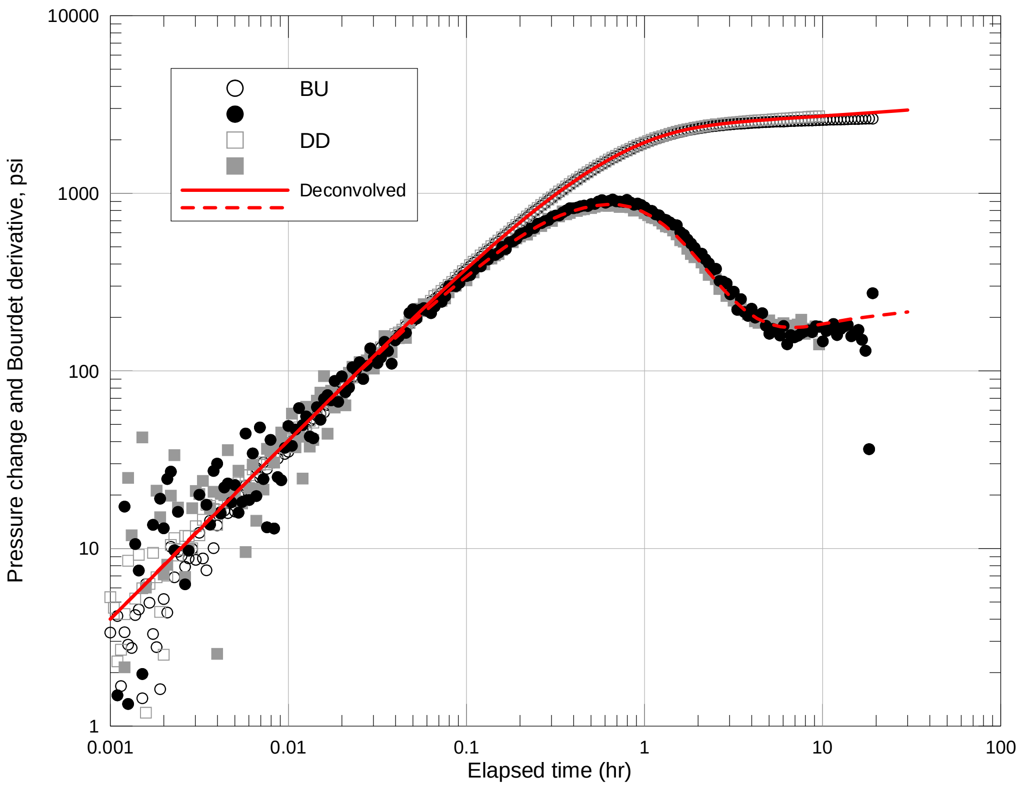

Figure 4.4: Comparison of deconvolved responses with the corresponding conventional responses for the BU and BU periods at the tested well.

The next task is to perform nonlinear regression analysis based on history matching of tested well pressure data in attempt to estimate the total of 4 well/reservoir parameters; *k*, *S*, C_w, *L*, by assuming that the values of ϕ, *h*, and c_t are known. It is well-known that it is not possible to uniquely determine the values of ϕ, *h*, and c_t from a single well test data set. In regression applications to be given next, it will be history matched only the buildup portion for the tested well, as this is typical application in real field test data applications, and use weighted least-squares estimation treating the data error variance (which is equal to 4 psi^2 in this test application) is known in history-matching. In regression applications, several different sets of the initial guesses is considered for the estimated parameters. For example, a set of initial guesses used is given in Table 4.2.

Table 4.2: Initial guesses, lower and upper constraint limits used for the parameters to be estimated by nonlinear regression.

Parameter	Initial Guess	Lower Limit	Upper Limit
k (md)	5	1	500
S (at the act. well, dimensionless)	1	-2	50
C_w (at the active well, B/psi)	1.0×10^{-2}	1.0×10^{-4}	1.0×10^{-1}
L (distance to the fault, ft)	500.	100	10000

First, a history matching application without a prior term in the objective function is considered for the parameters to be estimated. The lower and upper constraint limits for the parameters to be estimated are somewhat arbitrarily chosen as given in Table 4.2. It should be noted that it is used the imaging procedure of Carvalho et al. to keep the parameters to be estimated within their given lower and upper constraints (Table 4.2) during each iteration of the nonlinear regression optimization algorithm. This is necessary for example, to avoid permeability taking negative values, which halt the iteration procedure as the analytical solution used cannot accept negative permeability values during iteration. This non-linear regression application yielded the match of the entire pressure history as shown in Fig. 4.5 and the match of the buildup pressure change and its Bourdet derivatives as shown in Fig. 4.6, and the estimated values of the parameters as well as the value of Root-mean-Square (RMS) error for the pressure match are recorded in Table 4.3. As can be seen, the RMS for the match obtained is very close to the noise level (2.0 psi std.) in data, indicating that the model is had an acceptable match of the observed data with the corresponding model data as shown in Figs. 4.5 and 4.6. However, from Table 4.3, it is seen that all parameters except the distance to the fault (L) are determined well, and agree very well with the true, unknown values. Although not shown here, regression (without prior term in the objective function) applications was done with different sets of initial guesses of the parameters other than that given in Table 4.2 and the results pertaining to these applications showed that observed buildup pressure data well determine the values of the parameters k, S, and C_w, but not the distance to the fault L. In most of these regression applications, it was obtained that permeability ranging from 13 to 20 md, skin 3 to 5, the wellbore storage coefficient from 9.9×10^{-3} to 1.0×10^{-3} b/psi, and the distance to the fault widely ranging from 150 ft to 10000 ft.

So, next, it is investigated that whether adding a prior term in the objective function for the distance to the fault (L) provides a better means in nonlinear regression to obtain a reliable estimate of *L* that is consistent with the prior information and also provides a good match of the observed buildup pressure and derivative data. For this application, it will be assumed that the model has a priori information from geological and geophysical data indicating that the prior mean for the distance to the fault is 150 ft (which is the true, unknown value), but with an uncertainty (standard deviation) of 50 ft. Prior terms were not considered for the other parameters, i.e., infinite variance uncertainty was assumed for these parameters. Although different sets of initial guesses was considered for the parameters, the results is presented for the set of initial guesses considered in Table 4.2.

Table 4.3: Comparison of the values of parameters estimated from nonlinear regression application without prior term with the true values of the parameters.

