THE UNIVERSITY OF TULSA

THE GRADUATE SCHOOL

COMPARISON OF SAMPLING METHODS FOR UNCERTAINTY EVALUATION IN RESERVOIR FLOW PREDICTIONS

by

Soraya Sofia Betancourt Pocaterra

A thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science
in the Discipline of Petroleum Engineering
The Graduate School
The University of Tulsa
2000

THE UNIVERSITY OF TULSA

THE GRADUATE SCHOOL

COMPARISON OF SAMPLING METHODS FOR UNCERTAINTY EVALUATION IN RESERVOIR FLOW PREDICTIONS

by

Soraya Sofia Betancourt Pocaterra

A Thesis Approved for the Discipline of Petroleum Engineering

By Thesis Committee

_____, Chairperson
Dr. Dean S. Oliver

_____
Dr. Albert C. Reynolds

_____
Dr. William A. Coberly

A B S T R A C T

Betancourt Pocaterra, Soraya Sofia (Master of Science in Petroleum Engineering)
Comparison of Sampling Methods for Uncertainty Evaluation in Reservoir Flow Predictions (104 pp.- Chapter V)
Directed by Dr. Dean S. Oliver

(143 words)

A variety of methods for generating realizations conditional to production data for the reservoir characterization problem is available, and several studies have attempted to determine the most suitable technique. This study focuses on the application of five sampling algorithms to a synthetic, one-dimensional, single-phase flow problem, in order to establish the best algorithm under controlled conditions. A small test problem was chosen for this study in order to ensure that a large enough number of realizations could be generated from each method for statistical validity. Several thousand realizations from each sampling algorithm were generated to attempt the characterization of the probability density function. The methods considered were Linearization about the MAP (LMAP), Randomized Maximum Likelihood (RML), Pilot Point (PP), Markov Chain Monte Carlo (MCMC) and Rejection Algorithm (REJ). The distributions of the realizations were compared to evaluate the validity and usefulness of the methods.

A C K N O W L E D G E M E N T S

I want to express my gratitude to my advisor Dr. Dean Oliver for his patience during the elaboration of this project. I am thankful for the opportunity to work with him.

I want to recognize the job of all my professors at The University of Tulsa, for their valuable teachings. Especially, I would like to thank Dr. Albert Reynolds and Dr. Stefan Miska for their dedication and encouragement.

My sincere appreciation goes to the committee members Dr. Coberly, Dr. Reynolds and Dr. Oliver for taking time to evaluate this work and for their constructive recommendations that improved considerably the quality of this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

The economic viability of an oilfield development project is greatly influenced by the reservoir production performance under the current and future operating conditions. In order to analyze the reservoir performance and estimate reserves, engineers make use of numerical flow simulators that require a parameterization of the reservoir properties, i.e., a reservoir model, as input. The accuracy of the results obtained is determined by the quality of the reservoir model used to make the reservoir performance analysis.

The hydrocarbon reservoir is a physical system from which there is only available a limited amount of direct information. The main source of information from the system may come in the form of production data, such as production rates and pressure behavior, all of them indirect measurements of the physical parameters that describe the fluid flow through the reservoir. The goal is to obtain a map of the reservoir parameters to be used as an input in the flow simulator in order to describe satisfactorily the production performance.

A viable approach for the characterization of the physical parameters of a reservoir is through the use of inverse theory. In inverse problems, indirect measurements from a physical system are used to make inferences of that system. Since all measurements are

subject to error and the number of reservoir parameters is usually larger than the available data, the solution is non-unique and the inverse problem is ill posed. The non-uniqueness of the solution implies that there are an infinite number of reservoir models that honor the production history and the prior knowledge of the reservoir. This means that rather than generating a single model, it is more meaningful to generate a set of plausible realizations in order to evaluate the uncertainty in the reservoir performance.

There are several methods to sample; the a posteriori probability density function (pdf) of the reservoir model. The purpose of this study is to evaluate five of these methods on a fair basis. The same preliminary assumptions, such as discretization of the problem and geostatistical considerations, were used for all methods in this study in order to focus the analysis on the performance of the sampling algorithms. The objective is to investigate the reliability of the different methods to assess uncertainty in flow predictions.

The approximate sampling methods evaluated here belong to two types (Oliver et. al., 1996): those that add roughness to a smooth estimate and those that add a smooth correction to variable fields of properties. In the first category, the Linearization about the MAP (LMAP) method will be considered. Randomized Maximum Likelihood (RML), and the Pilot Point Method (PP) make a correction to an unconditional realization. Markov Chain Monte Carlo (MCMC) and Rejection (REJ) methods sample from the correct distribution, but may be too slow in practice. All these methods were

used to generate a large number of realizations from a single-phase, one-dimensional synthetic problem, where the observed data were in the form of dynamic pressure.

CHAPTER II

LITERATURE REVIEW

In reservoir characterization using inverse techniques, the objective is to create a mathematical model for the reservoir incorporating all possible information, and treating the uncertainty in reservoir properties as if the properties were random variables. The mathematical model should be based on clearly formulated assumptions, such as that it is possible to parameterize the reservoir by dividing it into segments, layers or facies. For the case of facies, the geometry could be parameterized as objects. We often assume that the porosity is represented by a Normal distribution within a single facies, and that the permeability is log-Normal. The mathematical model is chosen based on the understanding of the reservoir and the available data.

Reservoir models are required to satisfy both our previous knowledge of the system, as well as honoring the indirect and direct measurements from it. This is done using Bayesian techniques in the conditioning. The prior model is formulated using geological and geophysical information and a likelihood for the data. In our case, typical data could be well observations, seismic and production data. It is important to include the uncertainty of the most relevant parameters in the evaluation.

Since different reservoir models could reproduce the same observed data, we approach the problem in a probabilistic frame. Given the importance of the decisions based on reservoir studies, multiple realizations of the reservoir must be analyzed in order to assess the uncertainty in reservoir performance. The realizations from the model are generated by simulation algorithms. Typical simulation algorithms are MCMC methods, simulated annealing, sequential methods and FFT methods.

To the best of our knowledge, only three studies have addressed the validity of various sampling algorithms to generate realizations of the reservoir model conditional to flow data. Two of these studies were performed on synthetic cases based on real reservoirs and involved the participation of several research groups and different approaches for the analysis of the same problem.

The first study represents the first major attempt to compare inverse approaches and was performed in the area of groundwater hydrology (Zimmerman et. al. 1998). This study compared seven different stochastic inverse techniques for identifying aquifer transmissivity. The objective was to determine which of those techniques was the most appropriate for predicting the outcome of a solute transport problem in an aquifer on a probabilistic frame. The aforementioned aquifer was expected to safely contain a radioactive waste for a long period of time. The main challenge was the lack of

knowledge about the geology of the aquifer, which was suspected to be highly heterogeneous.

The seven inverse methods were compared on four synthetic data sets. The comparison criteria were the predicted travel times and the travel paths followed by conservative tracers over a distance of 5 km. The methods considered were as follows.

1. Fast Fourier Transform (FF). (Gutjahr and Wilson, 1989; Robin et. al., 1993; Gutjahr et. al., 1994). This method uses the Fast Fourier Transform Technique for the generation of parameter fields. Time dependent data were not considered. For the conditioning of the generated models to observed transmissivities and head measurements, an iterative co-kriging procedure was implemented. The FFT is very computationally efficient for generating unconditional realizations on a large grid.
2. Linearized Semi-Analytical Method (LS). (Dagan, 1985; Rubin and Dagan, 1987; Dagan and Rubin, 1988; Rubin 1991 a,b; Rubin and Dagan, 1992). This technique involved two stages: the solution of the inverse problem and secondly the solution of the transport problem using particle tracking. A maximum likelihood procedure with cokriging is used to generate head and transmissivity fields that are conditioned to observed data. This method does not require the numerical solution of flow equations, and therefore discretization errors are

avoided. The main limitation is the assumption that head measurements are linearly related to transmissivity.

3. Linearized Cokriging Method (LC). (Kitanidis and Vomvoris, 1983; Hoeksema and Kitanidis, 1984; Kitanidis and Lane, 1985). This technique uses maximum likelihood for the estimation of the structural parameters associated with the log-transmissivity covariance based on the two types of observed data (transmissivity and head). The method requires the direct inversion of a matrix equal to the number of gridblocks used in the flow simulator (the code used for this study, GEOINVS, was limited to 1600 gridblocks).
4. Fractal Simulation (FS). (Grindrod and Impey, 1991). The fractal parameters, *a* (amplitude) and *p* (phase), are fitted using maximum likelihood estimation and realizations (fractal fields) are generated using the Fast Fourier Transform method. A linear superposition of the unconditioned fields is used to condition them to the observed data, minimizing the difference between the variance of the final field and the data.
5. Pilot Point Method (PP). (Rama Rao et. al., 1995; La Venue et. al., 1995). After modeling the data variograms, unconditional realizations of transmissivity fields are generated using the turning bands method. The measured transmissivities are honored by adding a simulated kriging error based on the observed data. The calibration of the computed pressure to the observed pressures is done automatically through the minimization of an objective function.

6. Maximum Likelihood Method (ML). (Carrera and Neuman, 1986 a,b). This is a very general non-linear technique that estimates aquifer parameters (transmissivity, recharge, storage, leakage coefficients, heads or flow rates) using prior estimates of their values along with transient or steady-state head measurements. The non-linear flow equation was solved using a fully implicit finite-element flow simulator. The optimization part was performed by minimizing an objective function consisting of a likelihood term for the head data, and a prior estimate term with a weighting parameter for the other hydrologic parameters. Several minimizations were done for different weighting parameters.
7. Sequential Self Calibration (SS). (Gomez-Hernandez et. al., 1997). The transmissivity data are kriged and the kriging standard deviation is computed at each gridblock location. A grid, located in the mean flow direction, is constructed and a seed transmissivity field is generated according to the random function model chosen (multi-Gaussian or not) and conditional to observed data. A seed field that reproduces the data is determined by optimization and realizations are generated by perturbation of this seed field. The perturbation field is parameterized by a few values at selected master locations. Perturbation of the remaining cells is obtained by kriging interpolation of the master location values.

The information given to the participants consisted of steady-state hydraulic head data and transmissivity values at a few locations, and transient information in the form of

three independent aquifer tests. To evaluate the performance of the methods, the cumulative distribution functions (cdf) were compared with the "true" cdf of the field for ten quantitative evaluation measures. These measures included particle travel time from distinct locations, a measure of the reproduction of the real transmissivity field, a comparison of the transmissivity variograms and a measure of the reproduction of the real head field.

The best approaches for this study were selected among those techniques that showed a reduction in the uncertainty compared to unconditional realizations. Four algorithms satisfied this criterion for the problem analyzed: Linearized Semianalytical, Maximum Likelihood, Pilot Point, and Sequential Self Calibration. The main conclusion achieved on this study was the importance of the appropriate selection of the variogram and the time and experience devoted by the user of the method in analyzing and modeling the observed data. The construction of cross-variograms was found to be very useful.

Another interesting conclusion from this study is the importance of the identification of the appropriate structure of the transmissivity field, i.e. the spatial covariance for this parameter field. This was found to be more relevant than the estimation of the parameter values, since completely unrealistic models can reproduce the data satisfactorily. The participants of this study did not reach any final conclusion regarding the effects of the discretization of the problem (the SS method used a coarse grid and performed

satisfactorily). The choice of the grid is closely related to the scaling of parameters and domain of the measurements.

The authors of this groundwater study recommended the creation of a separate comparison committee, distinct from the participants, for future similar studies, to apply several tests to the algorithms without knowing the output of the preceding test, and to diversify the synthetic cases in order to reproduce more realistic situations. They also suggest a careful design of the comparison criteria prior the application of the algorithms to ensure that all the tests are suitable for all the methods.

The second study had as objective to evaluate several uncertainty quantification methods for production forecast in a hydrocarbon reservoir (Floris et. al. 1999). The synthetic reservoir model was based on a real field example. Similarly to the previous study, the participants received reservoir parameters (porosities and permeabilities) at well locations and 'historic' production data –all of them with noise. A general geologic description of the reservoir was also available. The production data were supplied for the first 8 years of field activity and participants were asked to extend their forecasts to 16.5 years (the objective was to compare the forecasts with the real performance over this period of time). Nine different techniques were evaluated having as a requirement the conditioning of the reservoir models to the production data, and results (production forecast for a certain period) were requested in the form of a cumulative distribution function (cdf).

Besides evaluating different methodologies for uncertainty quantification, distinct parameterization approaches and optimization procedures were applied by the participants. Parameterization deals with the spatial distribution of the reservoir properties (to be input in the flow simulator). The approaches used were grid block discretization, regional parameterization (geological layers, genetic units or draining areas) and pilot points. Optimization procedures included Gradient Optimization and Genetic Algorithms. The uncertainty quantification methods evaluated in this study were as follows.

1) Local Characterization of the objective function around the Maximum A Posteriori (MAP) solution. After minimizing an objective function with a likelihood term (data honoring) and a prior term, realizations are generated using the covariance matrix of the a posteriori distribution function. It is expected that the pdf will be characterized only around the MAP solution.
2) Local Characterization of the objective function around a Maximum Likelihood Solution. This method is similar to the above, with the difference that the objective function only contains the likelihood term.
3) Local Characterization of the objective function around several Maximum A Posteriori solutions (if the objective function is multimodal).
4) Local Characterization of the objective function around several Maximum Likelihood solutions (if the objective function is multimodal).

5) Randomized Maximum Likelihood. Unconditional realizations of the prior model are generated. This is followed by the minimization of an objective function containing the unconditional prior term and a likelihood term. In a variation of this method, Oliver et al. (1996) suggest the generation of unconditional realizations of the data to be used in the likelihood term, i.e. data sampling.
6) Markov chain Monte Carlo with local perturbation of the models.

The authors of the second study identified the variation in the parameterization of the problem as the main discriminating factor. The differences in the quality of the history matching and the production forecast caused by the distinct approaches to the problem, lead to major differences in the resulting cumulative distribution functions. Other aspects investigated were the use of prior sampling from the reservoir model, production sampling as introduced by Oliver et al. (1996), and the use of the quantification of extremes as proposed by the Scenario Test Method.

Approaches that used zonation to represent the parameter fields had large uncertainty ranges. These methods also had poor history matches and therefore, the production forecasts were biased. The comparison of the predictions and the true case showed that for five of the methods the real reservoir performance was not included in the cdf of the forecasts. This confirmed the hypothesis that uncertainties are often underestimated, and

reinforced the importance of the history matching process and the adequate modeling of the reservoir heterogeneities.

The sampling of the prior geological reservoir model was found to affect the extent of the uncertainty range. For the problem being analyzed, the data sampling (the likelihood term) was observed to have little contribution to the uncertainty range, and the quantification of extreme forecasts showed a wider range of uncertainty compared to other heterogeneous curves.

The third study focused on the comparison of a Markov Chain Monte Carlo algorithm with two approximate sampling methods (Omre et. al. 2000). The approximate methods considered were the Kitanidis-Oliver algorithm (Randomized Maximum Likelihood Method), and a reduced version of the Kitanidis-Oliver algorithm, (Pilot Point Method). The MCMC method was considered the benchmark for comparison purposes, since it can be shown to generate samples from the correct posterior pdf.

In order to obtain realizations of porosity, log-permeability and saturation, the algorithms were applied to a synthetic two-dimensional, two-phase problem. The porosity was considered constant throughout the reservoir.

For each algorithm nine independent samples were generated, and for comparison purposes histograms were constructed to show the distribution of the reservoir properties for each realization. The reduced algorithm was applied using three different spatial patterns of pilot points (18%, 6% and 2% of the gridblocks). The results showed that the RML method reproduces the average value of the reservoir properties, but the exact pdf was not obtained for the particular case considered. The results from PPM and RML were found to be similar, especially for a large number of pilot points (18%). For the second pattern (6%), results were also comparable but more heterogeneous. For the third pattern (2%), the effect of the pilot points in the properties map was noticeable. Even though the optimization procedure ensured the consistency of the models with the observed data, the realizations from the approximate methods were more heterogeneous and showed little resemblance to the real reservoir.

A second case was evaluated with an increase in the variance of the errors in the observed data. It was observed that the sampled models and the true models were more much alike in this case. The explanation for this was attributed to the weaker constraints in the likelihood term.

The authors recognized the MCMC method to be very resource demanding and recommend a careful analysis of the problem, especially for highly non-linear problems, before considering the use of the approximation algorithms.

CHAPTER III

INVERSE PROBLEM THEORY FOR UNCERTAIN DATA

Probability Theory

Probability is fundamentally about measuring sets. The sets can be finite as in the possible outcomes of a roulette or infinite as the possible distributions of porosity and permeability in a reservoir. The set of all possible outcomes of an experiment is known as the sample space. If A is an event or outcome of an experiment having a sample space S, then P(A) represents the probability of the event A. Probabilities are always positive and between 0 and 1 (the probability is zero for impossible events), and P(S) is equal to one (the probability of the sum of all the possible events in the sample space is equal to one). Probabilities also satisfy the additive property on mutually independent events.

Random Variables

Random variables are used to denote the possible outcomes of random trials. A given outcome of a random trial is called a realization. For example, a set of reservoir parameters (porosity, permeability, and/or saturation) that meets a certain requirement is called a realization.

"A random phenomenon is an empirical phenomenon characterized by the property that its observation under a given set of circumstances does not always lead to the same observed outcomes (so that there is no statistic regularity) but rather to different

outcomes in such a way that there is statistical regularity. By this is meant that a number exists between 0 and 1 that represents the relative frequency with which the different possible outcomes may be observed in a series of observations of independent occurrences of the phenomenon. … A random event is one whose relative frequency of occurrence, in a very long sequence of observations of randomly selected situations in which the event may occur, approaches a stable limit value as the number of observations is increased to infinity; the limit value of the relative frequency is called the probability of the random event"[1].

Bayes Theorem

The conditional probability P(B|A) (read probability that the event B will occur given that the event A has occurred), is the ratio of the probability of the intersection of these events with respect to probability of event A.

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Since the intersection of the event A and B is the same as the intersection of the events B and A, then:

$$P(A \cup B) = P(B|A)\ P(A) = P(BA) = P(A|B)\ P(B)\ .$$

Bayes Theorem is widely used in stochastic inverse theory, where the objective is to generate plausible realizations of a physical system conditioned to indirect measurements

[1] Parzen, E., Modern probability theory and its applications, Wiley 1960.

or to a prior knowledge of what the system might look like.

Probability Functions And Densities

For discrete random variables, the probability that the variable X will take the value of x, is the probability mass function p(x):

$$p(x) = P\{X = x\}\,.$$

It was stated before that the summation of the probabilities of all the possible outcomes of a discrete random process is equal to one. For random variables that can take values within an interval, i.e., continuous random variables, there is a nonnegative function called the probability density function f(x) such that the probability that the random variable X is contained within the interval C can be written as:

$$P\{X \in C\} = \int_C f(x)dx\,.$$

Expectation of a Function With a Probability Law

When dealing with sets of reservoir parameters, each parameter is considered a variable, and therefore we must consider multi-variate probabilities. Fortunately, any one dimensional distribution function defined on the R^1 space can be generalized to two (R^2) or more dimensions (R^N). If the outcomes of the random phenomena are real numbers, the probability law that defines them could be described as the distribution of a unit mass along the real line for R^1 spaces. This definition can be extended to vector fields and

functions. For the random phenomena P, having a probability density p(x) associated, the expectation of a function f(x) with respect to P is

$$E[f(x)] = \int_{-\infty}^{\infty} f(x)\, p(x)\, dx .$$

The mean of the probability P is the expectation of x is

$$\bar{x} \equiv E[X] = \int_{-\infty}^{\infty} x\, p(x)\, dx .$$

The variance of the probability P is defined as:

$$V(X) = E\left[(X - E(X))^2\right] = E\left[(X - \bar{x})^2\right] .$$

Central Limit Theorem

The sum of a large number of independent, identically distributed random variables, all with finite means and variances, is approximately normally distributed. This theorem is important for the following developments because the types of errors considered in stochastic inverse theory, errors of measurement and observation, are often adequately represented by a Normal distribution.

Variograms

To describe the spatial correlation between physical parameters in the reservoir we use variograms. The variogram is a measure of the similarity between properties (the co-variance between two values) that are located a distance h apart. The estimate of variance is repeated for many values of h to represent the spatial correlation of the

property being analyzed. This similarity measure is called γ (h), and it is plotted on an x-y plot with the x- axis being the distance h, and γ (h) on the y- axis.

If the variogram rises, and then levels off or stabilizes around some value, it is said to have reached a sill. This is theoretically the variance of the sample. The distance at which the rising variogram reaches the sill is called the range, and is symbolized by *a*. The range is the distance at which the covariance becomes zero, so it marks the limit of the zone of influence of a single sample. Beyond the range, samples are no longer correlated and are independent.

There are several types of variograms. For this study, the variogram used was the exponential model. The exponential model is a model that stabilizes around a sill and has a finite variance and covariance. This variogram is linear at the origin, but reaches the sill asymptotically, well beyond the value of the true range, i.e. it approaches the sill gradually without ever reaching it. The equation is:

$$\gamma (h) = C [1 - \exp(-h/a)] + Co .$$

As it is characterized by a gradual approach to the sill, then the true range *a* is one third of the value of the practical range *a*'. The practical range is the distance at which γ (h) approximates the sill. Then $3a = a'$. The true range may also be obtained by the intersection of the tangent at the origin with the sill.

Inverse Problem Theory

The characterization of a hydrocarbon reservoir can be approached as an inverse problem. "Inverse theory is concerned with the issue of making inferences about physical systems from data (usually remotely sensed). Since nearly all data are subject to some uncertainty, these inferences are usually statistical. Further, since one can only record finitely many (noisy) data and since physical systems are usually modeled by continuum equations, if there is a single model that fits the data there must be an infinity of them. A model is a parameterization of the system, usually a function."[2]

In reservoir characterization inverse theory, the data measured from the physical system (the reservoir and its fluids) will be production data (in our case, pressure data measured at the well locations at different times since the beginning of the production). The reservoir simulator provides the relation between the data (pressure data) and the model parameters (porosity and permeability). These relations are not linear and cannot be inverted directly.

We have mentioned that the solution of an inverse problem is not unique if the number of parameters to estimate is larger than the number of observed data, hence we may have to approach the solution of the problem by defining the a posteriori probability density

[2] Scales, J.A. and Smith M.L., Introductory Geophysical Inverse Theory, Samizdat Press, 1993.

function for the model parameters. In petroleum reservoir characterization, however, the complete description of the pdf is impractical given the large dimensions of the usual reservoir models (reservoir grids could contain millions of blocks). One way of approximately characterizing the pdf will be to construct a set of realizations of the model such that this set correctly reflects the uncertainty in these fields. This set of realizations that honor the observed data is considered a solution of the inverse problem.

Flow simulation

A petroleum reservoir simulation model is an algorithm that mathematically simulates the fluid flow in a porous media of a petroleum reservoir, using numerical techniques such as finite differences or finite volumes. The reservoir model is based on a specific mathematical model, which is intended to represent the expected reservoir behavior for a given period of time as determined by the reservoir engineer. A traditional reservoir study is divided into two parts.

- Analysis of past performance: in this part, the model uses the geological data, production data and petrophysical data to simulate the past performance. The obtained results are compared with real values of production history and manual adjustments of the model parameters are made by an application expert. This process is repeated until good agreement is obtained, when the second phase (forecast of future behavior) can start.

- Forecast of future behavior: in the second phase, the aim is for the reservoir engineer to study future possibilities and select those that maximize the profit.

Reservoir Engineering has as its goal the understanding of the subsurface movement and distribution of fluids, and the prediction of the performance of the reservoir. As with everything related to the subsurface, the reservoir is a black box that is never really known but is only understood by its response to inputs. With reservoir engineering that primary response is from production of the hydrocarbons.

Reservoir model

In order to be represented in the mathematical model, the petroleum reservoir is divided into cells that make up a simulation grid. A meticulously defined grid with an adequate degree of refinement will give greater precision to the simulation. On the other hand, the computation effort and precision requirements increase with the number of cells used.

Petroleum reservoir models are very important tools in the petroleum industry because reservoir engineers employ these models in order to estimate the amount of oil available from a specific reservoir and thus maximize the oil recovery. The accuracy of the result is dictated by the quality of the reservoir model used as input to make the performance analysis.

Traditionally, reservoir models were constructed by reservoir engineers with the input of geologists and geophysicists. The reservoir engineer was responsible for entering the data into the reservoir simulator and for obtaining a match of the production history. Most of the time, this match was achieved by modifying the rock properties in some regions of the reservoir. The parameters that were more likely to be modified were the rock properties, specifically, porosity and permeability. Porosity measures the capacity of the rock to store fluids, whereas permeability measures the capacity of the rock to allow the movement of fluids within it.

Recent trends in reservoir characterization aim to the automatic generation of models that honor the production history and satisfy our prior knowledge of the reservoir, without the subjective intervention of the user to achieve the match.

For the purpose of this study, a reservoir model is a map of porosity and permeability averaged over a previously defined geometry. In this case, the geometry of the reservoir is described by a 20 x 1 grid. Therefore, when we generate a model of the reservoir we are describing the average porosity and permeability in each one of the 20 gridblocks.

Prior model

Any previous knowledge that we have about the physical system is used to condition the realizations within logical boundaries. We want all the generated models to resemble

what we consider the reservoir to look like. Normally, geologists have a general idea of the characteristics of different types of reservoirs from examination of other reservoirs or from outcrops. We assume that the porosity in each gridblock is normally distributed while the permeability is lognormal. The mean value of the distributions is known as well as the standard deviation. We also know the correlation coefficient between porosity and permeability.

According to the assumption of multinormal distribution of the reservoir parameters, the prior probability density function is proportional to:

$$p(m) \propto \exp\left(-\frac{1}{2}(m - m_{prior})^T C^{-1}{}_M (m - m_{prior}) \right),$$

Equation 1

where m_{prior} is a vector that contains the estimates of the prior means of the reservoir parameters, and C_M is the prior covariance matrix obtained form the variogram model, i.e.:

$$C_M = \begin{bmatrix} Cf & Cfk \\ Ckf & Ck \end{bmatrix},$$

where C_ϕ is the covariance for the porosity in each gridblock, C_κ is the covariance of permeability and $C_{\phi\kappa}$ is the cross-covariance between porosity and permeability.

The basic assumption is that the prior distribution is multinormal with covariance C_M.

Inverse solution

New trends in reservoir characterization propose the treatment of the problem as an inverse problem. In an inverse problem, we try to make inferences of a physical system (the reservoir) through indirect data from it (the production history). Since these data have measurement errors associated, and the number of reservoir parameters that we want to estimate is larger that the observed data, the solution is non-unique. Our objective is to generate a set of realizations or models of the reservoir that automatically honor the production history and satisfy our previous knowledge of the reservoir.

Bayes estimation

The relationship between the model parameters and the observed data could be expressed as:

$$d_{obs} = g(m) + \varepsilon$$

where d_{obs} is a vector containing a N number of observed data; m is a vector containing a M number of model parameters; g represents the relation between d and m, and ε stands for the measurement errors.

For the flow through porous media, this relationship is non linear, and g(m) represents the flow equations used in the reservoir simulator.

For a given set of observed data, the likelihood function for the model is given by:

$$L(\mathrm{m}|\mathrm{d}_{obs}) \propto exp\left[-\tfrac{1}{2} (g(m)-d_{obs})^{T} C_D^{-1} (g(m)-d_{obs}) \right] .$$

Equation 2

Applying Bayes Theorem, the posteriori probability function of the model parameters conditioned to our prior knowledge and to the likelihood between the observed data and the output of the reservoir simulator is:

$$f(\mathrm{m}|\mathrm{d}_{obs}) \propto L(\mathrm{m}|\mathrm{d}_{obs})\, p(\mathrm{m}) .$$

Equation 3

Substituting equations 1 and 2 in equation 3, the expression for the posteriori distribution function is obtained:

$$f(m \mid d_{obs}) \mu \exp[- \tfrac{1}{2} (d_{obs} - g(m))^T C_D^{-1} (d_{obs} - g(m)) - \tfrac{1}{2} (m - m_{pr})^T C_M^{-1} (m - m_{pr})] .$$

Equation 4

The most probable model is the model that maximizes the function $f(m \mid d_{obs})$. This is known as the maximum a posteriori solution (MAP).

Gauss-Newton method

To maximize the function $f(m \mid d_{obs})$ is the same as minimizing the following function:

$$S(m) = \tfrac{1}{2} [(d_{obs} - g(m))^T C_D^{-1} (d_{obs} - g(m)) + (m - m_{pr})^T C_M^{-1} (m - m_{pr})] .$$

Equation 5

For the minimization of equation 5 we used the Gauss-Newton algorithm with Levenberg-Marquardt step controller. This method requires the computation of the gradient and the approximate Hessian of S(m).

The sensitivity coefficients represent the derivatives of the data (wellbore pressure) with respect to model parameters, i.e.,

$$G_{ij} = \frac{\partial g_i}{\partial m_j},$$

Equation 6

for $1 \leq i \leq N$ and $1 \leq j \leq M$. G is the sensitivity coefficient matrix and is N x M. N is the number of data and M is the number of model parameters. A sensitivity coefficient gives a measure of how strongly the data, $d_i = g_i(m)$ are affected by a change in model parameter m_j.

A one-dimensional, single-phase flow simulator was used, and given the small size of the problem the sensitivity coefficients were computed using the direct method.

$$\frac{\partial g_n}{\partial m_i} = \frac{p_n^{N-1} - p_n^{N}}{\Delta m_i},$$

Equation 7

where, i = 1, … , M, and n = 1, … , N.

The gradient and the approximate Hessian are:

$$\nabla_m S(m_l) = C_M^{-1} (m_l - m_{pr}) + G_l^T C_D^{-1} (g(m_l) - d_{obs})$$

Equation 8

$$H_l = C_M^{-1} + G_l^T C_D^{-1} G_l .$$

Equation 9

The Gauss Newton method is an iterative procedure defined by:

$$m_{l+1} = m_l - H_l^{-1} \nabla S_l ,$$

Equation 10

where l represents the iteration index, m_l is the estimate of the minimum of *S(m)* at the *lth* iteration, G_l is the corresponding matrix of sensitivity coefficients, H_l is the Hessian at the previous iteration, and ∇S_l is the gradient of S(m).

In order to avoid the inversion of H(m) in equation 10, we rewrite this equation introducing the vector *dm*, $dm = m_{l+1} - m_l$, which contains information about the search direction. We define the size of the step in that direction using a step controller factor μ:

$$H_l \, dm^{l+1} = -\nabla S_l \, .$$

Then equation 10 becomes:

$$m_{l+1} = m_l + \mu_l \, H_l^{-1} \, \nabla S_l = m_l + \mu_l \, dm^{l+1} \, .$$

Using matrix inversion lemmas, this expression could be simplified in order to invert a matrix that is only N × N.

$$m_{l+1} = \mu_l \, m_{prior} + (1 - \mu_l) m_l - \mu_l \left[C_M \; G_l^T \; (C_D + G_l \, C_M \; G_l^T)^{-1} \right.$$

$$\left. \times \left(g(m_l) - d_{obs} - G_l^T \; (m_l - m_{pr}) \right) \right] \, .$$

Equation 11

Posteriori covariance

Under the assumption that the relation between the observed data an the model parameters can be linearized around the maximum a posteriori solution, m_{MAP}, then we could express g(m) as:

$$g(m) = g\,(m_{MAP}) + G_{MAP}\,(m - m_{MAP}) + \varepsilon\,(m) \, ,$$

Equation 12

where ε (m) represents the error introduced by the linearization and G_{MAP} is the sensitivity matrix computed at the MAP solution.

If we use equation 12 in equation 5 we obtain:

$$S(m) = \tfrac{1}{2}\,[(m - m_{pr})^T C_M^{-1} (m - m_{pr}) + (G_{MAP}\, m - d_{MAP})^T C_D^{-1} (G_{MAP}\, m - d_{MAP})\,] +$$

$$(G_{MAP}\, m - d_{MAP})^T C_D^{-1} \varepsilon\,(m) + \tfrac{1}{2}\,\varepsilon\,(m)^T C_D^{-1} \varepsilon\,(m),$$

Equation 13

or:

$$S(m) = S\,(m) + (G_{MAP}\, m - d_{MAP})^T C_D^{-1} \varepsilon\,(m) + \tfrac{1}{2}\,\varepsilon\,(m)^T C_D^{-1} \varepsilon\,(m)$$

Equation 14

with

$$S\,(m) = \tfrac{1}{2}\,[(m - m_{pr})^T C_M^{-1} (m - m_{pr}) + (G_{MAP}\, m - d_{MAP})^T C_D^{-1} (G_{MAP}\, m - d_{MAP})\,] .$$

Equation 15

Equation 15 is quadratic and if expanded in a second order Taylor series about m_{MAP}, the result will be exact:

$$S\,(m) = S\,(m_{MAP}) + (\nabla S\,(m_{MAP}))^T (m - m_{MAP}) +$$

$$\tfrac{1}{2}\,(m - m_{MAP})^T \nabla\,(\nabla S\,(m_{MAP}))^T (m - m_{MAP})$$

$$= S\,(m_{MAP}) + (\nabla S\,(m_{MAP}))^T (m - m_{MAP}) +$$

$$\tfrac{1}{2}\,(m - m_{MAP})^T\, [C_M^{-1} + G_{MAP}{}^T C_D^{-1} G_{MAP}]\,(m - m_{MAP})$$

Equation 16

$$\nabla S(m_{MAP}) = C_M^{-1}(m_{MAP} - m_{pr}) + G_{MAP}^T C_D^{-1}(G_{MAP}\, m_{MAP} - d_{MAP})]$$

$$= C_M^{-1}(m_{MAP} - m_{pr}) + G_{MAP}^T C_D^{-1}(g(m_{MAP}) - d_{obs})] .$$

Equation 17

Equation 17 is equivalent to equation 8 evaluated at m_{MAP}. Since the MAP solution is a minimum of equation 5, then equation 8 evaluated at m_{MAP} is equal to zero. Then, equation 15 becomes:

$$S(m) = S(m_{MAP}) + \tfrac{1}{2}(m - m_{MAP})^T [C_M^{-1} + G_{MAP}^T C_D^{-1} G_{MAP}](m - m_{MAP}) .$$

Equation 18

The a posteriori covariance matrix is defined by:

$$C'_M = [C_M^{-1} + G_{MAP}^T C_D^{-1} G_{MAP}]^{-1} .$$

Equation 19

Equation 18 is used in equation 14 to obtain an expression for S(m). The a posteriori pdf is then:

$$f'(m \mid d_{obs}) \propto \exp[-\tfrac{1}{2} S(m)].$$

If the error introduced by the linearization, ε (m), is negligible, then

$$f'(m \mid d_{obs}) \propto \exp[-\tfrac{1}{2}(m - m_{MAP})^T C'^{-1}_{M} (m - m_{MAP})].$$

Equation 20

The algorithms used to draw samples from the prior and posterior pdf of the reservoir models conditioned to observed data will be described in the next chapter.

CHAPTER IV

SAMPLING ALGORITHMS

Linearization about the MAP

Once the most probable model is computed, the a posteriori covariance matrix C'_M of the model can be calculated as:

$$C_{M'} = (G_{\infty}^T C_D^{-1} G_{\infty} + C_M^{-1})^{-1},$$

Equation 19

where G_∞ is the sensitivity coefficient matrix corresponding to the MAP solution, m_∞.

When the relationship between the data and model parameters is linear, it is possible to generate models that honor the production data as

$$m_i = m_\infty + L\ z_i ,$$

Equation 21

where z_i is a vector of independent normal random deviates N [0,1] and L is the square root of C'_M obtained with the Cholesky method. This algorithm requires little computer resources once L and m_∞ are computed.

It is important to understand that the Linearization about the MAP method relies on the assumption that the probability density function of the model parameters approximates the normal distribution. This is true for linear models and normal prior probabilities, but

for the case of flow through porous media, the relation between model parameters and the reservoir response is not linear. It is possible that the probability distribution function for the model may be multimodal (multiple maximums) and because of the linear approximation, samples may be drawn from the vicinity of only one of the peaks.

Examples of the application of the LMAP method can be found in Oliver (1994) and He (1997).

Randomized Maximum Likelihood

Oliver (1996) demonstrated that conditional realizations from a Gaussian random field could be generated by finding the solution to a particular minimization problem. If the covariance of the model parameters and the variance of the observed data are known, samples can be drawn in the following way.

1. Generate an unconditional realization of the model parameters, m_{uc}, $N(m_{pr}, C_M)$.
2. Generate an unconditional realization of the data, d_{uc}, $N(d_{obs}, C_D)$.
3. Minimize the function:

$$S(m) = (m - m_{uc})^T C_M^{-1} (m - m_{uc}) + (g(m) - d_{uc})^T C_D^{-1} (g(m) - d_{uc})$$

Equation 22

The model that minimizes the function in Equation 22 is a realization drawn from a pdf that is an approximation to the posterior pdf for m conditioned to d_{obs}.

This step is similar to the minimization in the LMAP algorithm, with the difference that the regularization is with respect to an unconditional realization of the model and the simulated data instead of the prior model and the observed data. The computational cost of the RML method is higher than LMAP since it requires a minimization process for every new model proposed.

This method is an approximation to the MCMC algorithm in which the acceptance test is ignored. It can be demonstrated that when the relationship between the model parameters and the data is linear, the algorithm generates models from the a posteriori distribution function. This method has a high acceptance ratio, which makes it an efficient sampling algorithm.

Details of the RML method and the relationship to MCMC can be found in Oliver (1996). Examples of the application of the method are available in He (2000) and Reynolds et al. (1999).

Pilot Point Method

This method is a reduced version of the RML method, for the perturbation of the model parameters is done only on select locations called pilot points. The next step is to interpolate the model corrections between the pilot point locations using kriging techniques. The size of the system of equations that must be solved is $Np \times Np$, where

Np is the number of pilot points. This system of equations is smaller than for the RML method.

The algorithm is as follows.

1. Generate an unconditional realization of the model parameters, m_{uc}, $N(m_{pr}, C_M)$.
2. Generate an unconditional realization of the data, d_{uc}, $N(d_{obs}, C_D)$. (This step is neglected in many implementations).
3. Minimize the data misfit function:

$$J(m_p) = (d - g(m_p))^T C^{-1}_D (d - g(m_p)) ,$$

Equation 23

where:

$$m_p = m_{uc} + dm,$$

$$dm = C_M E a \quad .$$

The columns of the matrix E ($M \times Np$) are vectors with all entries equal to zero, except at the location of the gridblocks where pilot points are placed, where the entry is 1. α is the vector of coefficients of pilot points.

For this study, the PP algorithm was evaluated with six and nine pilot points (30% and 45% of the model parameters) distributed uniformly over the study region.

Various methods have been proposed to choose the locations of the pilot points. La Venue et al. (1995) used a coupled adjoint sensitivity analysis and kriging. In most cases, however, the pilot points are located uniformly throughout the reservoir (Xue and Datta-

Gupta, 1997). Examples of the application of this method can be found in De Marsily et al. (1984), Gomez-Hernandez et al. (1987), La Venue et al. (1995), RamaRao et al. (1995), Wen et al. (1996), and Xue and Datta-Gupta (1997).

Markov Chain Monte Carlo

In reservoir characterization, a set of realizations of reservoir properties could be considered a Markov Chain if the probability of generating a particular model depends only on the preceding model in the sequence. MCMC relies on relative probabilities; this is an advantage when the pdf cannot be characterized easily and consequently it is difficult to sample directly from it.