Parameter	Estimated	True
k (md)	19.9	20
S (at the active well, dimensionless)	4.99	5
C_w (at the active well, B/psi)	1.0×10^{-2}	1.0×10^{-2}
L (distance to the fault, ft)	9849.4	150
RMS, psi	2.01	

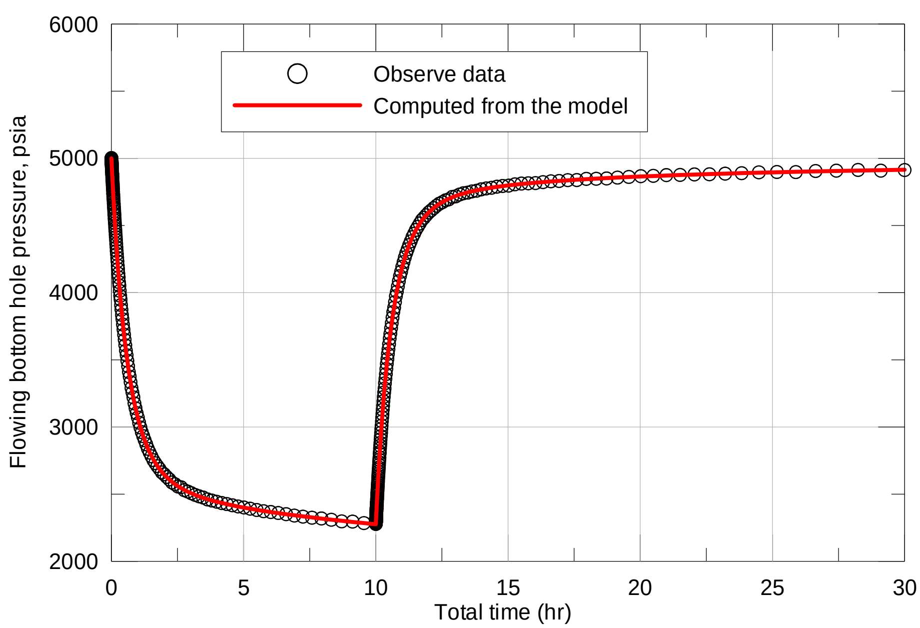

Figure 4.5: Match of the model pressures with the observed pressure data, nonlinear regression application without prior term in the objective function.

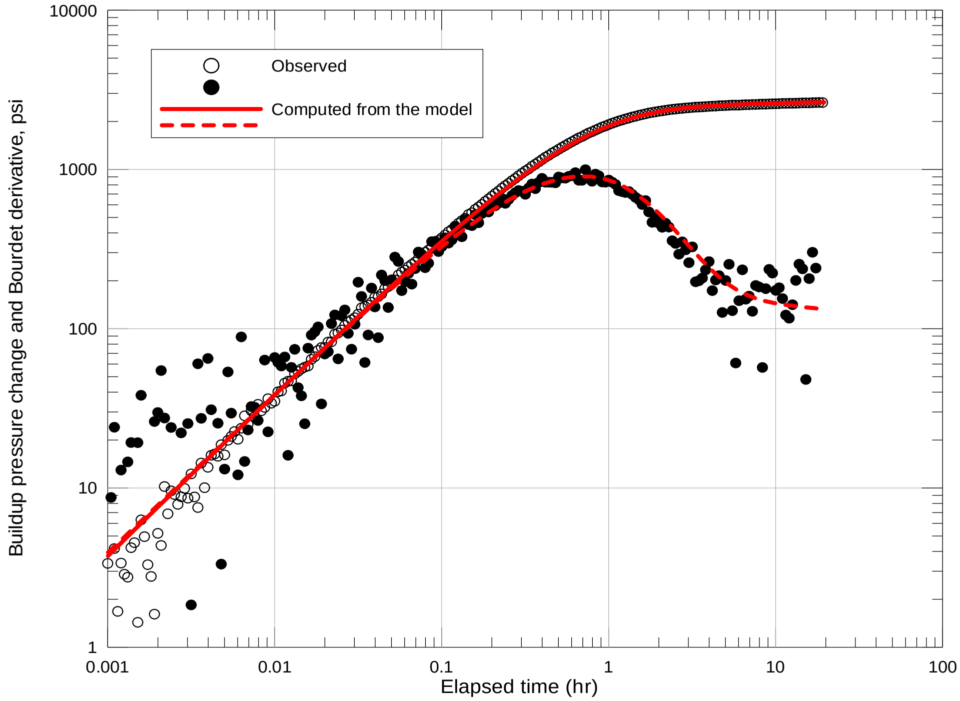

Figure 4.6: Match of the model buildup responses with the observed buildup responses, nonlinear regression application without prior term in the objective function.

This non-linear regression application (with a prior term having the correct mean for the distance to the fault) yielded the match of the entire pressure history as shown in Fig. 4.7 and the match of the buildup pressure change and its Bourdet derivatives as shown in Fig. 4.8, and the estimated values of the parameters as well as the value of Root-mean-Square (RMS) error for the pressure match are recorded in Table 4.4. As can be seen, the RMS for the match obtained is very close to the noise level (2.0 psi std.) in data, indicating that it is had an acceptable match of the observed data with the corresponding model data as shown in Figs. 4.7 and 4.8. The results of Table 4.4 indicate that that all parameters including the distance to the fault (L) are determined well as they agree very well with the true, unknown values. Although not shown here, regression (without prior term in the objective function) applications was performed with different sets of initial guesses of the parameters other than that given in Table 4.2 and the results pertaining to these applications showed that observed buildup pressure data well with a prior term having the correct mean for the distance to the fault determine well all the values of the parameters including the distance to the fault.

Table 4.4: Comparison of the values of parameters estimated from nonlinear regression application with a prior term for *l* with the true values of the parameters.

Parameter	Estimated	True
k(md)	19.9	20
S (at the active well, dimensionless)	4.99	5
C_w (at the active well, B/psi)	1.0×10^{-2}	1.0×10^{-2}
L (distance to the fault, ft)	150.6	150
RMS, psi	2.01	

Next, nonlinear regression is considered for a case where it is used an incorrect prior mean for the distance to the fault with a correction term for the distance to the fault in the objective function. Here, it is wanted to investigate whether it can be estimated the reliable estimates of the parameters including if it is had uncertainty in the prior mean of *L*. So, for this investigation, it will be assumed that it is had a priori information from geological and geophysical data indicating that the prior mean for the distance to the fault is 500 ft (which is radically different from the true, unknown value of 150 ft), but with an uncertainty (standard deviation) of 50 ft. Although somewhat arbitrarily chosen, it is considered that the standard deviation of the correction term (σ_{θ_L}) is as 500 ft. It was not considered prior terms for the other parameters, i.e., it was assumed infinite variance uncertainty for these parameters.

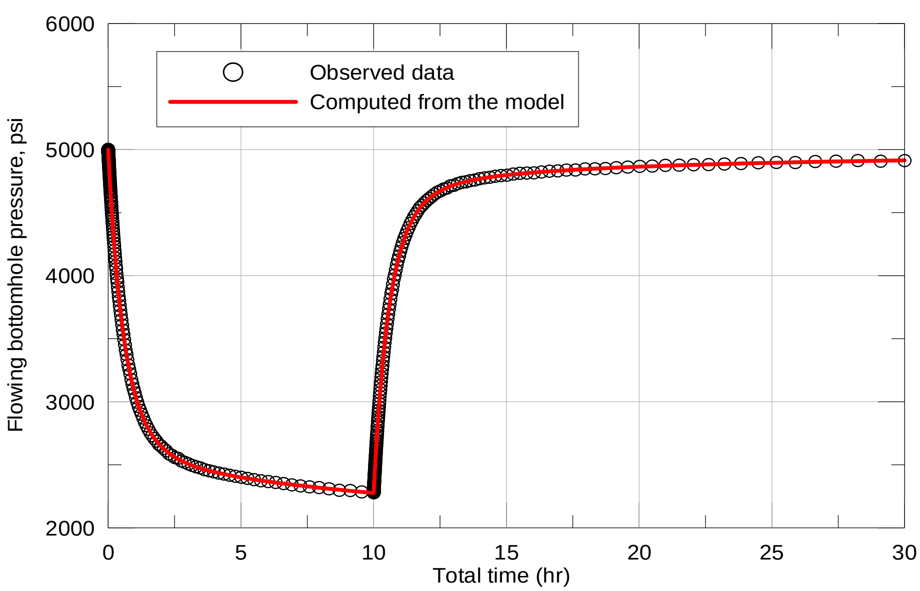

Figure 4.7: Match of the model pressures with the observed pressure data, nonlinear regression application with a prior term for *L* in the objective function.