We will denote a particular stochastic realization of reservoir properties as m^i, where the superscript i refers to the i^{th} possible realization (not the i^{th} element in the Markov chain). Each of these realizations has a probability π_i associated with it. π_i is the probability that the realization m^i is the "correct map of reservoir properties", Oliver et al., (1997). The transition probability, p_{ij}, is the probability of transition to state j from state i.

In order to obtain a stationary and ergodic Markov chain (independent of the initial state), two conditions must be met. The transition from any state to another must be possible in a finite number of steps, and the sum of probabilities of being in state m^i, times the probability of transition from state m^i to m^j, must be equal to the probability of state m^j.

$$p_j = \sum_i p_i \quad p_{ij} .$$

Equation 24

In the Metropolis-Hastings algorithm, the transition matrix is split into two components:

$$p_{ij} = \alpha_{ij}\, q_{ij}$$

Equation 25

q_{ij} is the probability of proposing a transition from state m^i to state m^j and α_{ij} is the probability of accepting the proposed transition. Finally, the acceptance probability is:

$$a_{ij} = min\left\{1, \frac{p_j q_{ji}}{p_i q_{ij}}\right\} .$$

Equation 26

See Chib and Greenberg (1995) and Brooks (1998) for a detailed explanation of the Markov chain Monte Carlo method and the Metropolis-Hastings algorithm.

The efficiency of the MCMC algorithm depends heavily on q_{ij}. If q_{ij} is a good approximation of π_j, the acceptance rate will be very high as can be seen from equation 26. In this study, q_{ij} will be based on the a posteriori covariance matrix,

$$C'_M = G^T_r C^{-1}_D G_r + C^{-1}_M .$$

In order to condition the realizations to a variogram, a mean (prior) value, and approximately condition to the data, we generate samples from:

$$m^i = m_{MAP} + L\, Z_i ,$$

Equation 27

where

$$C'_M = L\,L^T$$

and

$$Z^i = (Z_1, Z_2, Z_3, \ldots, Z_M)^T .$$

The correct probability density functions for m_i and m_j are:

$$\pi_j \propto \exp[-\tfrac{1}{2}\,(m^j - m_{pr})^T\, C'^{-1}_M\, (m^j - m_{pr}) - \tfrac{1}{2}\,(g(m^j) - d_{obs})^T\, C_D^{-1}\, (g(m^j) - d_{obs})]$$

$$\pi_i \propto \exp[-\tfrac{1}{2}\,(m^i - m_{pr})^T\, C'^{-1}_M\, (m^i - m_{pr}) - \tfrac{1}{2}\,(g(m^i) - d_{obs})^T\, C_D^{-1}\, (g(m^i) - d_{obs})].$$

Equation 28

And the proposing probabilities are:

$$q_{ij} = (2\pi)^{-M/2} \exp[-\tfrac{1}{2}\,(Z^j \cdot Z^j)]$$

$$q_{ji} = (2\pi)^{-M/2} \exp[-\tfrac{1}{2}\,(Z^i \cdot Z^i)].$$

Equation 29

In this study, local perturbations in the model parameter vector were used to propose states for the MCMC method. A local perturbation consists in the random selection and modification of one or a few elements of the parameter vector, as opposed to a global perturbation, where all M elements of the vector are modified. If the original vector is $Z^i = (Z_1, Z_2, \ldots., Z_k, \ldots, Z_M)^T$, and the k^{th} element was modified, the new proposed model is $Z^j = (Z_1, Z_2, \ldots., Z_k', \ldots, Z_M)^T$.

In order to obtain a more diversified set of samples, several short chains were used as suggested by Gelman and Rubin (1992), instead of a large chain. This is done to avoid

the risk of being trapped in a repetition. Other authors suggest the use of very long chains in order to obtain a good sampling.

For examples of the application of the MCMC method in petroleum engineering see Hegstad et al.(1994), and Oliver et al. (1997).

Rejection Algorithm

The principle of the rejection algorithm is to propose samples from some relatively simple distribution and then apply a test to decide whether or not to accept them. Basically the method is as follows.

If it is desired to sample from a distribution $f_X(x)$ which is difficult to sample from, it is possible to choose a function $h(x)$ ($0< h(x) \leq 1$) and a pdf g(x) that satisfy:

$$f_X(x) = C\, g(x)\, h(x)\,,$$

Equation 30

where C is a constant ≥ 1 and g(x) is a pdf that is easier to sample from.

Samples are generated according to this algorithm.

1. Generate Y from pdf $g(x)$
2. Generate U from $\mathbf{U}[0,1]$
3. If $U \leq h(y)$ return $X = Y$, otherwise return to 1.

In this case the pdf that we wish to sample is the a posteriori distribution function:

$$f(\mathrm{m}) = A\, exp[- \tfrac{1}{2} (m - m_{pr})^T C_M^{-1} (m - m_{pr}) - \tfrac{1}{2} (g(m) - d_{obs})^T C_D^{-1} (g(m) - d_{obs})].$$

Equation 31

The pdf from which samples will be drawn instead will be the linearized approximation of the a posteriori distribution

$$g(m) = B\, exp[- \tfrac{1}{2} (m - m_{MAP})^T C'_M{}^{-1} (m - m_{MAP})].$$

Equation 32

Therefore,

$$h(m) = \frac{f(m)}{C\, g(m)}.$$

Equation 33

If we define the following functions:

$$S_{m1}(m) = (m - m_{pr})^T C_M^{-1} (m - m_{pr})$$

Equation 34

$$S_{m2}(m) = (m - m_{MAP})^T C'_M{}^{-1} (m - m_{MAP})$$

Equation 35

$$S_{d1}(m) = (g(m) - d_{obs})^T C_D^{-1} (g(m) - d_{obs}).$$

Equation 36

Then equation 33 can be written as:

$$h(m) = A' \, exp\, [- S_{m1} - S_{d1} + S_{m2}\,]$$

Equation 37

where:

$$A' = \frac{A}{BC}.$$

Equation 37 is equivalent to:

$$h(m) = exp\,[-S_{m1} - S_{d1} + S_{m2} + Const].$$

The constant inside the exponential function must be chosen to ensure that $h(m) \leq 1$, or equivalently, that:

$$-S_{m1} - S_{d1} + S_{m2} + Const \pounds 0.$$

Equation 38

It is desired that the constant *Const* be as large as possible subject to the inequality 38. For this study, *Const* was computed evaluating S_{m1}, S_{d1}, and S_{m2} at the MAP solution. The selection of the constant is a critical issue in the application of this algorithm. By evaluating equation 38 at the MAP, it is assumed that the linearized approximation to the posteriori distribution is above the true distribution everywhere. Since the actual shape of the posteriori distribution is not known, there is a risk that the assumption is incorrect. If a low value of *Const* is chosen, the acceptance rate will be high but there is a high risk that the algorithm will not be sampling from the correct pdf. On the other hand, if as a safety factor *Const* is chosen too large, the acceptance rate will be significantly reduced. Under this situation, the application of the algorithm may not be feasible. This is the reason why, in practice, this algorithm is not used in reservoir characterization.

CHAPTER V

METHODOLOGY

Description of the Synthetic Cases

CASE 1.

The "one-dimensional" synthetic reservoir with a length of 1000 ft has been discretized into 20 uniform gridblocks of length Δx = 50 ft. The cross-sectional area is 1000 ft^2. Porosity, ϕ, and log permeability, ln(k), are considered stationary random functions of second order, with the following mean and variances:

$$\mu_\phi = 0.25, \sigma^2_\phi = 0.0025$$

$$\mu_{\ln(k)} = 4.5, \sigma^2_{\ln(k)} = 1.0 .$$

The prior model is multivariate Gaussian. For both porosity and permeability an exponential covariance is assumed. The range of the covariance is 175 ft,

$$C(h) \propto \exp (- | 3\ h | /175) .$$

The correlation coefficient between porosity and log-permeability is 0.5.

Wells are located at gridblocks 7, 13 and 18. Pressure data are recorded at each well location. The well at gridblock 13 is an active well with a constant production rate of 100 b/d, and the other two wells are observation wells. Pressure data are provided with measurement errors identically distributed independent Gaussian random variables with mean zero and variance 25.0.

The flow is single phase with an oil viscosity of 2 cp and a total compressibility of 4×10^{-6} psi^{-1}. The initial reservoir pressure is 3500 psi.

The distribution of the "real" parameters can be observed in figure 1 for porosity and figure 2 for permeability. Figure 3 shows a plot of ΔP versus time for the true case.

CASE 2.

Case 2 is similar to case 1 with the difference that the variance of the pressure data was changed to 0.25. This decrease in the variance of the errors in the observed data was intended to make the problem more nonlinear.

The observed data for cases 1 and 2, and the real pressure data are presented in Table 1.

Comparison Criteria

To evaluate the ability of the approximate methods to sample correctly, five functionals were computed for each realization. These functionals were chosen to be representative of important reservoir characteristics:

1. A measure of the steady-state productivity

$$f_1(m) = \left(\frac{1}{N} \sum_1^N \frac{1}{K} \right)^{-1} .$$

2. A measure of the pore volume and oil in place

$$f_2(m) = \frac{1}{N}\sum_1^N f \ .$$

3. A measure of the breakthrough time in a waterflood

$$f_3(m) = \frac{1}{N}\sum_1^N \frac{f}{K} \ .$$

4. Maximum value of permeability

$$f_4(m) = max_{1,N}\{K\} \ .$$

5. Minimum value of permeability

$$f_5(m) = min_{1,N}\{K\} \ .$$

For the case # 1, 5000 realizations were generated for LMAP, RML, PP6, and PP9. For MCMC method 37380 realizations were generated, and for Rejection Method 1589 realizations were obtained.

Similarly, for case # 2, 5000 realizations were generated for LMAP, RML, PP6, and PP9 and 685611 for MCMC. The Rejection method was not evaluated for case 2.

In addition to these five functionals, the mismatch of the models with respect to the prior model and the MAP estimate were computed. Also, the mismatch of the computed pressure with the observed data was calculated.

$$S_{m1}(m) = (m - m_{pr})^T C_M^{-1} (m - m_{pr})$$

$$S_{m2}(m) = (m - m_{MAP})^T C'_M{}^{-1} (m - m_{MAP})$$

$$S_{d1}(m) = (g(m) - d_{obs})^T C_D^{-1} (g(m) - d_{obs})$$

Description of the Program

The program used to solve this problem is in Fortran. The program contains a single-phase simulator, the subroutines to compute the sensitivity coefficients with the direct method, a matrix solver subroutine, and the sampling algorithms.

This study evaluates the application of the Linearization about the MAP algorithm, the Randomized Maximum Likelihood method, two versions of the Pilot Point method, the Markov Chain Monte Carlo method and the Rejection algorithm. The two versions of the Pilot Point method are defined by varying the control variable (number of pilot points). A single-phase, one-dimensional problem was chosen, because it allows the generation of a large number of realizations. It is expected that this will facilitate the construction of the pdf of the reservoir models, and therefore quantify better the uncertainty in reservoir

parameters. Larger problems require more computer resources and hence limit the number of realizations that can be sampled.

MAP Linearization Algorithm (LMAP)

The MAP linearization algorithm is used to sample from the approximation to the posterior pdf f (m, d | m_{pr}, d_{obs}). The MAP estimate is obtained through an optimization procedure (Gauss-Newton with Levenberg-Marquardt step controller) and the posterior covariance matrix is computed. After this, the computer resources required to sample from the approximated pdf are lower than for any other method tested. The number of realizations generated with this algorithm was 5198 for the first case and 5000 for the second case.

Randomized Maximum Likelihood (RML)

It is expected that this algorithm will sample from the a posteriori pdf f (m, d | m_{pr}, d_{obs}), or a close approximation. The two steps of this algorithm consist in the generation of unconditional realizations of the model and the data, and second the optimization procedure to condition the realizations to prior knowledge and observed data. A total of 5000 realizations were sampled.

Pilot Point Method (PP)

The pilot point method is used to try to sample from the posterior pdf f (m, d | m_{pr}, d_{obs}). This algorithm is a reduced version of the RML method, since the optimization step is done only on a subspace of the previous. The algorithm was run with two different patterns for the location of the pilot points. First, 33% of the gridblocks were used as pilot points (6 gridblocks), and then 45% of the gridblocks (9 pilot points). For both patterns, the blocks containing the observation and production wells were included as pilot points. Also, pilot points were within the range of the variogram, but not adjacent. 5000 independent realizations were generated for each variation of the PP method.

Markov Chain Monte Carlo Method (MCMC)

The MCMC algorithm is used to sample from the posterior pdf. It can be shown that after convergence this algorithm samples from the correct pdf. The number of realizations generated with this method was 37380 for the first case and 685611 (these figures include the repeated elements in the chain). Realizations were produced with local perturbations choosing one element of Zi at a time.

For the first case, the starting elements of the chains were conditional realizations from the MAP ($m^0 = m_{MAP} + L\ Z_i$), where $L^T\ L = C_M'$. In order to improve the mixing of the

samples, 21 chains of 1000 distinct elements were created; each chain had a different initial element. Elements with high data mismatch occurred at the beginning of the chain (about 3 percent of the chain elements), but this mismatch was consistently reduced as new elements entered the chain. Realizations with high data mismatch ($S_{d1} > 40.0$) were eliminated. The value of 40.0 was chosen after observing the trend of Sd1 for the elements of a chain. It was noted that Sd1 started with high values in the first elements and stabilized around 40.

Local perturbations were chosen instead of global perturbations due to the high acceptance ratio. An average of 0.78 elements was rejected in order to accept a new realization. For this case, the requirements in computer resources with this method were higher than LMAP but lower than the other three algorithms.

For the second case, where the variance of the errors in the observed data was decreased in order to make the problem more non-linear, the initial elements for the Markov chains were samples accepted from the RML method. The totals of 60 samples from RML were used in order to generate 60 chains. After an elimination process to discard elements with a high data mismatch term ($S_{D1} > 40.0$), usually the first elements in the chain, a total of 685611 samples were left. The rejection rate increase for this case from an average of 0.78 to 11.7 proposed realizations in order to accept a model.

Rejection Algorithm (REJ)

This algorithm samples from the posterior pdf f (m, d | m_{pr}, d_{obs}), by sampling from a pdf easier to sample and then testing the proposed realization before acceptance. As described on page 39, realizations were generated from the approximation to the posteriori pdf (conditional realizations from the MAP). The acceptance test requires the evaluation of data and model mismatches.

The efficiency of this algorithm is very low. For the base case, the number of rejected models with this method was very high, with an average of 740 rejections in order to accept a realization. This efficiency issue makes this model the most demanding in computer resources, and limited the number of realizations produced to 1589. It was not feasible to use this method for the more non-linear case 2.

CHAPTER VI

RESULTS

Figures 5 to 21 show the results for the two study cases in the form of histograms. Histograms of all the methods are presented for the five functionals and the mismatch functions: S_{M1}, S_{M2} and S_{D1}. At the top of the figures is a table containing a summary of the statistical parameters for the data sets. For the first five figures (corresponding to the five functionals), the "real" value of the functional (evaluated with the "real case") is also shown in the upper part of the summary table.

In order to compare the methods, features such as the shape and the span of the distribution, as well as the inclusion of the "real" value, were considered. It was observed that for the first case, a problem that is close to linear, the MCMC method and the Rejection algorithm performed similarly, and therefore confidence was given to the fact that these methods were sampling from the real pdf. For the first case, MCMC and Rejection were used as a benchmark to evaluate the performance of the other algorithms.

Figure 4 shows the data mismatch term for 59 Markov Chains for case number 2. It can be observed in this figure that elements of the same chain have a high correlation. It was not possible to obtain a well mixed sample set with the MCMC method for this second case, and therefore it was not possible to use the histograms of this method as a benchmark.

Functional No. 1

CASE 1

This is measure of the steady state productivity. It can be observed in Figure 5 that for this functional the distribution curves for all methods seem Gaussian. The less symmetric envelopes are those corresponding to the pilot point method. The average for all the methods is higher than the real value, which means that the production forecast will be higher with all methods.

CASE 2

For the second case (Figure 6), histograms for LMAP, RML, MCMC, and PP9 are skewed to the right. For MCMC the true value for this functional is not even included in the histogram, while for the other four methods the true value is in the lower quartile. The histogram for PP6 is the most spread and shows no resemblance to the other histograms.

Functional No. 2

CASE 1

This functional is a measure of the pore volume and oil in place. Again the histograms in Figure 7 show symmetry for all the methods, and all of them are close to the real value. The two versions of the pilot point method have means that are farther for the real value (3% difference), besides being the most spread and asymmetric curves.

It is interesting to observe how for all methods the mean approximated the real value of this functional, taking in consideration that the a priori value given (0.25) had a difference with the true value of 7%.

CASE 2

For the second case, the means of the histograms are closest to the priori value, and the true value is in the lower quartile for all the distributions (figure 8). Histograms show some symmetry, and the most spread distribution corresponds to PP9.

Functional No. 3

CASE 1

This functional is a measure of the breakthrough time in a waterflood. All methods have average values of this functional close to each other, and all of them are lower than the real value. This implies that all methods will –on average– predict a shorter breakthrough time than the truth case. Figure 9 shows that all curves are skewed to the left, the curve with the least spread corresponding to PP9.

CASE 2

For the second case (figure 10) all the histograms are skewed to the left. For MCMC the true value for this functional is not even included in the histogram, while for the other four methods the true value is in the upper quartile. The histogram for PP6 has the most spread.

Functional No. 4

CASE 1

This functional is an indicator of the maximum value of permeability in each realization. It can be observed in Figure 11 that in average, all the methods overestimate the maximum permeability, in particular the pilot point methods. This algorithm is affected by the presence of extreme values that affect the mean, however the median for all methods is still higher than the true value. The methods that are closest to the real value are the Rejection algorithm and MCMC, and the difference was still 51%.

CASE 2

For the second case (figure 12) all the histograms are skewed to the left in what is suspected to be a log-normal distribution. For the MCMC and PP6 methods the true value for this functional is not even included in the histograms, while for the other four methods the true value is in the lower quartile. The histogram for PP6 is the largest spread and indicates that the method did not perform well in discriminating models with unrealistically high values of permeability. The difference with the real value is much more higher than for the first case for all the methods.

Functional No. 5

CASE 1

This functional is an indicator of the minimum value of permeability in each realization. For all methods, the histograms in figure 13 show curves with some symmetry, but again

all of them overestimate —in average— the real value, which is in the lower quartile for all the distributions.

CASE 2

Figure 14 shows that for all methods, except MCMC, the histograms present some symmetry, but again the real value is in the lower quartile for all the distributions.

Model Mismatch with respect to Prior

CASE 1

Most of the methods show curves with some symmetry, except PP6, which is skewed to the left (Figure 15). MCMC and rejection have the smallest values, while PP6 has the highest mismatch in average and shows a more spread distribution.

CASE 2

All the distributions in Figure 16 are slightly skewed to the left. MCMC is the least spread while the two pilot point methods show the highest standard deviations.

Model Mismatch wrt MAP Estimate

CASE 1

For this comparison parameter (Sm2), the histograms of the realizations generated using methods that involved an optimization process (RML, PP6, and PP9) are much different from the linearization, rejection, and MCMC methods, all of which seem to generate symmetric histograms (Figure 17).

CASE 2

All methods but LMAP show a high spread and very high mismatches with respect to the MAP estimate (Figure 18). The MCMC distribution shows a very heterogeneous histogram. It is supposed that the heterogeneity in the histogram is showing the effects of using 20 independent chains. Diffusion from the starting points does not appear to have been very efficient. It was attempted to change the procedure for the perturbations to 3 parameters at a time in order to improve the distribution of the samples, but the results obtained showed little improvement and the computing time was very high. Another possibility was to implement global perturbations, but the low acceptance rate of the algorithm would not allow generating a large number of independent realizations in a reasonable time.

Data Mismatch

CASE 1

It can be observed in Figure 19 that all curves are skewed to the left for this comparison parameter. MCMC, REJ and RML are the algorithms with the lowest values of data mismatch. The method with the highest data mismatch was LMAP. It was observed that in the two pilot point methods as well as in RML, the optimization procedure accepts some models with high values of the data mismatch term. A simple acceptance criterion was implemented to eliminate realizations that did not honor the data satisfactorily, and would otherwise bias the histograms.

CASE 2

Figure 20 shows that all histograms are skewed to the left. The realizations from the LMAP method showed again a poor performance honoring the observed data. The most spread distributions correspond to the two pilot point methods.

CHAPTER VII

CONCLUSIONS

Important decisions that will affect the development of an oilfield are the final product of reservoir characterization studies. Sampling from the posterior pdf of all plausible reservoir models in order to estimate uncertainty is the best way to address this problem. Given the non-linear nature of the relations involved and the fact that the inverse problem is ill posed, the stochastic algorithms that honor both the production data and the prior knowledge of the reservoir place a high demand in computer resources. For this reason, for practical applications, engineers have to work with only a few realizations if not a single one. This study was intended to extensively evaluate five different sampling algorithms in order to derive some observations that could be useful in the selection of a sampling method for a real problem. Indeed, the dimensions and type of problem studied allowed this extensive sampling.

In the present, the requirement in computer resources is an issue that may dictate the ultimate selection of a particular method in many cases. Even though the linearization about the MAP proved to be a fast sampling method, it is important to highlight that this algorithm provided poor matches to the observed data for both cases. As a result, it would be difficult to place much confidence in realizations obtained with this method in spite of their lower computing cost.

For the first case, the Markov chain Monte Carlo and Rejection methods gave satisfactory and very similar results in estimating uncertainty. This gave us confidence in determining that these two methods are sampling from the real pdf. Unfortunately, the low efficiency of these methods makes them poor candidates for practical assessment of uncertainty.

For the second case the variance in the errors in the observed data was decreased. It was observed that the true values were farther away from the means for the functionals evaluated. This is believed to be due to the stronger constraints in the likelihood term.

The Markov chain Monte Carlo method did not perform satisfactorily in the second case, where for some instances the real values of the functionals were not even included in the histograms. Therefore there is little confidence in the use of this method as a benchmark. Alternatives as the use of more elements in the perturbation process showed not to be efficient, and a global perturbation process would have required an unfeasible amount of time in order to generate a large number of realizations. It is suggested to evaluate the use of a hybrid Markov Chain Monte Carlo method (Bonet-Cunha, 1998) for this problem.

In an overall evaluation, the reduced version of the RML algorithm, i.e. the Pilot Point method performed worse than all the other methods.

None of the approximate algorithms had a good performance sampling the extreme values in the reservoir for both cases. This requires getting the tails of the distribution correct. If estimation of extreme values, or estimation of an extreme event is of interest, it may be necessary to take a different approach.

Models that perform the best honoring production data are the MCMC, Rejection and RML algorithms for the first case, and MCMC and RML for the second case. Of the approximate methods, randomized maximum likelihood appears to have the best performance for sampling. Estimation of the mean and standard deviation is improved if outliers are rejected. A simple acceptance criterion based on the magnitude of the data mismatch term, proved to be effective in this study. It is recommended to develop a better acceptance criterion for RML and PP methods.

It is advisable to implement this study in a higher dimensional problem in order to validate the results obtained.

R E F E R E N C E S

Bonet-Cunha, L., Oliver, D.S., Redner, R.A., Reynolds, A.C., 1998, "A Hybrid Markov Chain Monte Carlo Method for Generating Permeability Fields Conditioned to Multiwell Pressure Data and Prior Information," *SPE Journal*, September 1998, p 261-271.

Brooks, S. P., 1998, "Markov Chain Monte Carlo Method and its Application," *The Statistician*, 1998, Vol. 47, No. 1, p 69-100.

Carrera, J., and Neuman, S.P., "Estimation of Aquifer Parameters under Transient and Steady State Conditions: 1. Maximum Likelihood Method Incorporating Prior Information," *Water Resources Research*, 1986a, 22(2), p.199-210.

Carrera, J., and Neuman, S.P., "Estimation of Aquifer Parameters under Transient and Steady State Conditions: 2. Uniqueness, Stability, and Solution Algorithms," *Water Resources Research*, 1986b, 22(2), p. 211-227.

Chib, S. and Greenberg, E., 1995, "Understanding the Metropolis Hastings Algorithm," *The American Statistician*, Nov. 1995, Vol. 49, No. 4, p 327-335.

Dagan, G., "Stochastic Modeling of Groundwater Flow by Unconditional and Conditional Probabilities: The Inverse Problem," *Water Resources Research*, 1985, 21(1), p.65-72.

Dagan, G., and Rubin, Y., "Stochastic Identification of Recharge, Transmissivity and Storativity in Aquifer Transient Flow," *Water Resources Research*, 1988, 24(10), p.1698-1710.

De Marsily G., Lavedan, G., Boucher, M., Fasanino, G., 1984, "Interpretation of Interference Tests in a Well Field Using Geostatistical Techniques to Fit the Permeability Distribution in a Reservoir Model," *Geostatistics for Natural Resources Characterization*, Part 2, D. Reidel, 1984, p 831-849.

Floris, F.J.T., Bush, M.D., Cuypers, M., Roggero, F., Syversveen, A.R., 1999, "Comparison of production forecast uncertainty quantification methods - an integrated study," 1st Conference on Petroleum Geostatistics, 20-23 April, Toulouse. (Paper submitted to the Petroleum Geoscience magazine of EAGE, July 1999).

Gelman, A., and Rubin, D.B., 1992, "Inference from Iterative Simulation Using Multiple Sequences (with discussion)," *Statistical Science*, Vol. 7, p 457-511.

Gómez-Hernández, J., Sahuquillo, A., Capilla, J., 1997, "Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Piezometric Data, 1. Theory," *Journal of Hydrology*, 1997, No. 203, p 162-174.

Grinrod, P., and Impey, M.D., "Fractal Field Simulations of Tracer Migration within the WIPP Culebra Dolomite," Report, Intera Information Technologies, Dec. 1991.

Gutjahr, A.L., and Wilson, J.R., "Cokriging for Stochastic Flow Models," *Transport in Porous Media*, 1989, 4(6), p. 585-598.

Gutjahr, A.L., Bullard, B., Hatch, S., Hughson, L., "Joint Conditional Simulations of the Spectral Method Approach for Flow Modeling," *Stochastic Hydrology Hydraulics*, 1994, 8(1), p.79-108.

He, N., Oliver, D.S., and Reynolds, A.C., 2000, "Conditioning Stochastic Reservoir Models to Well Test Data," *SPE Reservoir Evaluation Engineering*, 2000, Vol. 3, No. 1, p. 74-79.

He, N., 1997, "Three Dimensional Reservoir Description by Inverse Problem Theory Using Well-Test Pressure and Geostatistical Data," PhD Dissertation, The University of Tulsa.

Hegstad, B.K., More, H., Tjelmeland, H., Tyler, K, 1994, "Stochastic Simulation and Conditioning by Annealing in Reservoir Description," *Geostatistical Simulation*, M. Armstrong and P.A. Dowd, editors, p 43-55. Kluwer Acad.

Hoeksema, R.J., and Kitanidis, P.K., "An Application of the Geostatistical Approach to the Inverse Problem in Two-Dimensional Groundwater Modeling," *Water Resources Research*, 1984, 20(7), p. 1003-1020.

Kalita, R. and Betancourt S.S., 1999, "Markov Chain Monte Carlo Methods for Conditioning Permeability Fields to Pressure Data," The University of Tulsa, Internal Report, December 1999.

Kalos, M.H., Whitlock P.A., *Monte Carlo Methods. Volume I: Basics*, New York, John Wiley and Sons, 1986.

Kitanidis, P.K., and Lane, R.W., "Maximum Likelihood Parameter Estimation of Hydrologic Spatial Processes by the Gauss-Newton Method," *Journal of Hydrology*, 1985, 79(1-2), p. 53-71.

Kitanidis, P.K., and Vomvoris, E.G., "A Geostatistical Approach to the Inverse Problem in Groundwater Modeling (Steady State) and One-Dimensional Simulations," *Water Resources Research*, 1983, 19(3), p. 677-690.

La Venue, A.M., RamaRao, B.S., Marsily, G., Marietta, M.G., 1995, "Pilot Point Methodology for Automated Calibration of an Ensemble of Conditionally Simulated Transmissivity Fields, 2. Application," *Water Resouces Research*, 1995, Vol. 31, No. 3, p 495-516.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Orlando, Fla. Academic Press 1984.

Oliver, D.S., 1994, "Incorporation of Transient Pressure Data into Reservoir Characterization," In Situ, Vol.18, No.13, p. 243-275.

Oliver, D.S., 1996, "Multiple Realizations of the Permeability Field from Well Test Data," SPE 27970, *SPEJ*, 1(2), p. 145-154.

Oliver, D.S., He, N. and Reynolds, A.C., 1996, "Conditioning Permeability Fields to Pressure Data," V European Conference for the Mathematics of Oil Recovery, p. 1-11.

Oliver, D.S., Cunha, L.B. and Reynolds, A.C., 1997, "Markov Chain Monte Carlo Methods for Conditioning Permeability Fields to Pressure Data," *Mathematical Geology*. Vol. 29, No. 1, p. 61-91.

Oliver, D.S., 1999, "The Pilot Point Method," Personal Notes, p. 1-12.

Oliver, D.S., 1999, "RML Sampling," Personal Notes, p. 1-9.

Omre, H., Tjelmeland, H. and Wist, H.T., 1999, "Uncertainty in History Matching - Model Specification and Sampling Algorithms," *Technical Report Statistics*, No. 6/1999, Norwegian University of Science and Technology.

Omre, H., Tjelmeland, H. and Wist, H.T., 2000, "Sampling Algorithms for Stochastic Reservoir Models Conditioned to Production History," *Technical Report Statistics*, No. ?/2000, Norwegian University of Science and Technology.

Press, W.H., Teukolsky,S.A., Vetterling, W.T., and Flannery, B.P., *Numerical Recipes in C : The Art of Scientific Computing*, 2nd Edition, Cambridge University Press, 1993.

RamaRao, B.S., La Venue, A.M., Marsily, G., 1995, "Pilot Point Methodology for Automated Calibration of an Ensemble of Conditionally Simulated Transmissivity Fields,

1. Theory and Computational Experiments," *Water Resouces Research*, 1995, Vol. 31, No. 3, p 475-493.

Reynolds A.C., He, N., Oliver, D.S., 1999, "Reducing Uncertainty in Geostatistical Description with Well Testing Pressure Data," *Reservoir Characterization – Recent Advances, American Association of Petroleum Geologists*, 1999, p 149-162.

Ripley, B.D., *Stochastic Simulation*, New York, John Wiley and Sons, 1987.

Robin, C., de Fouquet, C., Chiles, J.P., and, Matheron, G., "Cross-Correlated Random Field Generation with the Direct Fourier Transform Method," *Water Resources Res.*, 1993, 29(7), p.2385-2397.

Ross S.M., *Simulation (Statistical Modeling and Decision Science)*, 2nd edition, Academic Press, 1996.

Rubin, Y., and Dagan, G., "Stochastic Identification of Transmissivity and Effective Recharge in Steady Groundwater Flow 1. Theory," *Water Resources Research*, 1987, 23(7), p.1185-1192.

Rubin, Y., "Prediction of Tracer Plume Migration in Disordered Porous Media by the Method of Conditional Probabilities," *Water Resources Research,* 1991a, 27(6), p.1291-1308.

Rubin, Y., "Transport in Heterogeneous Porous Media: Prediction and Uncertainty," *Water Resources Research,* 1991b, 27(7), p.1723-1738.

Rubin, Y., and Dagan, G., "Conditional Estimation of Solute Travel Time in Heterogeneous Formations: Impact of Transmissivity Measurements," *Water Resources Research,* 1987, 28(4), p.1033-1040.

Scales, J.A. and Smith M.L., 1993, *Introductory Geophysical Inverse Theory*, Samizdat Press.

Wen, X., G, Gómez-Hernández, J., Capilla, J., Sahuquillo, A., 1996, "Significance of Conditioning to Piezometric Head Data for Predictions of Mass Transport in Groundwater Modelling," *Mathematical Geology*, 1996, Vol. 28, No. 7, p 951-968.

Wu, Z., Reynolds, A.C. and Oliver, D.S., 1998, "Conditioning Geostatistical Models to Two Phase Production Data," SPE 49003, SPE Annual Technical Conference and Exhibition, New Orleans, LA, September 27-30, 1998.

Xue, G. and Datta-Gupta, A., 1997, "Structure Preserving Inversion: An Efficient Approach to Conditioning Stochastic Reservoir Models to Dynamic Data," SPE 35412, SPE Annual Technical Conference and Exhibition, San Antonio, TX, October 5-8, 1997.

Zimmerman, D.A., de Marsily, G., Gotway, C.A., Marietta, M.G., Axness, C.L., Beauheim, R., Bras, R., Carrera, J., Dagan, G., Davies, P.B., Gallegos, D.P., Galli, A., Gomez-Hernandez, J., Gorelick, S.M., Grindrod, P., Gutjahr, A.L., Kitanidis, P.K., Lavenue, A.M., Mc Laughlin, D., Neuman, S.P., Ramarao, B.S., Ravenne, C. And Rubin, Y., 1998, "A Comparison of Seven Geostatisically Based Inverse Approaches to Estimate Transmissivities for Modeling Advective Transport by Groundwater Flow," *Water Resources Report*, 34(6), p.1373.1413.

NOMENCLATURE

d_{obs}: vector of observed data
d_{uc}: unconditional realization of the data
h: lag distance
m: vector of model parameters
m_{MAP}: maximum a posteriori solution
m_{pr}: prior model
m_{uc}: unconditional model
q: probability of proposing a transition from one state to another (MCMC)
C_M: prior covariance matrix of the model parameters
C'_M: a posteriori covariance matrix
C_D: covariance matrix of the observed data
G: sensitivity coefficient matrix
S_{d1}: data mismatch
S_{m1}: model mismatch with respect to prior model
S_{m2}: model mismatch with respect to MAP solution
H: Hessian
J: data misfit function (PP)
L: lower diagonal matrix
M: number of model parameters
N: number of observed data
N_p: number of pilot points
Z: vector of normal random deviates
∇S: gradient

Greek

α: acceptance probability (MCMC) or vector of pilot points coefficients
μ: mean or step controller factor
π: probability of a state (MCMC)
γ: variogram
ϕ: porosity
κ: permeability
σ: standard deviation
ε: error term

	Time	Real Pressure		Simulated Pressure with variance = 25		Simulated Pressure with variance = 0.25
Well # 7	4.00E-04	3500		3511.66		3500.56
	1.20E-03	3500		3493.86		3500.30
	2.00E-03	3500		3501.7		3499.76
	4.00E-03	3499.999		3491.82		3498.84
	8.00E-03	3499.95		3497.62		3498.40
	1.20E-02	3499.694		3497.31		3499.15
	2.00E-02	3497.652		3497.66		3498.28
	3.60E-02	3484.642		3482.42		3485.15
	6.00E-02	3445.014		3454.62		3446.70
	1.02E-01	3341.721		3335.4		3343.97
Well # 13	4.00E-04	3462.251		3458.63		3462.45
	1.20E-03	3404.852		3401.9		3404.55
	2.00E-03	3361.074		3356.9		3361.30
	4.00E-03	3276.185		3277.45		3275.00
	8.00E-03	3142.248		3142.18		3141.84
	1.20E-02	3028.475		3024.59		3028.60
	2.00E-02	2834.61		2833.29		2835.03
	3.60E-02	2516.052		2515.78		2517.02
	6.00E-02	2128.254		2136.08		2130.13
	1.02E-01	1595.021		1588.98		1598.24
Well # 18	4.00E-04	3500		3505.9		3500.21
	1.20E-03	3500		3507.73		3499.95
	2.00E-03	3499.998		3495.8		3499.34
	4.00E-03	3499.966		3494.36		3498.43
	8.00E-03	3499.309		3502		3498.59
	1.20E-02	3496.817		3497.7		3497.29
	2.00E-02	3482.827		3480		3483.46
	3.60E-02	3420.182		3419.5		3420.98
	6.00E-02	3273.167		3275.12		3275.01
	1.02E-01	2954.232		2952.71		2956.68

Table 1. “True” Pressure Data and “observed” Data for Cases 1 and 2.

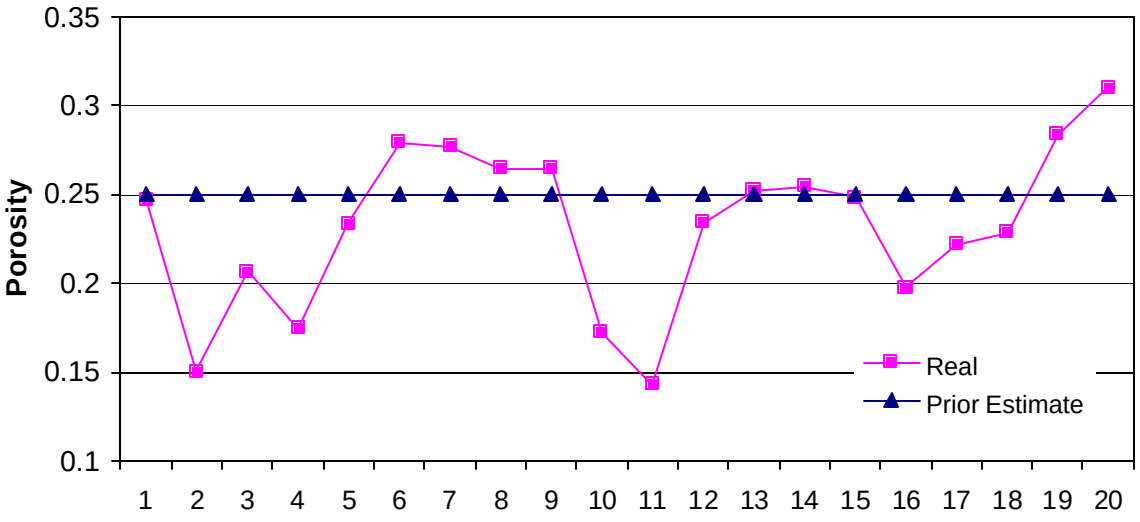


Figure 1. Porosity distribution for the "true" case.

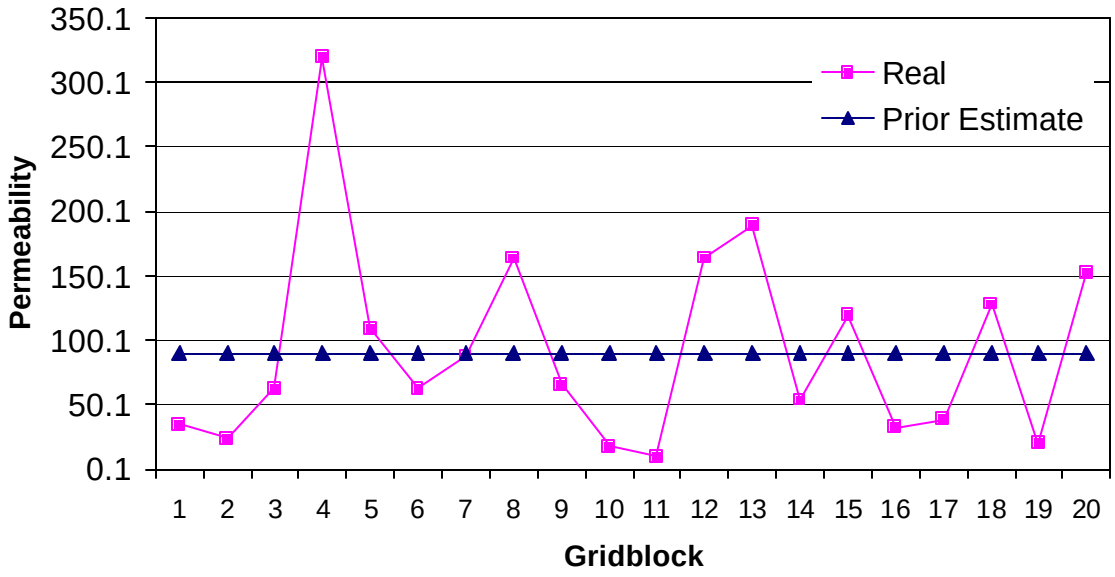


Figure 2. Permeability distribution for the "true" case.