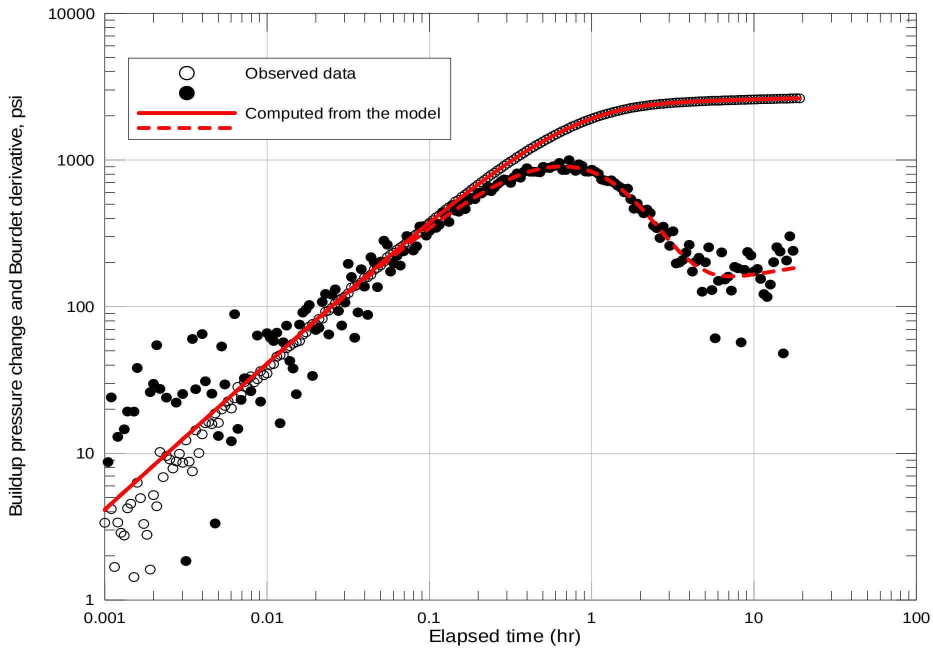

Figure 4.8: Match of the model buildup responses with the observed buildup responses, nonlinear regression application with a prior term for *L* in the objective function.

The initial guesses considered for the parameters are the same as those given in Table 4.2. The estimated values of the parameters as well as the value of Root-mean-Square (RMS) error for the pressure match for this application are recorded in Table 4.5. The matches of entire observed pressure data and buildup pressure data with the corresponding model responses were as shown in Figs. 4.7 and 4.8.

Table 4.5: Comparison of the values of parameters estimated from nonlinear regression application with a prior term for *l* with uncertainty in the prior mean and correction with the true values of the parameters.

Parameter	Estimated	True
k (md)	19.9	20
S (at the active well, dimensionless)	4.99	5
C_w (at the active well, B/psi)	1.0×10^{-2}	1.0×10^{-2}
L (distance to the fault, ft)	150.6	150
Correction term for L, θ_L	-346.	NA
RMS, psi	2.01	

As a final nonlinear regression application, nonlinear regression matching is considered of the equivalent constant-rate pressure change data reconstructed from the variable rate data by deconvolution. Those data were previously shown in Fig. 4.4. The objective of this exercise is to show that if we have good quality drawdown data (as those reconstructed by deconvolution procedure) that show the sufficient sensitivity to the all parameters of interest in the interpretation model chosen for the over-determined problem, it can be determined all the parameters reliably from the data itself without using a prior term with or without correction term. Nonlinear regression was performed of the deconvolved pressure data by considering the initial guesses given in Table 4.2 and obtained the results for the estimated values of the parameter as given in Table 4.6. The match of the deconvolved drawdown data with the model drawdown data is shown in Fig. 4.9.

Table 4.6: Comparison of the values of parameters estimated from nonlinear regression application with a prior term for *l* with uncertainty in the prior mean and correction with the true values of the parameters.

Parameter	Estimated	True
k (md)	20.	20
S (at the active well, dimensionless)	4.99	5
C_w (at the active well, B/psi)	1.0×10^{-2}	1.0×10^{-2}
L (distance to the fault, ft)	150.2	150
RMS, psi	0.6	

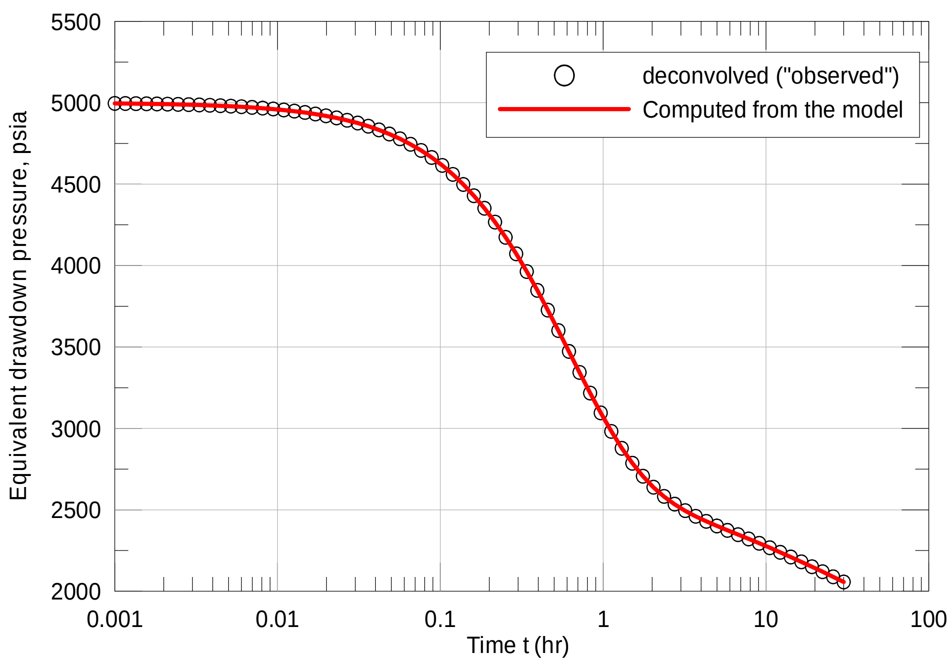

Figure 4.9: Match of the deconvolved equivalent drawdown data with the corresponding model data obtained by nonlinear regression.

5. CONCLUSIONS AND RECOMMENDATIONS

Within the concept of Bayesian framework, the prior information obtained from both static and dynamic data is added to the probability density function, and posterior pdf is obtained in order to reduce the uncertainty on sample.

In this work, the objective was to account the error on prior mean and reduce the uncertainty of the prior mean. Therefore, a new approach which is called partially doubly stochastic model, is considered and investigated on the over- and under-determined problems of reservoir characterization. This model provides to reduce uncertainty on prior mean by using a correction vector in posterior pdf. In this model, it is assumed that the prior mean is uncertain or incorrect but the covariance or variance of prior model is known.

The partially doubly stochastic model is applied to well-known parameter estimation methods, which are weighted least squares, unweighted least squares and maximum likelihood. A simple linear model is implemented for investigating these three parameter estimation methods. Besides, for underdetermined and overdetermined problems, it is applied WLS parameter estimation method and both static and dynamic realizations are obtained in order to estimate intended model parameters. The results obtained from all given problems are discussed, respectively.

5.1 Conclusions

1. In the partially doubly stochastic model, it seems that the variances of both the prior mean and correction term significantly affect the posterior estimates of the model parameters.
2. It was found for both over and under-determined problems that increasing the value of the variance of the correction (e.g, for the examples studies, a value of correction variance equal to 10 or larger) for given values of incorrect mean and the variance of the prior distribution improves the match with the true model.

3. In the case of true prior mean used in estimation, the estimated model parameters match closely to the true regardless of the input values of variances of prior mean and correction.
4. The results obtained with WLS and ML estimations indicate that ML estimation is useful in cases if the errors or variances in observed data are not known a priori.
5. It was shown that if one uses an incorrect prior mean in a estimation model that does not account for error in prior mean, then he/she will obtain incorrect posterior estimates of the model parameter. Using this incorrect posterior for generating conditional realizations (e.g., using GSlib) lead to incorrect realizations conditional to observed data. On the other hand, using the partially doubly stochastic model for such cases prevents one to generate inappropriate realizations of the parameters. Note that such realizations are used to assess the uncertainty in performance predictions.
6. For the case where we considered an overdetermined nonlinear problem of pressure transient test data (Chapter 4), we observed incorporating prior information, particularly when estimating model parameters that are not very sensitive to pressure data during the span of the test, is useful. In cases of incorrect prior mean chosen for such parameters, the partially doubly stochastic method seems to be useful not to lead an incorrect estimate of such model parameters.

5.2 Recommendations

1. In this thesis, it is assumed that the variance of prior mean is known. It is suggested that the model be estimated with the doubly stochastic method, which accounts for uncertainty in both the prior mean and variance. Thus, such method could be more flexible and realistic.
2. For underdetermined problems, a linear model is assumed in this work. It is recommended that the partially doubly stochastic model be considered for solving nonlinear models.
3. For both underdetermined and overdetermined problems, we mainly focused on WLS estimation to generate model parameters. However, in cases of errors in

observed data are uncertain, it is recommended that the maximum likelihood method be considered for estimation.

REFERENCES

Bard, Y. , 1974, *Nonlinear Parameter Estimation*. Academic Press. New York, USA.

Brummbert, A. C., Pool, S. E., Portman, M. E., Hancock, J.S., and Ammer, J. R., 1991, *Determining Optimum Estimation Methods for Interpolation and Extrapolation of Reservoir Properties: A Case Study, in Reservoir Characterization II*. Academic Press, Inc., 1991, California, USA.

Ceyhan, A. G., 1997, Jeoistatistiksel, Statik ve Kararsız Basınç Testi Verilerine Koşullandırılmış Heterojen Geçirgenlik ve Gözeneklilik Sahalarının Türetilmesi. *PhD Thesis,* The Technical University of Istanbul, Istanbul, Turkey.

Chu, L., Reynolds, A. C. and Oliver, D. S., 1995, Reservoir Description From Static And Well – Test Data Using Efficient Gradient Methods. SPE 2999, presented at the *International Meeting on Petroleum Engineering*, Beijing, China, November 14 -17.

Damsleth, E., 1994, Mixed Reservoir Characterization Methods. SPE 27969, presented at *The University of Tulsa Continental Petroleum Engineering Symposium*, Tulsa, USA.

Feller, W., 1966, *An Introduction to Probability Theory and Its Applications*. Wiley, Vol. **II**. New York, USA.

Fisher, R., 1950, *Contributions to mathematical statics*. Wiley, New York, USA.

Fletcher, R., 1991, *Practical Methods of Optimization*. John Wiley & Sons, 2nd edition, Chichester, UK.

He, N., 1997, Three Dimensional Reservoir Description by Inverse Problem Theory Using Well – Test Pressure and Geostatistical Data. *PhD Thesis,* The University of Tulsa, Tulsa, USA.

He, N., Oliver, D., and Reynolds, A. C., 2000, Conditioning Stochastic Reservoir Models to Well - Test Data. SPE 60687, presented at the *1997 SPE Annual Technical Conference and Exhibition*, San Antonio, Texas, USA, October 5- 8.

He, N., Reynolds, A. C., and Oliver, D., 1997, Three – Dimensional Reservoir Description From Multiwell Pressure Data and Prior Information. SPE 36509, presented at the *SPE Annual Technical Conference and Exhibition*, Colorado, USA, 6 – 9 October.

Hegstad, B. K., and Omre, H., 1998, Reservoir Characterization Integrating Well Observations, Seismic Data and Production History. *Norwegian University of Science and Technology Final Report*, Trondheim, Norway.

Holden, L., Omre, H., and Tjelmeland, H., 1992, Integrated Reservoir Description. SPE 24261, presented at the *SPE European Petroleum Computer Conference*. Stavanger, Norway.

Gavalas, G. R., Shah, P. C., and Seinfeld, J. H., 1976, Reservoir History Matching by Bayesian Estimation. SPE 5740, presented at the *SPE – AIME Fourth Symposium on Numerical Simulation of Reservoir Performance*. Los Angeles, USA, 19 – 20 February.

Gavin, H., 2010, CE281 Experimental Systems (The Levenberg – Marquardt Method For Nonlinear Least Squares Curve – Fitting Problems). Duke University, North Carolina, USA.

Kelkar, M. and Perez, G., 2002, *Applied Geostatistics for Reservoir Characterization*. Society of Petroleum Engineers Inc., Texsas, USA.

Kuchuk, F. J., Onur, M., and Hollaender, F., 2010, *Pressure Transient Formation and Well Testing: Convolution, Deconvolution and Nonlinear Estimation*. Vol. **57**, Elsevier, Oxford, UK.

Journel, A.G. and Huijbregts, C.J., 1978, *Mining Geostatistics*. The Blackburn Press, Caldwell, New Jersy (1978).

Levitan, M.M., 2005, Practical Application of Pressure/Rate Deconvolution to Analysis of Real Well Tests. SPE 84290, presented at the *2003 SPE Annual Technical Conference and Exhibition*, Denver, USA, 5 – 8 October.

Li, G., Han, M., Banerjee, R., and Reynolds, A.C., 2009, Integration of Well Test Pressure Data Into Heterogeneous Geological Reservoir Models. SPE 124055, presented at the *2009 SPE Annual Technical Conference and Exhibition*, Louisiana, USA, October 4 – 7.

Marquardt, D. W., 1963, An Algorithm for Least – Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*. Vol. **11**, No. 2, pp. 431 – 441. USA.