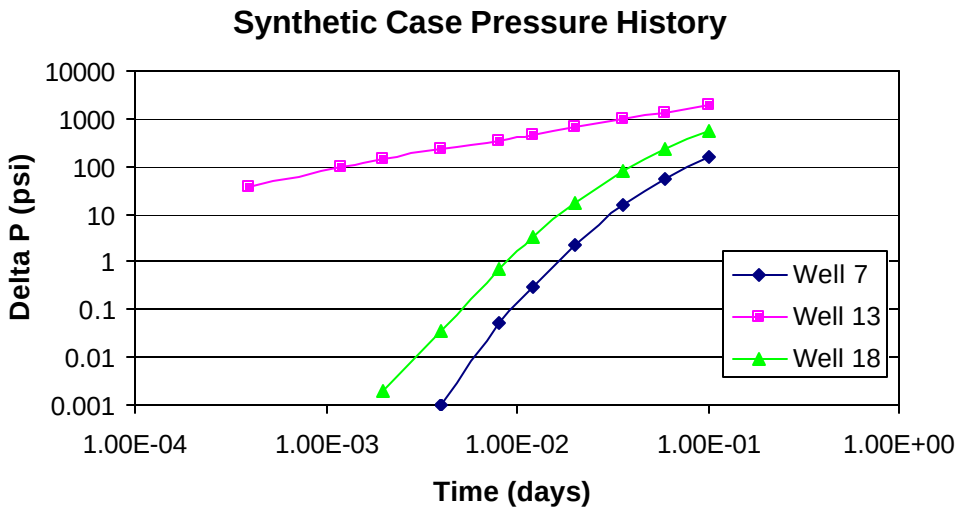


Figure 3. Delta Pressure vs. Time for the "true" case

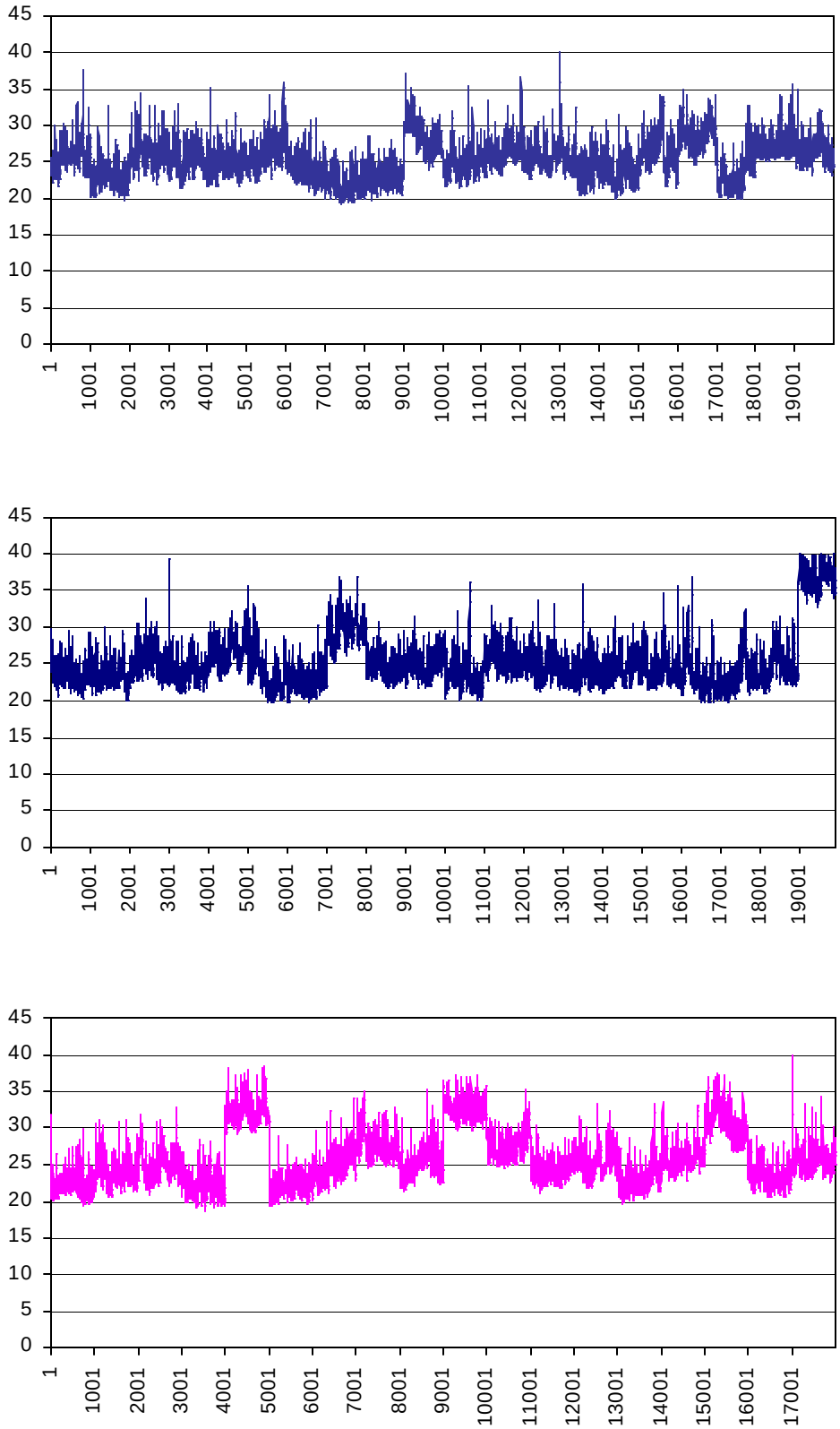
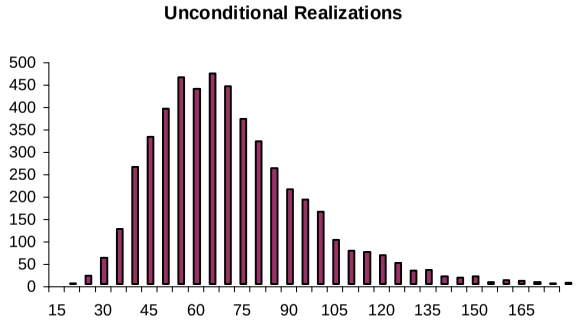
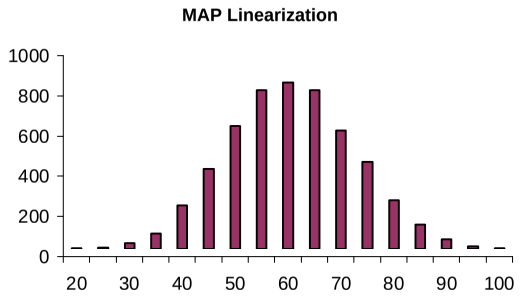


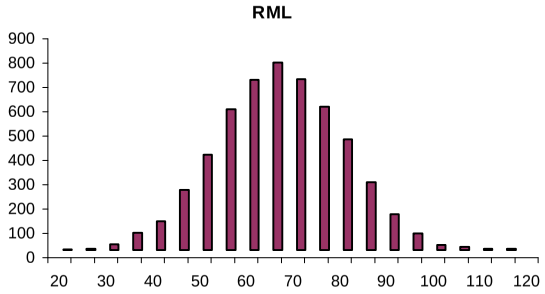
Figure 4. Data Mismatch for 59 Markov Chains (Case 2)

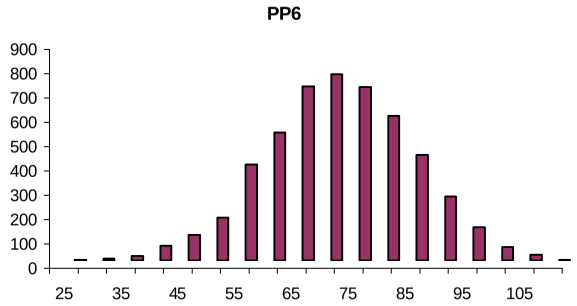
Table No. 2. Statistical Summary for Functional # 1. Case 1 (Data Variance = 25)

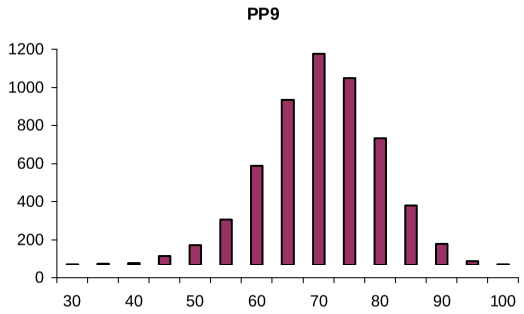
Real Value	43.1289						
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**REJ**	**UNCOND**
Number of Realizations	5198	5000	5000	5000	20000	1589	5000
Mean	58.0269	63.0780	68.2038	67.9248	64.2519	63.6045	63.2920
Standard Deviation	11.9599	13.4840	12.8703	9.2464	12.1808	12.3049	24.5610
Median	57.8795	63.0026	68.2978	68.1890	64.2492	63.6600	59.3462
Coefficient of Assymetry	0.0642	0.0245	-0.0789	-0.3144	-0.0507	-0.0430	1.0634
Minimum Value	15.1475	16.3452	20.1046	26.1275	21.7340	22.3238	12.5281
Lower Quartile	49.7274	53.9260	59.6551	62.1910	55.9987	55.2194	45.7865
Upper Quartile	66.1734	72.2664	77.0960	74.1688	72.5235	72.2600	76.3188
Maximum Value	99.6830	113.6302	114.2672	97.8673	112.2306	109.1336	220.3421

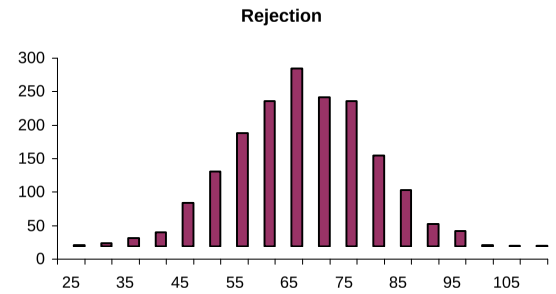












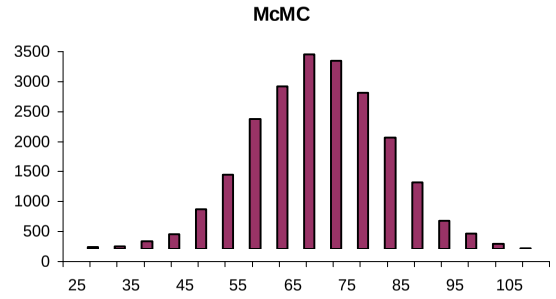
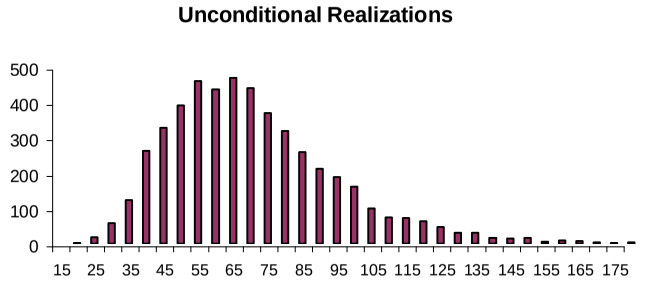
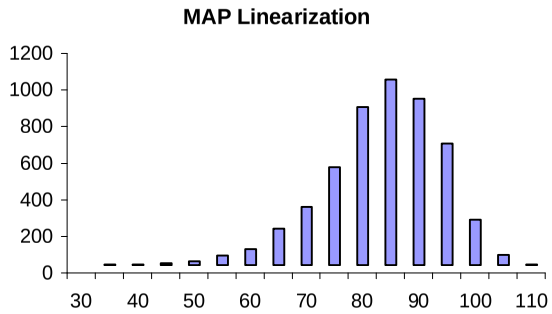


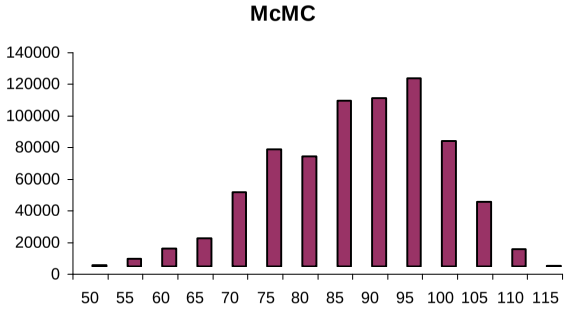
Figure 5. Histograms for Functional # 1. Case 1.

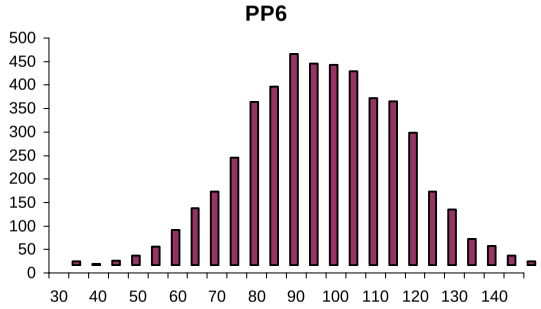
Table No. 3. Statistical Summary for Functional # 1. Case 2 (Data Variance = 0.25)

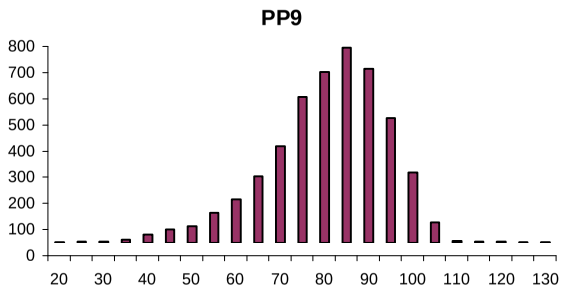
Real Value	43.129					
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**UNCOND**
Number of Realizations	5000	5000	5000	5000	685611	5000
Mean	80.952	84.018	89.969	78.879	85.646	63.292
Standard Deviation	10.359	12.750	18.583	13.692	10.848	24.561
Median	81.944	85.937	89.602	80.599	86.682	59.346
Coefficient of Assymetry	1.849	-0.061	-0.156	-0.575	1.459	1.063
Minimum Value	27.511	16.600	20.274	23.106	43.410	12.528
Lower Quartile	75.060	76.879	77.096	70.758	77.925	45.786
Upper Quartile	88.320	93.245	104.213	88.661	93.877	76.319
Maximum Value	106.342	108.099	140.015	121.327	110.933	220.342











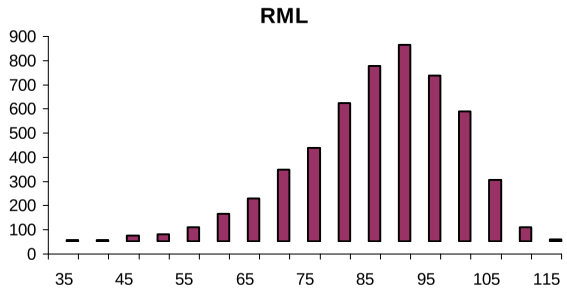
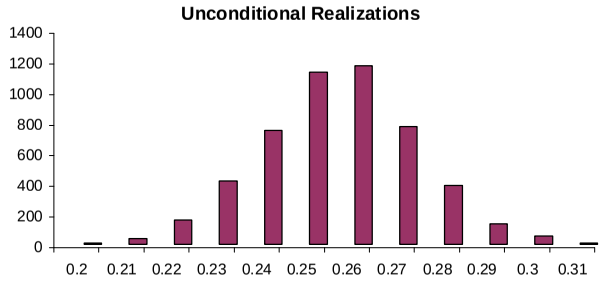
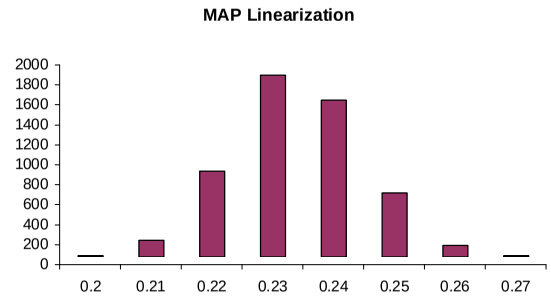


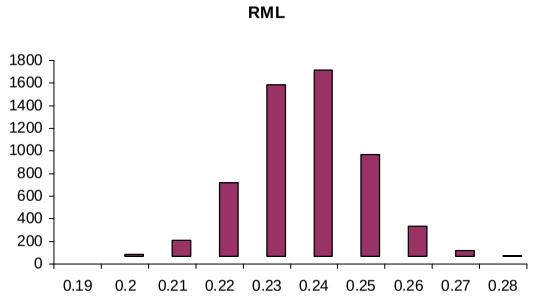
Figure 6. Histograms for Functional # 1. Case 2.

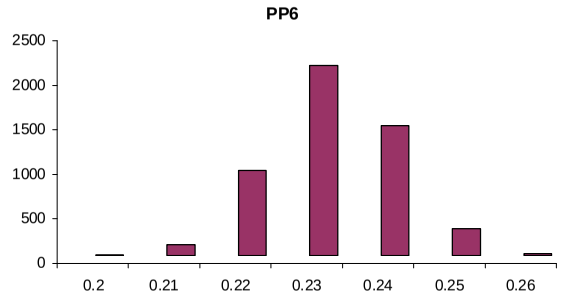
Table No. 4. Statistical Summary for Functional # 2. Case 1 (Data Variance = 25)

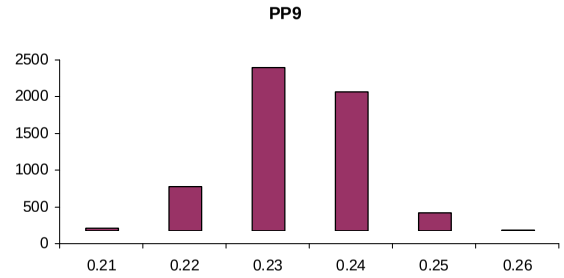
Real Value	0.23237						
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**REJ**	**UNCOND**
Number of Realizations	5198	5000	5000	5000	20000	1589	5000
Mean	0.22890	0.23172	0.22691	0.22845	0.22929	0.22859	0.24997
Standard Deviation	0.01057	0.01175	0.00869	0.00734	0.00981	0.00998	0.01704
Median	0.22873	0.23155	0.22678	0.22868	0.22901	0.22868	0.25013
Coefficient of Assymetry	0.07222	0.08322	0.05876	-0.10926	0.14404	0.00338	-0.00340
Minimum Value	0.19005	0.18837	0.19655	0.20027	0.19780	0.19729	0.18916
Lower Quartile	0.22163	0.22371	0.22111	0.22356	0.22250	0.22201	0.23866
Upper Quartile	0.23593	0.23948	0.23288	0.23339	0.23582	0.23557	0.26107
Maximum Value	0.26877	0.27735	0.25803	0.25931	0.26311	0.25930	0.30969

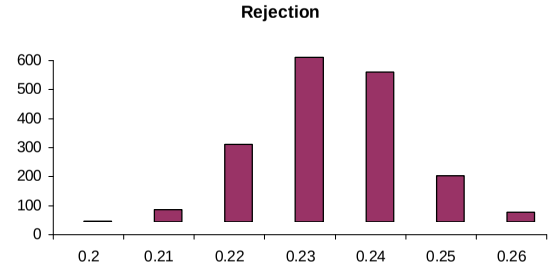












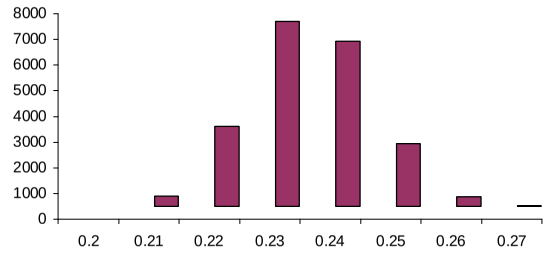
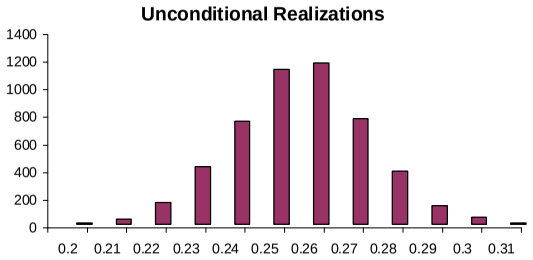
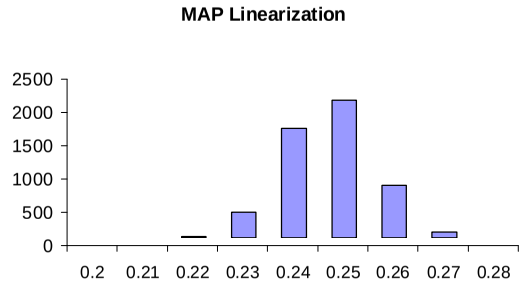


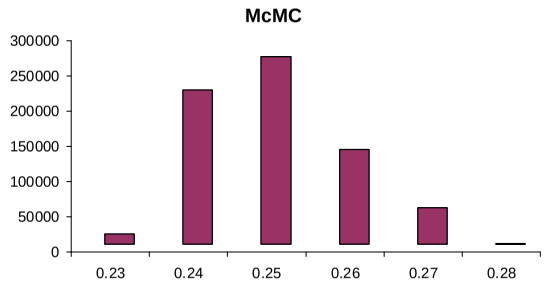
Figure 7. Histograms for Functional # 2. Case 1.