Menekşe, K., 1996, Doğal Çatlaklı Rezervuarlara Ait Kuyu Testi Verilerinin Doğrusal Olmayan Regresyon Yöntemleri İle Analizi. *PhD Thesis*, The Technical University of Istanbul, Istanbul, Turkey.

Olea, R. A., 1994, *Fundamentals of Semivariogram Estimation, Modeling, and Usage in Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies***.** American Association of Petroleum Geologists No: 3, pp. 27-36, Tulsa, USA.

Oliver, D. S., Cunha, L. B. and Reynolds, A. C., 1997, Markov Chains Monte Carlo Methods for Conditioning a Permeability Field to Pressure Data. *Mathematical Geology*. Vol. **29**, No.1, pp. 61 – 91. Heidelberg, Germany.

Oliver, D. S., Reynolds, A. C. and Liu, N., 2008: *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge University Press, Cambridge, UK.

Omre, H., Tjelmeland, H. and Wist, H. T., 1999, Uncertainty in History Matching–Model Specification and Sampling Algorithms. Norwegian University Science and Technology. Preprint Statistics No. 6/1999, Trondheim, Norway.

Onur, M., Çınar, M., Ilk, D., Valko, P. P., Blasingame, T. A., and Hegemean, P. S., 2008, An Investigation of Recent Deconvolution Methods for Well Test Data Analysis. SPE 102575, presented at the *2006 SPE Annual Technical Conference and Exhibition*, Texas, USA, 24 -27 September.

Onur, M., 2009, PET 604E Optimization Methods in Reservoir Engineering Class Notes. The Technical University of Istanbul, Istanbul, Turkey.

Onur, M., 2010, PET 467E Analysis of Well Pressure Tests Class Notes. The Technical University of Istanbul, Istanbul, Turkey.

Pimonov, E., Ayan, C., Onur, M., and Kuchuk, F. J., 2009, New Pressure/Rate Deconvolution Algorithm to Analyze Wireline Formation Tester and Well-Test Data. SPE 123982, presented at the *SPE Annual Technical Conference and Exhibition*, New Orleans, USA, 4 -7 October.

Press, W. H., Teukolsky, S. A, Vetterling, W. T. and Flannery, B. P., 1992, *Numerical Recipes in Fortran: The Art of Scientific Computing*. Cambridge University Press, 2nd edition. New York, USA.

Reynolds, A.C., He, N., and Oliver, D.S., 1999, *Reducing Uncertainty in Geostatistical Description with Well Testing Pressure Data, in Reservoir Characterization - Recent Advances*. American Association of Petroleum Geologists Memoir 71, pp. 149-162, Tulsa, USA.

Shah, P. C., Gavalas, G. R. and Seinfeld J. H., 1978, Error Analysis In History Matching: Level of Parameterization. SPE 6508, *Society of Petroleum Engineers Journal*. Vol. **18**, No. 3, pp. 219 – 228.

Tarantola, A, 2005: *Inverse Problem Theory and Methods for Model Parameter Estimation*. p.2, Siam, Philadelphia, USA.

Tjelmeland, H., Omre, H., and Hegstad, B.K., 1994, Sampling from Bayesian Models in Reservoir Characterization, *Technical Report Statistics*, No.2/1994, Norwegian Institute of Technology, Trondheim, Norway.

Url-1 < *http://en.wikipedia.org* >, accessed at 04.16.2011.

Url-2 < http://www.robertnowlan.com >, accessed at 05. 01.2011.

Url-3 < http://free-books-online.org >, accessed at 05. 01.2011.

Url-4 < http://www.gslib.com >, accessed at 05. 01.2011.

von Schroeter, T., Hollaender, F. and Gringarten, A.C., 2004, Deconvolution of Well Test Data as a Nonlinear Total Least-Squares Problem. SPE 77688, presented at the *2002 SPE Annual Technical Conference and Exhibition*, Texas, USA, 29 September – 2 October.

Zhang, F., Skjervheim, J. A., Reynolds, A. C., and Oliver, D. S., 2003, Automatic History Matching in a Bayesian Framework, Example Applications. SPE 106229-MS, presented at the *2003 SPE Annual Technical Conference and Exhibition*, Denver, USA, 5 – 8 October.

APPENDICES

APPENDIX A.1 : Vector - Matrix Calculus.

APPENDIX A.2: Cases for Synthetic Example.

APPENDIX A.1

Eqs. 2.24 and 2.25 are obtained using vector calculus based on the following formulas.

You will need to use the formula given by

$$\nabla_{\mathbf{x}}(\mathbf{BC}) = (\nabla_{\mathbf{x}}\mathbf{B})\mathbf{C} + (\nabla_{\mathbf{x}}\mathbf{C}^T)\mathbf{B}^T \quad \textbf{(A.1)}$$

where **B** is a (1xn) matrix (or also regarded as a row vector) with elements that are function of coordinates $x_1,x_2,\ldots,x_n$ and **C** is a (nx1) matrix (or also regarded as a column vector) with elements are function of coordinates $x_1,x_2,\ldots,x_n$.

Suppose we take the gradient of the scalar function f given by

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Ax} \quad \textbf{(A.2)}$$

Where, **A** is a constant square matrix.

Therefore, taking the gradient of f gives

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \nabla_{\mathbf{x}}\left(\frac{1}{2}\mathbf{x}^T\mathbf{Ax}\right) = \frac{1}{2}\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{Ax}) \quad \textbf{(A.3)}$$

Now, we need to use the formula given by

$$\nabla(\mathbf{BC}) = (\nabla\mathbf{B})\mathbf{C} + (\nabla\mathbf{C}^T)\mathbf{B}^T \quad \textbf{(A.4)}$$

Similarly, applying the formula given by Eq. A.4 to the term $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{Ax})$ in Eq. A.4 and letting $\mathbf{B} = \mathbf{x}^T$ and $\mathbf{C} = \mathbf{Ax}$ in Eq. A.4 gives

$$\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{Ax}) = (\nabla_{\mathbf{x}}\mathbf{x}^T)\mathbf{Ax} + \nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}^T)\mathbf{x} = \mathbf{Ax} + \mathbf{A}^T\mathbf{x} \quad \textbf{(A.5)}$$

but **A** is a symmetric matrix, i.e., $\mathbf{A}^T = \mathbf{A}$, then Eq. A.5 becomes

$$\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{Ax}) = 2\mathbf{Ax} \quad \textbf{(A.6)}$$

Now using Eq. A.6 in Eq. A.2 gives

$$\nabla_{\mathbf{x}} f(\boldsymbol{x}) = \mathbf{A}\mathbf{x} \tag{A.7}$$

Eq. 2.36 is obtained using vector calculus based on the following formulas.

First, the objective function Eq. A. 8 is differentiated by ***m***,

$$O(m) = \frac{1}{2}\sum_{j=1}^{K} N_{dj} \ln\left(\sum_{i=1}^{N_{dj}} \left[y_{ij} - f_{ij}(m)\right]^2\right) \tag{A.8}$$

$$\frac{\partial O}{\partial m} = -\sum_{j=1}^{K} \frac{N_{dj}}{\sum_{i=1}^{N_{dj}} \left[y_{ij} - f_{ij}(m)\right]^2} \sum_{i=1}^{N_{dj}} \left[y_{ij} - f_{ij}(m)\right] \frac{\partial f_{ij}}{\partial m} \tag{A.9}$$

Then the second derivation is applied to Eq. A.9 gives

$$\begin{aligned}
\frac{\partial}{\partial m}\left(\frac{\partial O}{\partial m}\right) = & -2\sum_{j=1}^{K} \frac{N_{dj}}{\left(\sum_{i=1}^{N_{dj}} \left[y_{ij} - f_{ij}(m)\right]^2\right)^2} \sum_{i=1}^{N_{dj}} \left[y_{ij} - f_{ij}(m)\right]\left(\frac{\partial f_{ij}}{\partial m}\right)\left[y_{ij} - f_{ij}(m)\right]\left(\frac{\partial f_{ij}}{\partial m}\right) \\
& + \sum_{j=1}^{K} \frac{N_{dj}}{\sum_{i=1}^{N_{dj}} \left[y_{ij} - f_{ij}(m)\right]^2} \sum_{i=1}^{N_{dj}} \left(\frac{\partial f_{ij}}{\partial m}\right)^2 \\
& + \sum_{j=1}^{K} \frac{N_{dj}}{\sum_{i=1}^{N_{dj}} \left[y_{ij} - f_{ij}(m)\right]^2} \sum_{i=1}^{N_{dj}} \left[y_{ij} - f_{ij}(m)\right]\left(\frac{\partial^2 f_{ij}}{\partial m^2}\right)
\end{aligned} \tag{A.10}$$

The vectors S and r are put in Eq. A.10, and then it gives

$$\begin{aligned}
\nabla_{\mathbf{m}}\left[\nabla_{\mathbf{m}} O(\hat{\mathbf{m}})\right]^T & = \nabla_{\mathbf{m}}\left[\sum_{j=1}^{K} -\frac{N_{dj}}{S_j}\left[\mathbf{y} - \mathbf{f}_j(\mathbf{m})\right]^T \mathbf{G}_j + \left(\mathbf{m} - \mathbf{m}_p - \boldsymbol{\theta}\right)^T \mathbf{C}_M^{-1}\right] \\
& = \mathbf{C}_M^{-1} + \sum_{j=1}^{K} \frac{N_{dj}}{S_j}\mathbf{G}_j^T\mathbf{G}_j - 2\sum_{j=1}^{K} \frac{N_{dj}}{S_j^2}\mathbf{G}_j^T\mathbf{r}_j\mathbf{r}_j^T\mathbf{G}_j + \sum_{j=1}^{K} \frac{N_{dj}}{S_j}\sum_{i=1}^{N_{dj}} r_{ij}\nabla\mathbf{G}_j^T
\end{aligned} \tag{A.11}$$

APPENDIX A.2

Some additional cases is applied to clarify the doubly stochastic (DS) estimation within MLE, WLS and UWLS. As it is mentioned before, in concept of this thesis prior means are assumed wrong and some of these cases show the behavior of model parameter when prior means are correct. These cases are shown below, respectively.

MLE:

Table A.2.1: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameters	$m_{1initial}$	$m_{2initial}$	$\theta_{1initial}$	$\theta_{1initial}$	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta1}^2$	$\sigma_{\theta2}^2$
Prior Model 1	10	5	0	0	10	5	0	0	0.1	0.1	0.001	0.001
Prior Model 2	10	5	0	0	10	5	0	0	1	1	0.001	0.001
Prior Model 3	10	5	0	0	10	5	0	0	10	10	0.001	0.001
Prior Model 4	10	5	0	0	10	5	0	0	0.1	0.1	0.1	0.1
Prior Model 5	10	5	0	0	10	5	0	0	1	1	0.1	0.1
Prior Model 6	10	5	0	0	10	5	0	0	0.1	0.1	1	1
Prior Model 7	10	5	0	0	10	5	0	0	1	1	1	1
Prior Model 8	10	5	0	0	10	5	0	0	10	10	1	1
Prior Model 9	10	5	0	0	10	5	0	0	100	100	1	1

Table A.2.2: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	σ_1^2	σ_2^2	σ_3^2	Obj_i	Obj_f
Prior Model 1	9.9461	5.0294	-0.001	0.000	4.0107	0.2507	0.00008	-14.45	-15.91
Prior Model 2	9.9454	5.0297	0.000	0.000	4.0108	0.2507	0.00008	-14.45	-15.93
Prior Model 3	9.9453	5.0298	0.000	0.000	4.0108	0.2507	0.00008	-14.45	-15.93
Prior Model 4	9.9457	5.0296	-0.027	0.015	4.0107	0.2507	0.00008	-14.45	-15.92
Prior Model 5	9.9454	5.0297	-0.005	0.003	4.0108	0.2507	0.00008	-14.45	-15.93
Prior Model 6	9.9454	5.0297	-0.050	0.027	4.0108	0.2507	0.00008	-14.45	-15.93
Prior Model 7	9.9454	5.0297	-0.027	0.015	4.0108	0.2507	0.00008	-14.45	-15.93
Prior Model 8	9.9453	5.0298	-0.005	0.003	4.0108	0.2507	0.00008	-14.45	-15.93
Prior Model 9	9.9453	5.0298	-0.001	0.000	4.0108	0.2507	0.00008	-14.45	-15.93

Table A.2.3: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameters	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 10	100	50	0	0	100	50	0	0	10	10	0.001	0.001
Prior Model 11	100	50	0	0	100	50	0	0	0.1	0.1	1	1
Prior Model 12	100	50	0	0	100	50	0	0	10	10	1	1
Prior Model 13	100	50	0	0	100	50	0	0	0.1	0.1	10	10
Prior Model 14	100	50	0	0	100	50	0	0	10	10	10	10
Prior Model 15	100	50	0	0	100	50	0	0	0.1	0.1	100	100
Prior Model 16	100	50	0	0	100	50	0	0	0.1	0.1	10000	10000
Prior Model 17	100	50	0	0	100	50	0	0	10	10	10000	10000

Table A.2.4: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	σ_1^2	σ_2^2	σ_3^2	Obj_i	Obj_f
Prior Model 10	97.005	46.637	0	0	10894	22648	28597	250	248.9
Prior Model 11	99.685	49.646	-0.286	-0.322	11757	24851	31555	250	249.8
Prior Model 12	96.688	46.282	-0.301	-0.338	10794	22395	28257	250	248.8
Prior Model 13	96.974	46.602	-2.996	-3.364	10884	22623	28563	250	248.9
Prior Model 14	93.646	42.883	-3.177	-3.559	9863	20038	25105	250	247.9
Prior Model 15	9.946	5.0294	-89.96	-44.93	4.011	0.251	8E-05	250	34.68
Prior Model 16	9.9453	5.0298	-90.05	-44.97	4.011	0.251	8E-05	250	-15.4
Prior Model 17	9.9453	5.0298	-89.97	-44.93	4.011	0.251	8E-05	250	-15.4