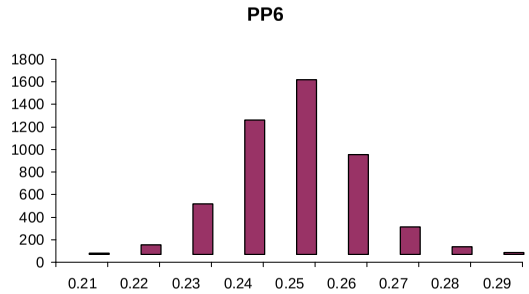
Table No. 5. Statistical Summary for Functional # 2. Case 2 (Data Variance = 0.25)

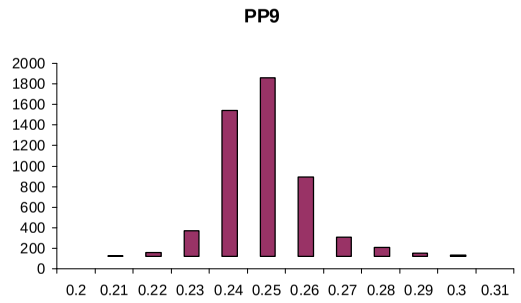
Real Value	0.2324					
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**UNCOND**
Number of Realizations	5000	5000	5000	5000	685611	5000
Mean	0.2419	0.2447	0.2442	0.2500	0.2447	0.2500
Standard Deviation	0.0086	0.0098	0.0121	0.0173	0.0088	0.0170
Median	0.2420	0.2445	0.2442	0.2502	0.2438	0.2501
Coefficient of Assymetry	2.3508	2.6564	0.1941	0.0868	3.8199	-0.0034
Minimum Value	0.2096	0.2167	0.2020	0.2023	0.2214	0.1892
Lower Quartile	0.2360	0.2378	0.2362	0.2380	0.2383	0.2387
Upper Quartile	0.2478	0.2515	0.2514	0.2609	0.2505	0.2611
Maximum Value	0.2731	0.2782	0.2893	0.3097	0.2713	0.3097











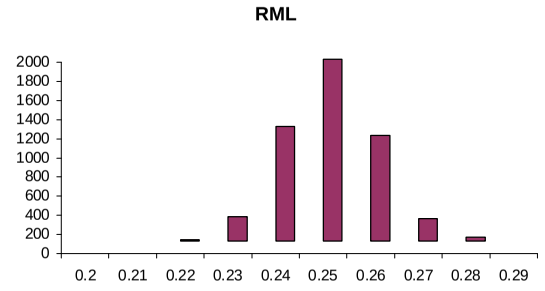
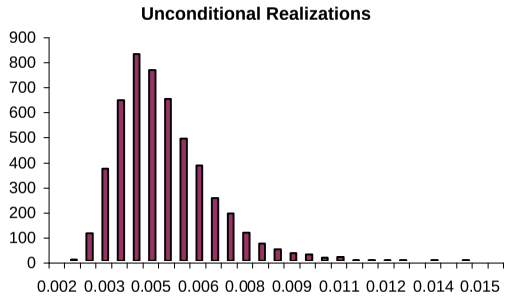
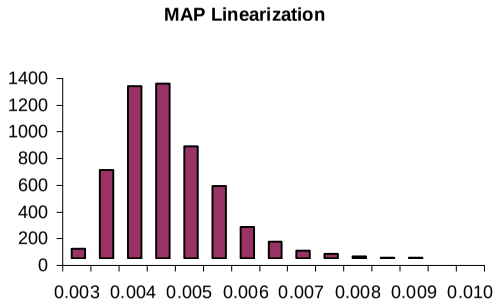


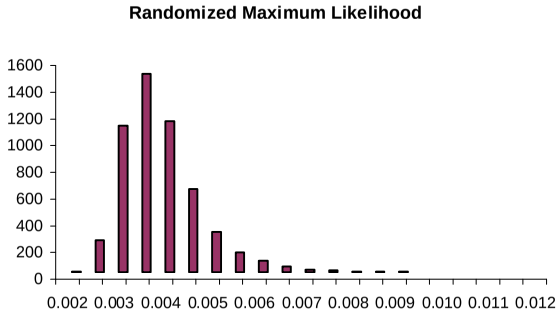
Figure 8. Histograms for Functional # 2. Case 2.

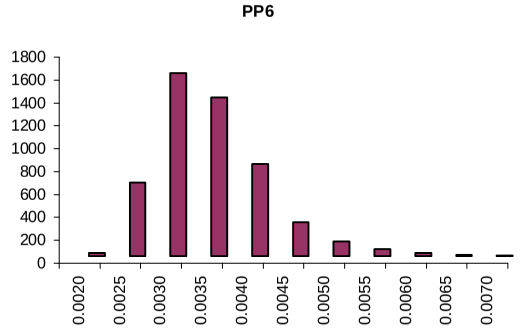
Table No. 6. Statistical Summary for Functional # 3. Case 1 (Data Variance = 25)

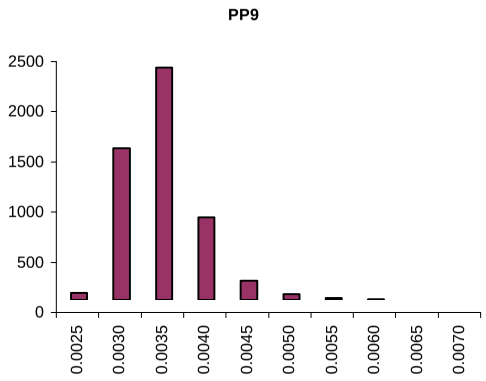
Real Value	0.004810						
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**REJ**	**UNCOND**
Number of Realizations	5198	5000	5000	5000	20000	1589	5000
Mean	0.0038535	0.0035745	0.0031859	0.0032373	0.0034727	0.0035281	0.0041210
Standard Deviation	0.0008710	0.0008593	0.0007052	0.0004603	0.0007369	0.0007775	0.0014902
Median	0.0037121	0.0034270	0.0030763	0.0031750	0.0033457	0.0033922	0.0038569
Coefficient of Assymetry	1.2654	1.5629	1.4490	1.6724	1.3063	1.4973	1.3425
Minimum Value	0.0021666	0.0018442	0.0016397	0.0021800	0.0017664	0.0019708	0.0010830
Lower Quartile	0.0032463	0.0029851	0.0026983	0.0029357	0.0029625	0.0029819	0.0030731
Upper Quartile	0.0043097	0.0039790	0.0035410	0.0034529	0.0038322	0.0039129	0.0048954
Maximum Value	0.0107456	0.0116208	0.0096962	0.0084718	0.0079033	0.0086114	0.0151270

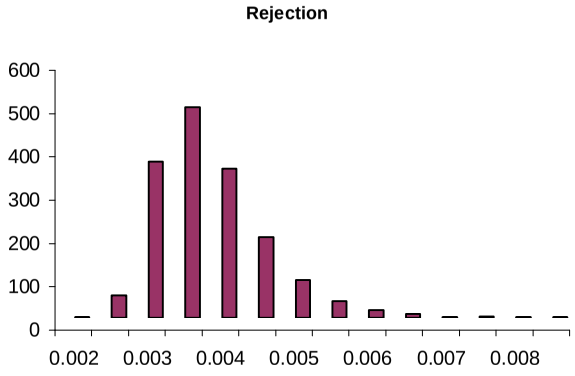












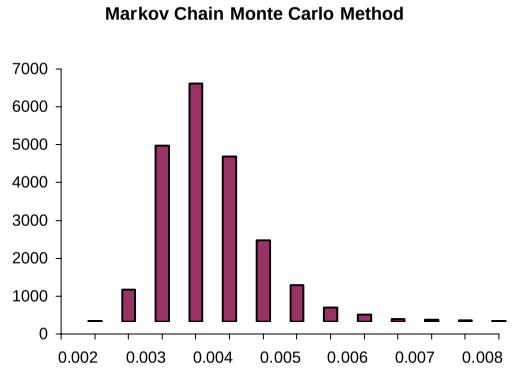
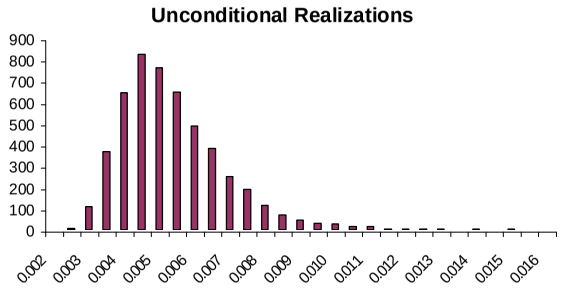
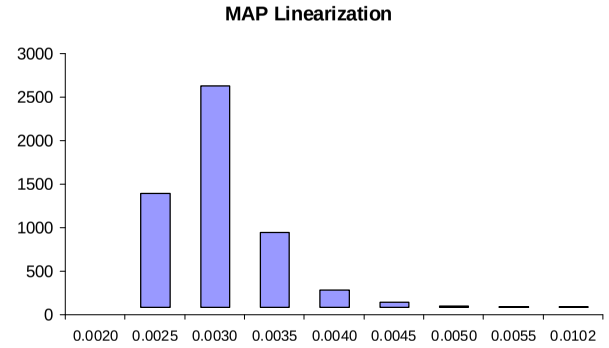


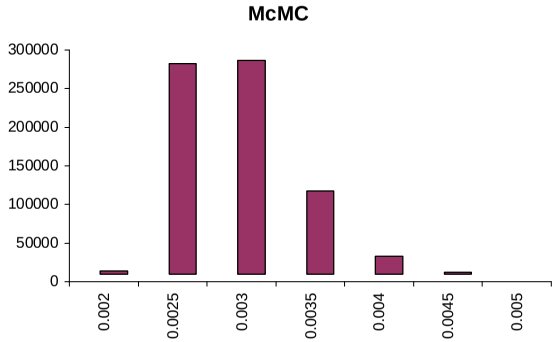
Figure 9. Histograms for Functional # 3. Case 1.

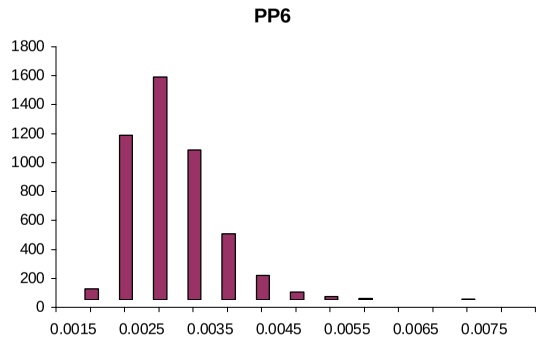
Table No. 7. Statistical Summary for Functional # 3. Case 2 (Data Variance = 0.25)

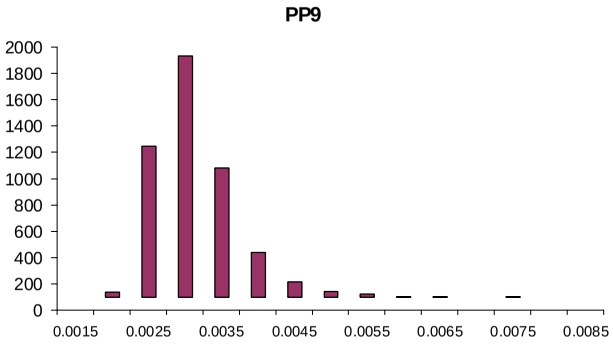
Real Value	0.004810					
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**UNCOND**
Number of Realizations	5000	5000	5000	5000	685611	5000
Mean	0.00278	0.00271	0.00243	0.00286	0.00262	0.00412
Standard Deviation	0.00043	0.00054	0.00062	0.00058	0.00038	0.00149
Median	0.00270	0.00259	0.00233	0.00276	0.00256	0.00386
Coefficient of Assymetry	2.81108	3.42631	1.50781	1.80811	1.81235	1.34248
Minimum Value	0.00197	0.00183	0.00123	0.00165	0.00191	0.00108
Lower Quartile	0.00249	0.00236	0.00197	0.00246	0.00234	0.00307
Upper Quartile	0.00297	0.00292	0.00276	0.00313	0.00285	0.00490
Maximum Value	0.00714	0.00874	0.00779	0.00831	0.00463	0.01513











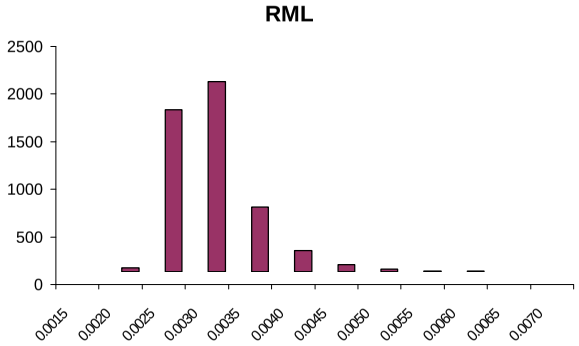


Figure10. Histograms for Functional # 3. Case 2.

Table No. 8. Statistical Summary for Functional # 4. Case 1 (Data Variance = 25)

Real Value	319.5488						
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**REJ**	**UNCOND**
Number of Realizations	5198	5000	5000	5000	20000	1589	5000
Mean	574.3524	600.9704	964.3010	850.3969	503.9249	484.4336	633.0011
Standard Deviation	286.1131	300.6585	1117.5088	864.9985	308.3283	270.5075	484.0640
Median	500.8280	526.2720	671.4998	611.4696	421.6475	419.9539	510.3867
Coefficient of Assymetry	2.7649	3.1813	8.2898	5.9597	4.1991	5.9663	5.0793
Minimum Value	190.8471	203.2533	204.5125	215.8020	220.6428	227.8605	102.4663
Lower Quartile	388.9389	415.3576	480.9463	444.2266	351.8479	350.5063	352.7150
Upper Quartile	675.6919	691.3722	1032.6158	916.8884	522.9447	527.2549	760.6333
Maximum Value	3571.8074	4791.9437	28416.4914	13622.0446	4059.5915	4648.4786	10868.7395

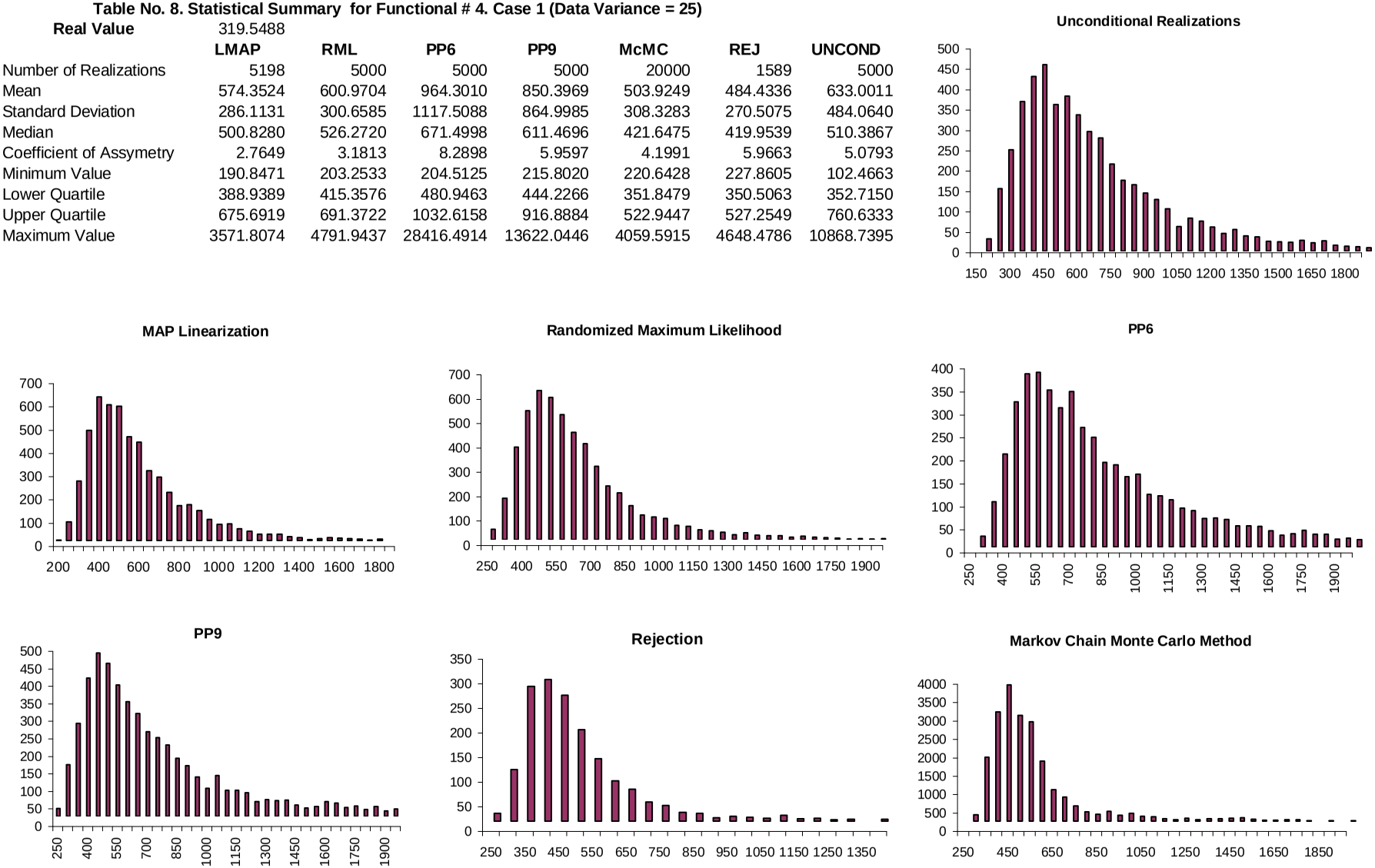
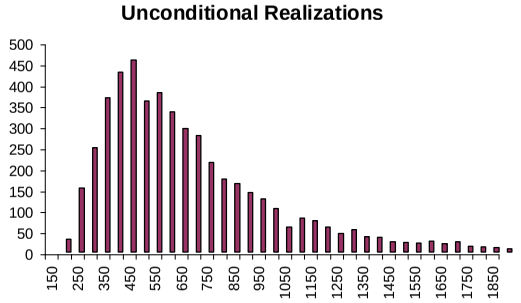
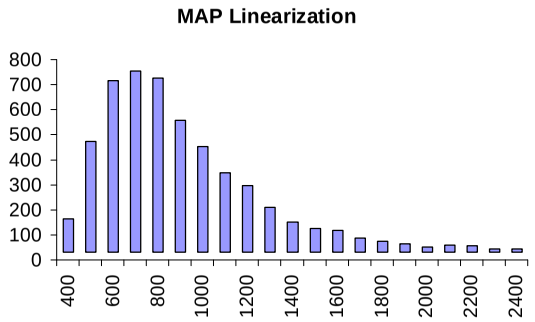


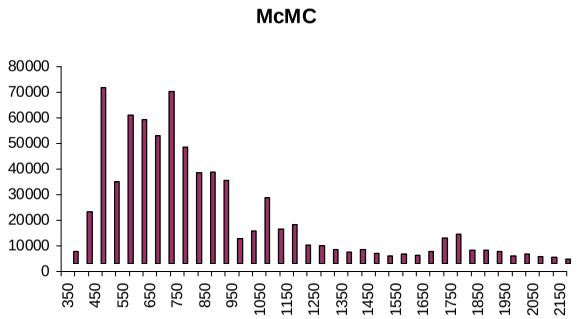
Figure 11. Histograms for Functional # 4. Case 1.

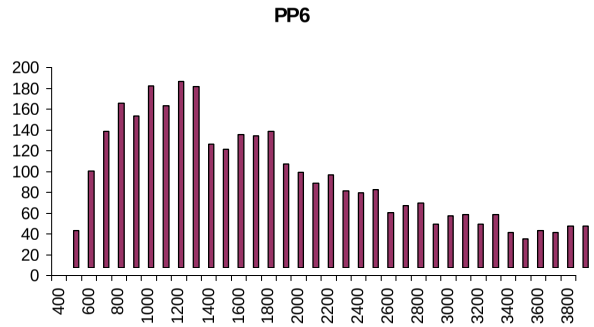
Table No. 9. Statistical Summary for Functional # 4. Case 2 (Data Variance = 0.25)

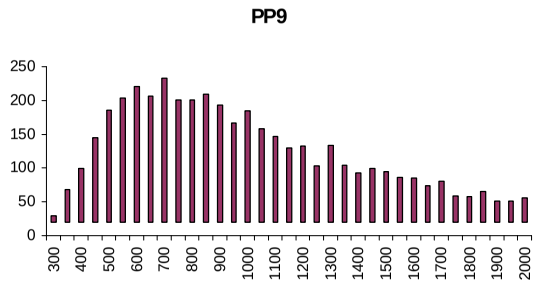
Real Value	319.5488					
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**UNCOND**
Number of Realizations	5000	5000	5000	5000	685611	5000
Mean	887.8963	806.9534	5984.5612	1666.8432	823.8827	633.0011
Standard Deviation	466.2531	448.9366	11310.2131	2679.3599	403.1151	484.0640
Median	771.8593	684.1683	2072.5129	945.9245	690.9491	510.3867
Coefficient of Assymetry	0.5550	0.7020	4.1517	7.7473	1.0778	5.0793
Minimum Value	265.9260	295.0254	330.0055	274.7939	329.4430	102.4663
Lower Quartile	598.5410	548.5412	1147.0742	676.5456	541.7629	352.7150
Upper Quartile	1037.4563	899.9986	4978.5065	1567.4458	947.0584	760.6333
Maximum Value	6321.2001	4423.1772	98317.6775	44339.9433	2948.5091	10868.7395











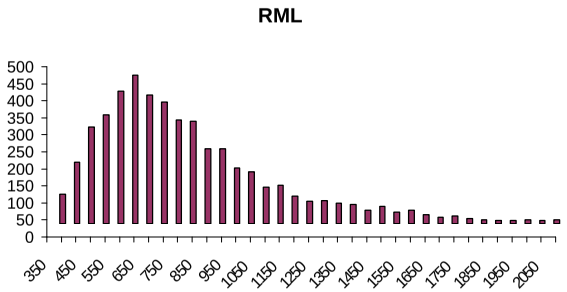
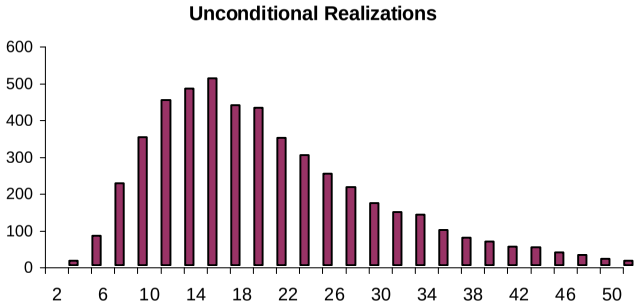
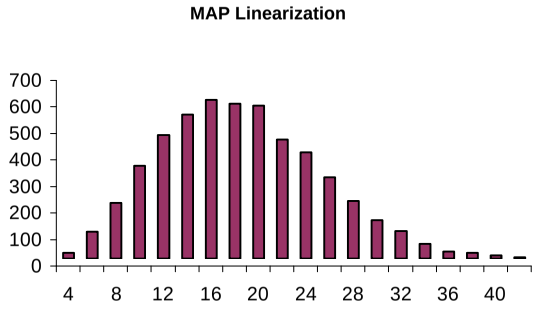


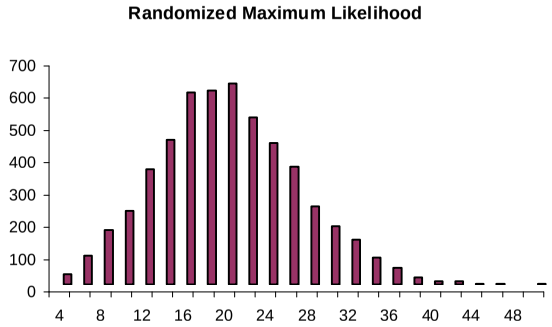
Figure 12. Histograms for Functional # 4. Case 2.

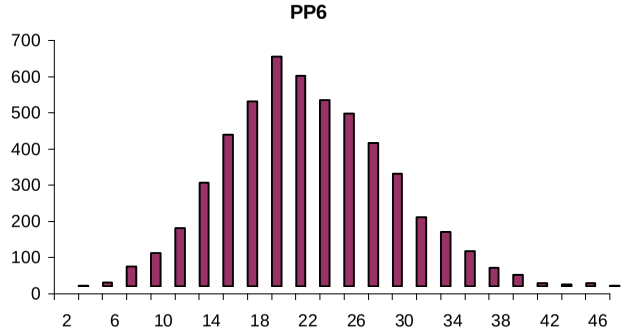
Table No. 10. Statistical Summary for Functional # 5. Case 1 (Data Variance = 25)

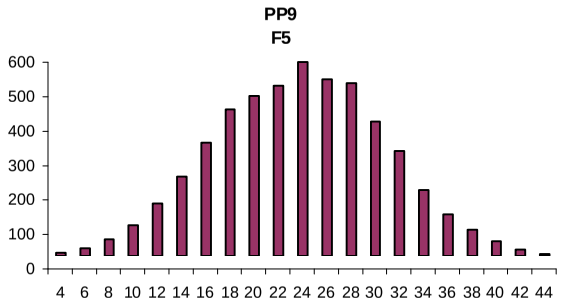
Real Value	9.9604						
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**REJ**	**UNCOND**
Number of Realizations	5198	5000	5000	5000	20000	1589	5000
Mean	17.5376	18.6384	19.5911	22.7562	22.2042	22.0723	18.0262
Standard Deviation	6.7150	6.8571	6.7015	7.0378	8.3328	8.2890	10.3454
Median	17.0186	18.2122	19.0817	22.8200	22.0355	21.9671	15.8529
Coefficient of Assymetry	0.4237	0.3109	0.3096	0.0225	0.2300	0.0745	1.3718
Minimum Value	1.4295	1.3582	1.8791	2.6751	2.5934	2.0380	1.2827
Lower Quartile	12.5946	13.9035	14.9020	17.8015	16.0318	15.8921	10.5534
Upper Quartile	21.9659	23.0250	24.0228	27.6426	27.8311	27.9994	23.1309
Maximum Value	44.0007	49.3872	46.2823	45.7063	50.6225	48.4396	100.2932

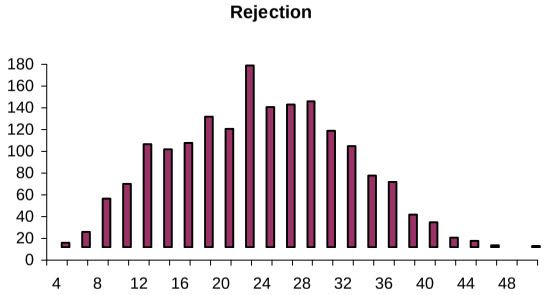












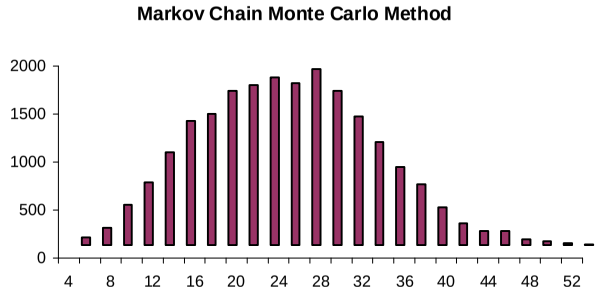
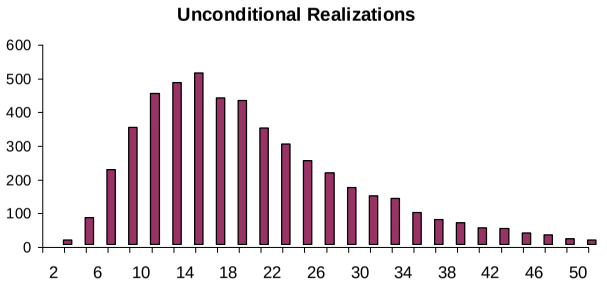
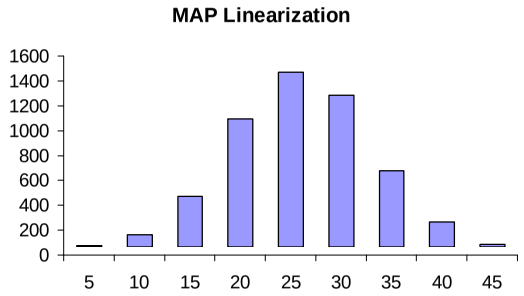


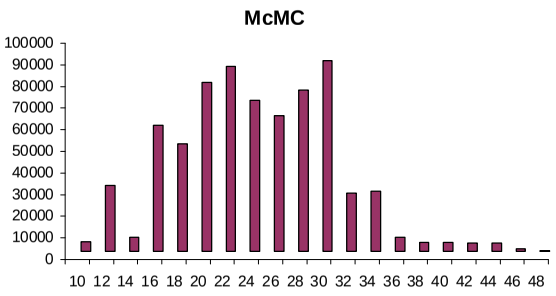
Figure 13. Histograms for Functional # 5. Case 1.

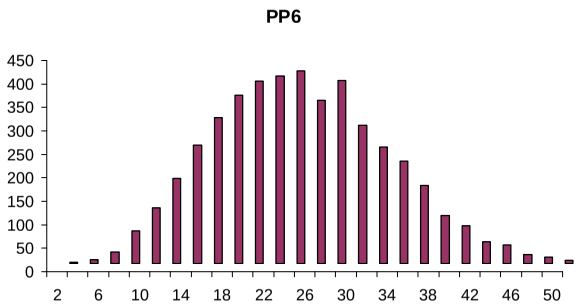
Table No. 11. Statistical Summary for Functional # 5. Case 2 (Data Variance = 0.25)

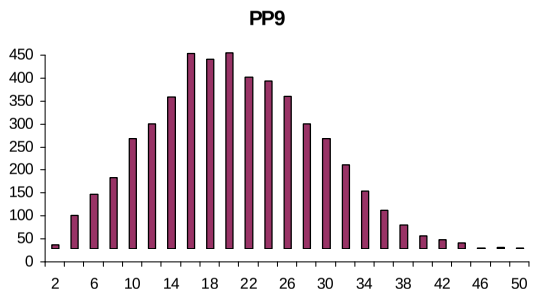
Real Value	9.9604					
	LMAP	**RML**	**PP6**	**PP9**	**McMC**	**UNCOND**
Number of Realizations	5000	5000	5000	5000	685611	5000
Mean	23.4799	22.1985	23.3794	19.9497	23.8048	18.0262
Standard Deviation	6.6693	6.7893	8.5402	8.3317	6.1580	10.3454
Median	23.4873	21.9413	23.1218	19.3923	23.8482	15.8529
Coefficient of Assymetry	1.6326	1.2673	0.2229	0.2112	2.9266	1.3718
Minimum Value	2.2915	1.2578	1.4660	1.5674	4.8854	1.2827
Lower Quartile	18.8409	17.7497	17.3475	13.9979	19.5661	10.5534
Upper Quartile	28.0431	26.0325	29.1682	26.0845	28.2055	23.1309
Maximum Value	44.4775	44.5930	52.3111	47.1227	47.3023	100.2932











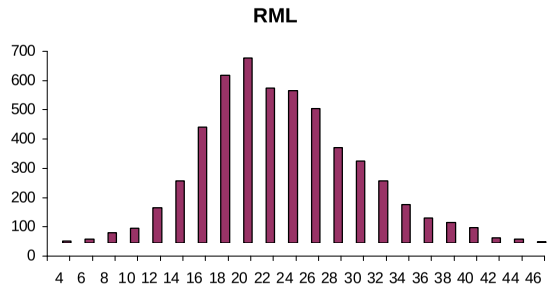


Figure 14. Histograms for Functional # 5. Case 2.

Table No. 12. Statistical Summary for SM1. Case 1 (Data Variance = 25)

	LMAP	RML	PP6	PP9	McMC	REJ	UNCOND
Number of Realizations	5198	5000	5000	5000	20000	1589	5000
Mean	20.2430	20.9743	23.2508	17.6020	17.2662	17.5717	19.9796
Standard Deviation	4.0706	4.3026	6.5884	4.7858	3.5773	3.7232	4.5697
Median	19.8604	20.6701	22.0947	16.9412	17.0320	17.2745	19.6028
Coefficient of Assymetry	0.4886	0.4882	1.0576	0.9038	0.5494	0.4555	0.4836
Minimum Value	9.3323	9.9370	9.5886	6.9636	8.6560	7.8392	6.3943
Lower Quartile	17.3928	17.8642	18.6108	14.1374	14.7215	14.8419	16.7071
Upper Quartile	22.7518	23.6644	26.7297	20.3142	19.4333	19.8814	22.7759
Maximum Value	35.8446	42.2150	57.2766	46.1154	34.7209	34.1649	41.8829

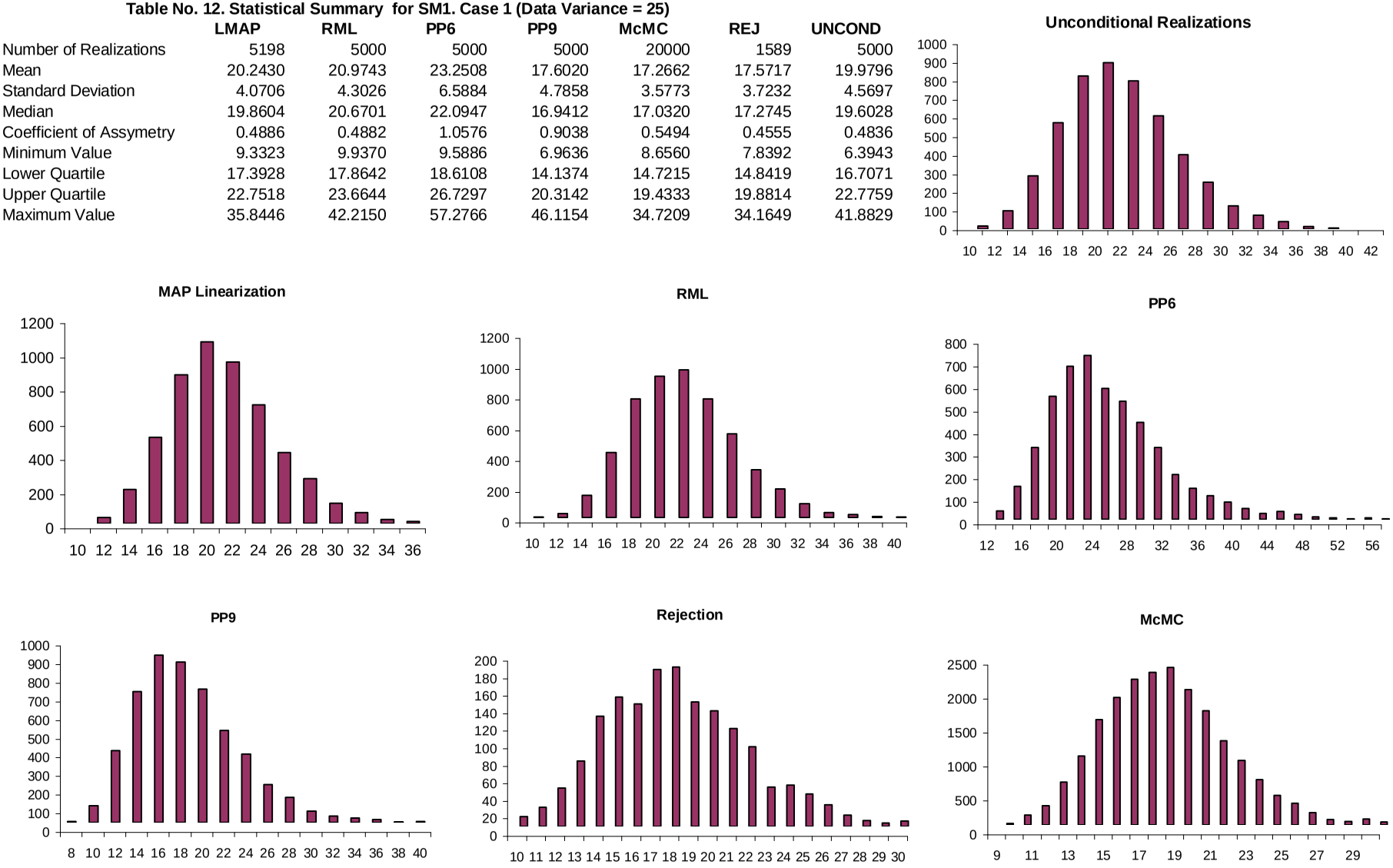
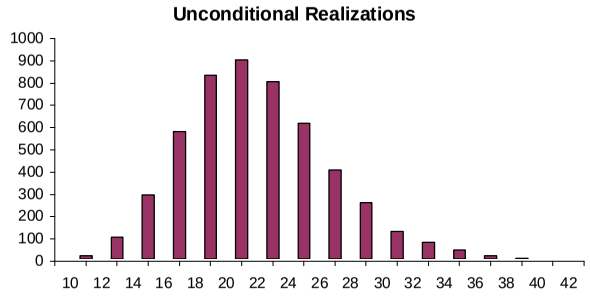
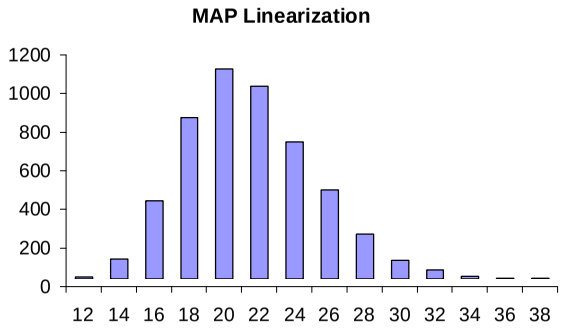


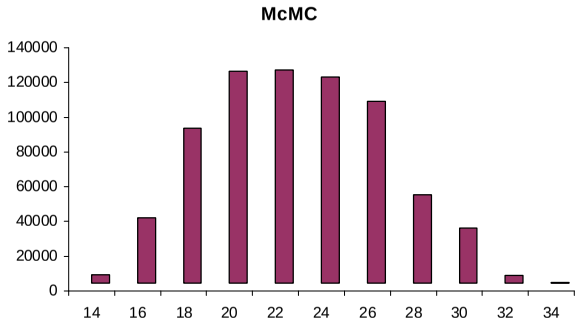
Figure 15. Histograms for Model Mismatch wrt Prior Model for Case 1.