Table A.2.5: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameters	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 18	10	5	0	0	8	3	0.1	0.1	0.1	0.1	0.001	0.001
Prior Model 19	10	5	0	0	8	3	0.1	0.1	10	10	0.001	0.001
Prior Model 20	10	5	0	0	8	3	0.1	0.1	0.1	0.1	1	1
Prior Model 21	10	5	0	0	8	3	0.1	0.1	100	100	1	1
Prior Model 22	10	5	0	0	8	3	0.1	0.1	0.1	0.1	10	10
Prior Model 23	10	5	0	0	8	3	0.1	0.1	0.1	0.1	100	100

Table A.2.6: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	σ_1^2	σ_2^2	σ_3^2	Obj_i	Obj_f
Prior Model 18	9.9368	5.0342	0.118	0.119	4.0108	0.2507	0.00008	35.55	19.33
Prior Model 19	9.9452	5.0298	0.1	0.1	4.0108	0.2507	0.00008	-4.05	-15.58
Prior Model 20	9.9445	5.0302	1.777	1.855	4.0108	0.2507	0.00008	25.56	-12.69
Prior Model 21	9.9453	5.0298	0.118	0.119	4.0108	0.2507	0.00008	-14.4	-15.9
Prior Model 22	9.9452	5.0298	1.927	2.011	4.0108	0.2507	0.00008	25.55	-15.58
Prior Model 23	9.9453	5.0298	1.943	2.028	4.0108	0.2507	0.00008	25.55	-15.9

Table A.2.7: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameters	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 24	10	5	0	0	8	3	10	10	0.1	0.1	0.001	0.001
Prior Model 25	10	5	0	0	8	3	10	10	10	10	0.001	0.001
Prior Model 26	10	5	0	0	8	3	10	10	0.1	0.1	0.1	0.1
Prior Model 27	10	5	0	0	8	3	10	10	0.1	0.1	1	1
Prior Model 28	10	5	0	0	8	3	10	10	10	10	1	1

Table A.2.8: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	σ_1^2	σ_2^2	σ_3^2	Obj_i	Obj_f
Prior Model 24	9.9895	5.0065	9.921	9.921	4.0062	0.2513	0.00009	100026	618.93
Prior Model 25	9.9457	5.0295	9.999	9.999	4.0108	0.2507	0.00008	99986	-9.51
Prior Model 26	9.9659	5.0189	5.983	6.009	4.0108	0.2507	0.00008	1025.55	304.88
Prior Model 27	9.949	5.0278	2.681	2.753	4.0108	0.2507	0.00008	125.55	42.43
Prior Model 28	9.9457	5.0296	9.268	9.275	4.0108	0.2507	0.00008	85.95	-10.1

WLSE:

Table A.2.9: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameter	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 1	10	5	0	0	10	5	0	0	0.1	0.1	0.001	0.001
Prior Model 2	10	5	0	0	10	5	0	0	1	1	0.001	0.001
Prior Model 3	10	5	0	0	10	5	0	0	10	10	0.001	0.001
Prior Model 4	10	5	0	0	10	5	0	0	0.1	0.1	0.1	0.1
Prior Model 5	10	5	0	0	10	5	0	0	1	1	0.1	0.1
Prior Model 6	10	5	0	0	10	5	0	0	0.1	0.1	1	1
Prior Model 7	10	5	0	0	10	5	0	0	1	1	1	1
Prior Model 8	10	5	0	0	10	5	0	0	10	10	1	1
Prior Model 9	10	5	0	0	10	5	0	0	100	100	1	1
Prior Model 10	10	5	0	0	10	5	0	0	0.1	0.1	100	100
Prior Model 11	10	5	0	0	10	5	0	0	100	100	100	100
Prior Model 12	10	5	0	0	10	5	0	0	0.1	0.1	10000	10000

Table A.2.10: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	Tol.	İter.	Obj_i	Obj_f
Prior Model 1	9.9463	5.0293	-0.001	0.000	1E-10	2	19.052	17.782
Prior Model 2	9.9454	5.0297	0.000	0.000	1E-10	2	19.052	17.765
Prior Model 3	9.9454	5.0297	0.000	0.000	1E-10	2	19.052	17.763
Prior Model 4	9.9458	5.0295	-0.027	0.015	1E-10	2	19.052	17.772
Prior Model 5	9.9454	5.0297	-0.005	0.003	1E-10	2	19.052	17.765
Prior Model 6	9.9454	5.0297	-0.050	0.027	1E-10	4	19.052	17.765
Prior Model 7	9.9454	5.0297	-0.0273	0.0149	1E-10	3	19.052	17.764
Prior Model 8	9.9454	5.0297	-0.005	0.003	1E-10	2	19.052	17.763
Prior Model 9	9.9454	5.0298	-0.001	0.000	1E-10	2	19.052	17.763
Prior Model 10	9.945	5.030	-0.055	0.030	1E-10	3	19.052	17.763
Prior Model 11	9.9454	5.0298	-0.027	0.015	1E-10	3	19.052	17.763
Prior Model 12	9.9454	5.0298	-0.055	0.030	1E-10	4	19.052	17.763

Table A.2.11: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameter	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 13	10	5	0	0	8	3	10	10	0.1	0.1	0.001	0.001
Prior Model 14	10	5	0	0	8	3	10	10	10	10	0.001	0.001
Prior Model 15	10	5	0	0	8	3	10	10	0.1	0.1	0.1	0.1
Prior Model 16	10	5	0	0	8	3	10	10	0.1	0.1	1	1
Prior Model 17	10	5	0	0	8	3	10	10	10	10	1	1
Prior Model 18	10	5	0	0	8	3	10	10	0.1	0.1	10000	10000
Prior Model 19	10	5	0	0	8	3	10	10	1	1	10000	10000
Prior Model 20	10	5	0	0	8	3	10	10	10	10	10000	10000
Prior Model 21	10	5	0	0	8	3	10	10	100	100	10000	10000

Table A.2.12: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	Tol.	İter.	Obj_i	Obj_f
Prior Model 13	9.9952	5.0035	9.921	9.921	1E-10	7	100059	652.467
Prior Model 14	9.9459	5.0295	9.999	9.999	1E-10	6	100019	24.182
Prior Model 15	9.9707	5.0164	5.985	6.008	1E-10	3	1059	338.524
Prior Model 16	9.9500	5.0273	2.682	2.752	1E-10	4	159.052	76.119
Prior Model 17	9.9458	5.0295	9.268	9.275	1E-10	3	119.452	23.599
Prior Model 18	9.9454	5.0298	1.945	2.030	1E-10	3	59.062	17.769
Prior Model 19	9.9454	5.0298	1.946	2.031	1E-10	4	23.062	17.769
Prior Model 20	9.9454	5.0298	1.953	2.038	1E-10	3	19.462	17.769
Prior Model 21	9.9454	5.0298	2.025	2.109	1E-10	5	19.102	17.769

UWLS:

Table A.2.13: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameter	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta1}^2$	$\sigma_{\theta2}^2$
Prior Model 1	10	5	0	0	10	5	0	0	0.1	0.1	0.001	0.001
Prior Model 2	10	5	0	0	10	5	0	0	1	1	0.001	0.001
Prior Model 3	10	5	0	0	10	5	0	0	10	10	0.001	0.001
Prior Model 4	10	5	0	0	10	5	0	0	0.1	0.1	0.1	0.1
Prior Model 5	10	5	0	0	10	5	0	0	1	1	0.1	0.1
Prior Model 6	10	5	0	0	10	5	0	0	0.1	0.1	1	1
Prior Model 7	10	5	0	0	10	5	0	0	1	1	1	1
Prior Model 8	10	5	0	0	10	5	0	0	10	10	1	1
Prior Model 9	10	5	0	0	10	5	0	0	100	100	1	1
Prior Model 10	10	5	0	0	10	5	0	0	0.1	0.1	100	100
Prior Model 11	10	5	0	0	10	5	0	0	100	100	100	100
Prior Model 12	10	5	0	0	10	5	0	0	0.1	0.1	10000	10000

Table A.2.14: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	Tol.	İter.	Obj_i	Obj_f
Prior Model 1	10.0163	5.0745	0.000	0.001	1E-10	2	29.889	29.570
Prior Model 2	10.0023	5.0902	0.000	0.000	1E-10	5	29.889	29.540
Prior Model 3	9.9993	5.0928	0.000	0.000	1E-10	3	29.889	29.536
Prior Model 4	10.0109	5.0815	0.005	0.041	1E-10	4	29.889	29.554
Prior Model 5	10.0020	5.0904	0.000	0.008	1E-10	4	29.889	29.539
Prior Model 6	10.0020	5.0904	0.002	0.082	1E-10	3	29.889	29.539
Prior Model 7	10.0007	5.0916	0.0003	0.0458	1E-10	3	29.889	29.538
Prior Model 8	9.9993	5.0928	0.000	0.008	1E-10	4	29.889	29.536
Prior Model 9	9.9990	5.0931	0.000	0.001	1E-10	4	29.889	29.536
Prior Model 10	9.999	5.093	-0.001	0.093	1E-10	5	29.889	29.536
Prior Model 11	9.9990	5.0931	-0.001	0.047	1E-10	3	29.889	29.535
Prior Model 12	9.9989	5.0931	-0.001	0.093	1E-10	5	29.889	29.535

Table A.2.15: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameter	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{pr1}	θ_{pr2}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta1}^2$	$\sigma_{\theta2}^2$
Prior Model 13	10	5	0	0	8	3	0.1	0.1	0.1	0.1	0.001	0.001
Prior Model 14	10	5	0	0	8	3	0.1	0.1	10	10	0.001	0.001
Prior Model 15	10	5	0	0	8	3	0.1	0.1	0.1	0.1	0.1	0.1
Prior Model 16	10	5	0	0	8	3	0.1	0.1	0.1	0.1	1	1
Prior Model 17	10	5	0	0	8	3	0.1	0.1	10	10	1	1
Prior Model 18	10	5	0	0	8	3	0.1	0.1	100	100	1	1
Prior Model 19	10	5	0	0	8	3	0.1	0.1	0.1	0.1	10	10
Prior Model 20	10	5	0	0	8	3	0.1	0.1	0.1	0.1	100	100
Prior Model 21	10	5	0	0	8	3	0.1	0.1	0.1	0.1	10000	10000

Prior Model 22	10	5	0	0	8	3	0.1	0.1	100	100	10000	10000

Table A.2.16: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	Tol.	İter.	Obj_i	Obj_f
Prior Model 13	9.6667	5.0447	0.116	0.119	1E-10	4	79.889	63.452
Prior Model 14	9.9935	5.0937	0.100	0.100	1E-10	4	40.289	29.914
Prior Model 15	9.7948	5.0864	0.947	1.093	1E-10	4	69.989	47.479
Prior Model 16	9.9523	5.0971	1.784	1.916	1E-10	5	69.899	32.944
Prior Model 17	9.9940	5.0937	0.272	0.281	1E-10	4	30.299	29.880
Prior Model 18	9.9984	5.0932	0.119	0.120	1E-10	5	29.939	29.573
Prior Model 19	9.9936	5.0937	1.975	2.074	1E-10	8	69.890	29.910
Prior Model 20	9.9984	5.0932	1.997	2.091	1E-10	4	69.889	29.573
Prior Model 21	9.9989	5.0931	1.999	2.093	1E-10	5	69.889	29.536
Prior Model 22	9.9989	5.0931	1.980	2.073	1E-10	4	29.929	29.536

Table A.2.17: The initial values of model parameters and corrections, the values of prior means, correction means, prior and correction variances.

Parameter	m_{1i}	m_{2i}	θ_{1i}	θ_{1i}	m_{pr1}	m_{pr2}	θ_{01}	θ_{02}	σ_{m1}^2	σ_{m2}^2	$\sigma_{\theta 1}^2$	$\sigma_{\theta 2}^2$
Prior Model 23	10	5	0	0	8	3	10	10	0.1	0.1	0.001	0.001
Prior Model 24	10	5	0	0	8	3	10	10	10	10	0.001	0.001
Prior Model 25	10	5	0	0	8	3	10	10	0.1	0.1	0.1	0.1
Prior Model 26	10	5	0	0	8	3	10	10	0.1	0.1	1	1
Prior Model 27	10	5	0	0	8	3	10	10	10	10	1	1
Prior Model 28	10	5	0	0	8	3	10	10	0.1	0.1	10000	10000
Prior Model 29	10	5	0	0	8	3	10	10	1	1	10000	10000
Prior Model 30	10	5	0	0	8	3	10	10	10	10	10000	10000

Table A.2.18: The results are obtained from program, respectively, model parameters, corrections, variances and minimized objective functions.

Parameters	m_1	m_2	θ_1	θ_2	Tol.	İter.	Obj_i	Obj_f
Prior Model 23	11.4882	5.1999	9.936	9.923	1E-10	16	100070	592.786
Prior Model 24	10.0237	5.0889	9.999	9.999	1E-10	3	100030	35.853
Prior Model 25	10.9209	5.0609	6.460	6.030	1E-10	7	1070	328.072
Prior Model 26	10.2113	5.0623	2.919	2.784	1E-10	8	169.889	86.390
Prior Model 27	10.0215	5.0893	9.275	9.281	1E-10	3	130.289	35.280
Prior Model 28	9.9990	5.0931	1.999	2.093	1E-10	4	69.899	29.542
Prior Model 29	9.9990	5.0931	2.000	2.094	1E-10	3	33.899	29.542
Prior Model 30	9.9990	5.0931	2.007	2.101	1E-10	3	30.299	29.542

CURRICULUM VITAE

Melek Deniz

06 Mayıs 2011

She was born 13 April 1983 in Kayseri. *She* completed Kdz. Ereğli Foreign Language Oriented High school between 1999 and 2002. After high school education, she joined Istanbul Technical University Petroleum and Natural Gas Engineering Department in 2002-2007. After that step she worked as a reservoir simulation engineer during one year at Binagadi Oil Co. in Baku. She has decide to move from the Istanbul Technical University Science Engineering and Technology, Petroleum and Natural Gas Engineering Department as a research assistant and student since 2008. She has a publication by the name of "Importance of Oil Composition in Oil/Brine/Rock System and Its Effect on Waterflooding" , presented at the 16th International Petroleum and Natural Gas Congress and Exhibition of Turkey, Ankara, May 29 - 31, 2007