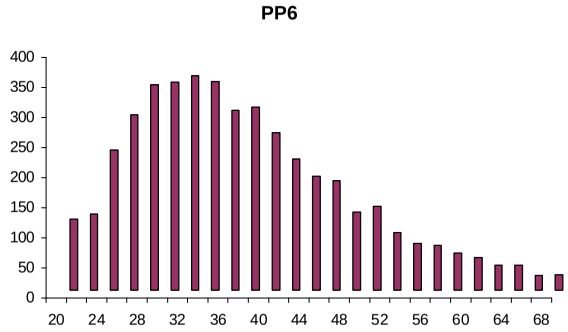
Table No. 13. Statistical Summary for SM1. Case 2 (Data Variance = 0.25)

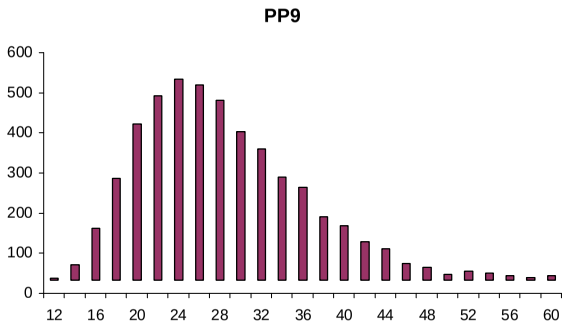
	LMAP	RML	PP6	PP9	McMC	UNCOND
Number of Realizations	5000	5000	5000	5000	685611	5000
Mean	20.4803	21.2178	36.7526	27.2348	21.2381	19.9796
Standard Deviation	3.7255	3.8389	11.9209	8.5460	3.6009	4.5697
Median	20.1327	20.9408	34.2938	25.5894	20.9494	19.6028
Coefficient of Assymetry	1.8759	1.8734	1.1743	1.1098	0.9651	0.4836
Minimum Value	11.2055	11.2435	16.4114	11.7907	12.3546	6.3943
Lower Quartile	17.7898	18.6258	27.9686	21.3351	18.3399	16.7071
Upper Quartile	22.7477	23.5659	43.2716	31.7374	23.8844	22.7759
Maximum Value	36.6510	40.8427	97.3897	65.8264	32.7346	41.8829











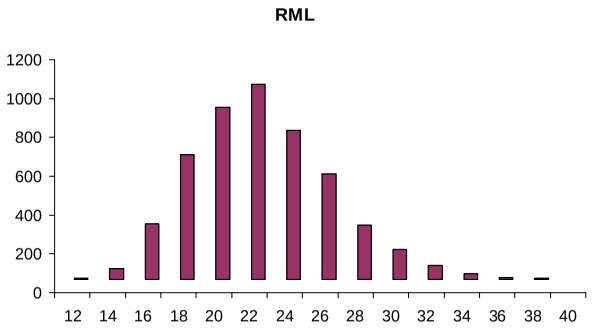
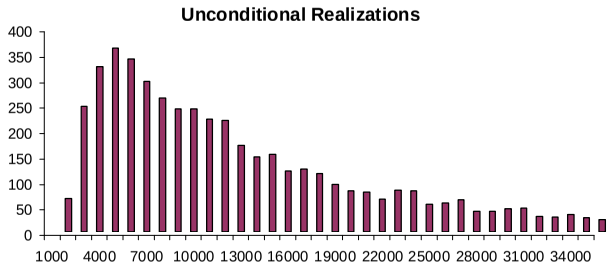
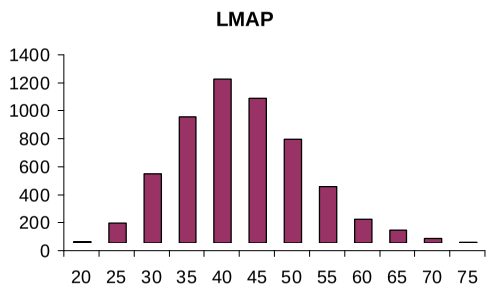


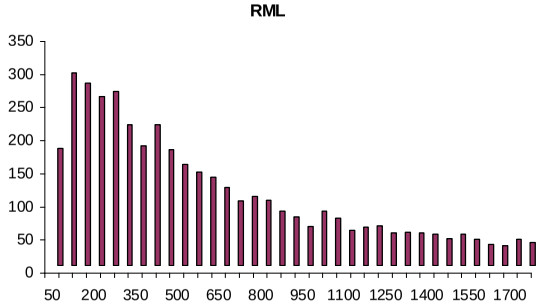
Figure 16. Histograms for Model Mismatch wrt Prior Model for Case 2.

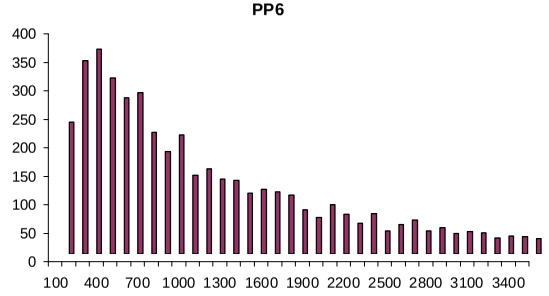
Table No. 14. Statistical Summary for SM2. Case 1 (Data Variance = 25)

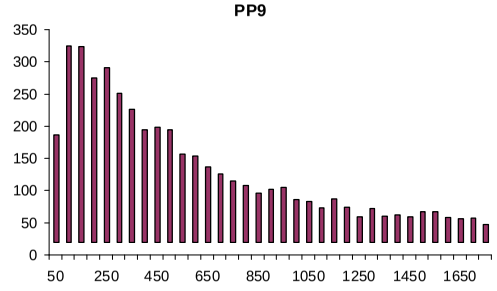
	LMAP	RML	PP6	PP9	McMC	REJ	UNCOND
Number of Realizations	5198	5000	5000	5000	20000	1589	5000
Mean	40.1366	1242.0814	2003.0619	1375.1526	19.7877	18.0520	14008.7797
Standard Deviation	8.9788	1556.4602	3722.1661	2434.2517	5.3481	4.3912	14032.2929
Median	39.5727	610.2628	962.4165	582.6721	19.1338	17.6927	9536.5029
Coefficient of Assymetry	0.4333	0.9327	8.3294	5.4811	0.9368	0.4641	2.4532
Minimum Value	13.70	12.4242	16.28	7.38	7.28	6.83	287.27
Lower Quartile	33.76	247.0743	402.34	239.13	16.12	14.78	4783.68
Upper Quartile	45.87	1545.9839	2204.58	1480.02	22.70	20.84	18149.78
Maximum Value	77.25	7985.1997	84302.56	37591.70	44.91	34.48	116071.10

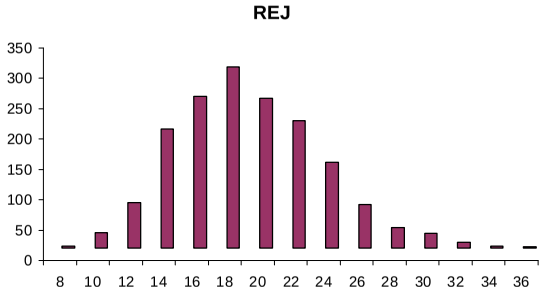












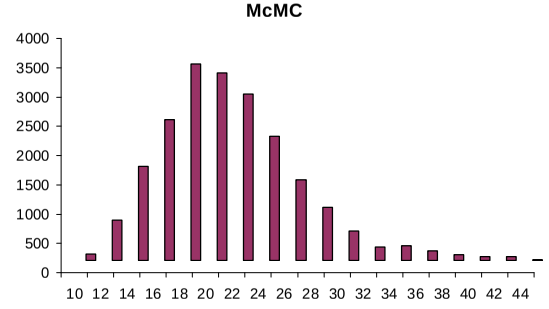


Figure17. Histograms for Model Mismatch wrt MAP for Case 1.

Table No. 15. Statistical Summary for SM2. Case 2 (Data Variance = 0.25)

	LMAP	RML	PP6	PP9	McMC
Number of Realizations	5000	5000	5000	5000	685611
Mean	20.04	154486.81	152431.11	139857.00	165096.86
Standard Deviation	4.477	265727.350	256005.273	225248.186	220781.773
Median	19.72	58842.03	71473.29	67417.85	78024.40
Coefficient of Assymetry	1.529246483	0.928124099	5.585019307	5.290123807	0.833
Minimum Value	7.324907	113.279205	509.537506	579.207825	1272.188
Lower Quartile	16.86715625	22143.04492	28416.67187	25961.05176	27617.824
Upper Quartile	22.93731325	148477.5313	176180.9531	160184.375	176650.550
Maximum Value	39.35498	1858867.5	3389717.75	3314385.75	959703.938

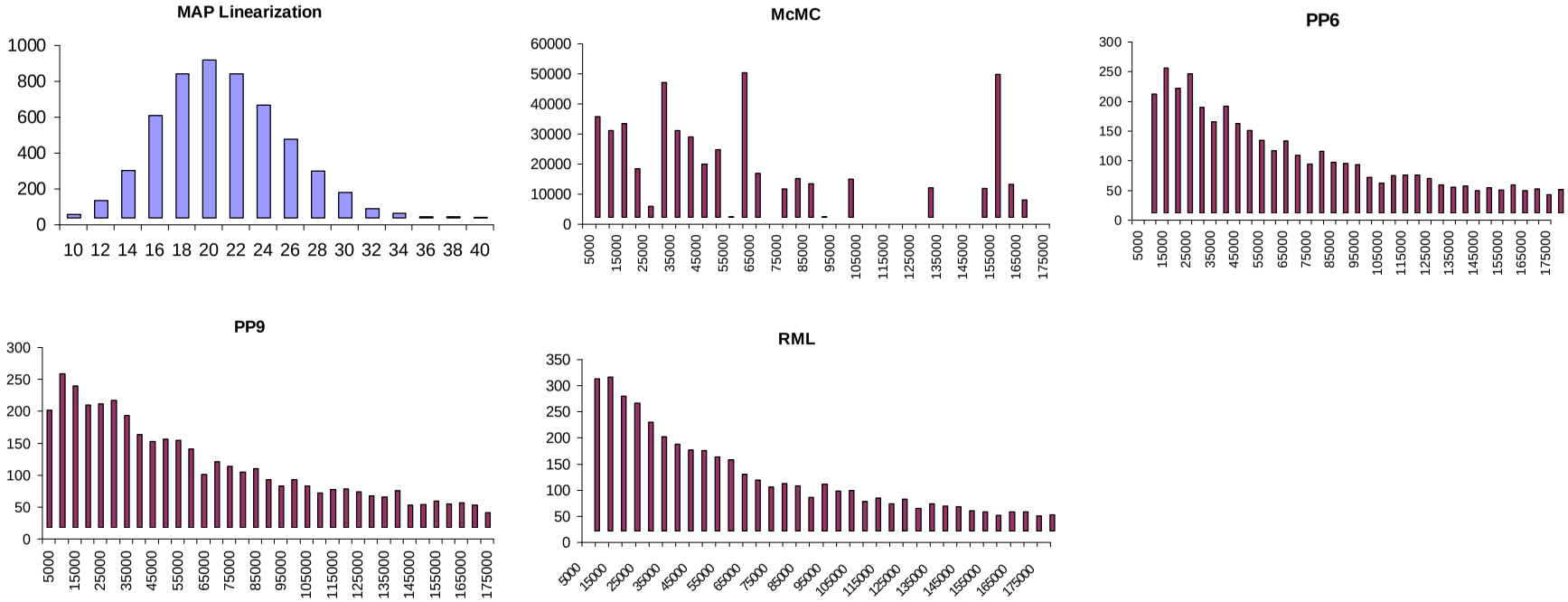
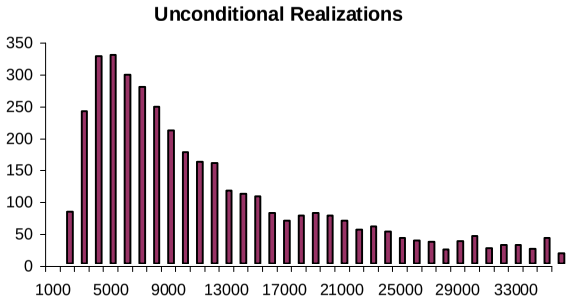
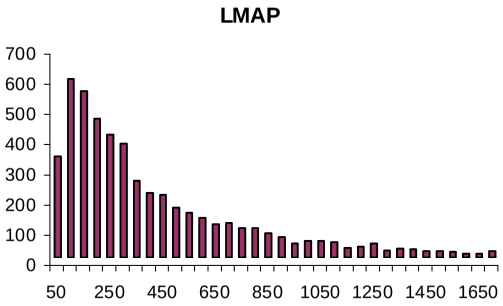


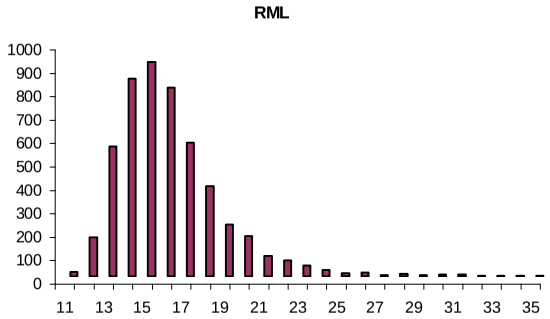
Figure 18. Histograms for Model Mismatch wrt MAP for Case 2.

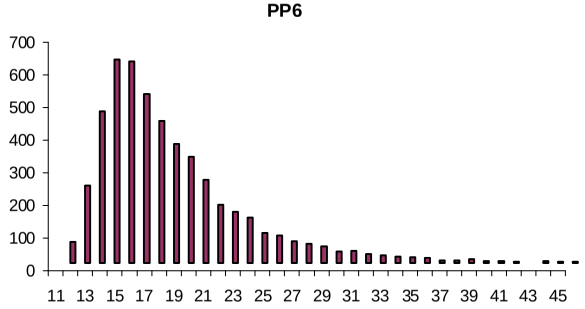
Table No. 16. Statistical Summary for SD1. Case 1 (Data Variance = 25)

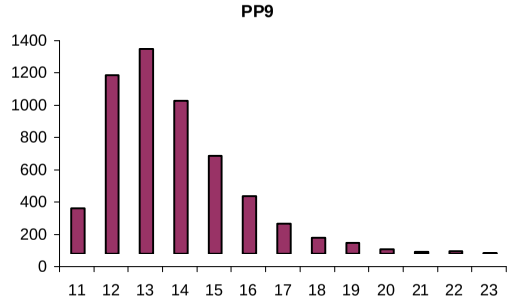
	LMAP	RML	PP6	PP9	McMC	REJ	UNCOND
Number of Realizations	5198	5000	5000	5000	20000	1589	5000
Mean	504.6221	15.6799	17.4047	13.3005	14.9104	15.5709	40107.1608
Standard Deviation	634.4409	3.5692	5.6102	2.1210	1.7498	2.2105	85796.1585
Median	283.7997	14.9953	15.9334	12.8475	14.5986	15.2631	11020.0879
Coefficient of Assymetry	3.3009	1.4514	2.6905	2.7662	0.9830	0.8483	4.8290
Minimum Value	13.3769	10.0545	10.1011	10.0242	12.3783	10.8179	137.7042
Lower Quartile	133.5492	13.6213	13.7677	11.8718	13.5649	13.9523	4932.4900
Upper Quartile	622.5130	16.7547	19.3228	14.2059	15.8962	16.8064	33020.3457
Maximum Value	6964.6777	39.6816	88.3370	38.3798	22.2324	27.2165	929202.3750

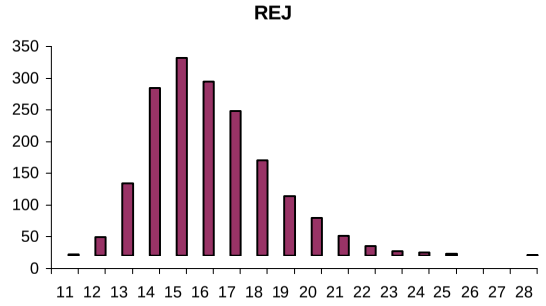












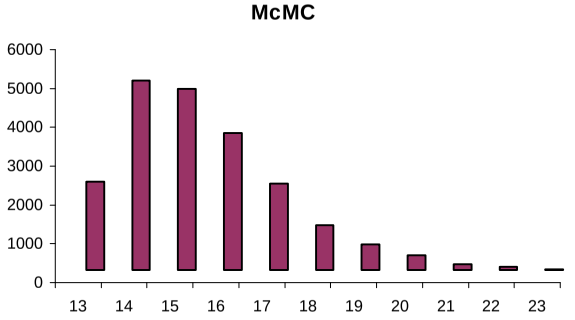
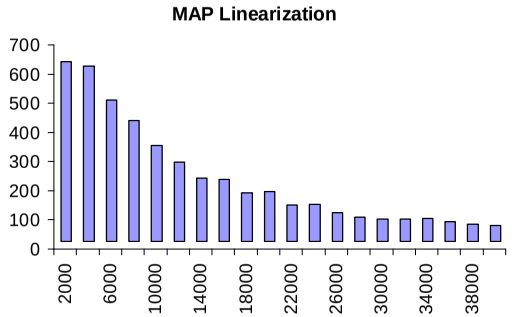
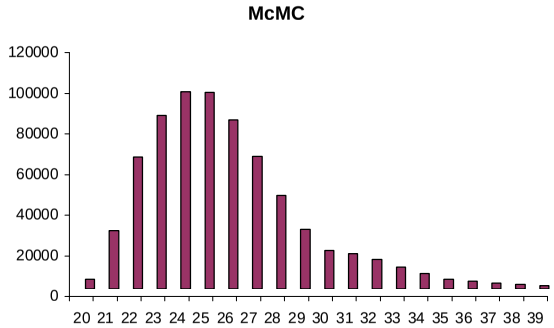


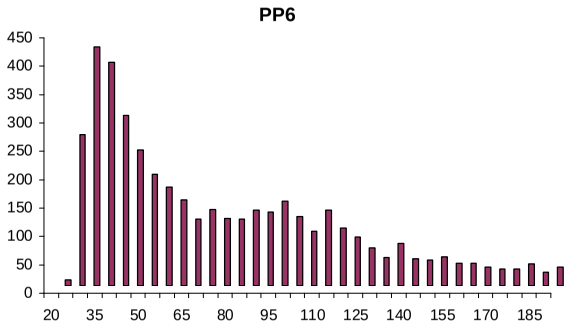
Figure 19. Histograms for Data Mismatch for Case 1.

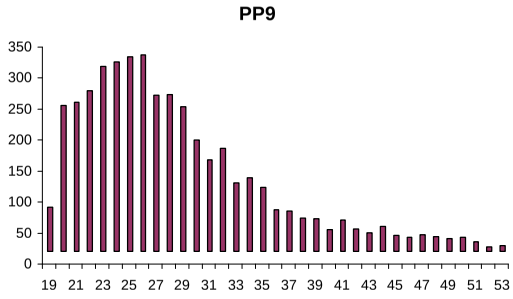
Table No. 17. Statistical Summary for SD1. Case 2 (Data Variance = 0.25)

	LMAP	RML	PP6	PP9	McMC
Number of Realizations	5000	5000	5000	5000	685611
Mean	20244.5737	25.8633	79.1571	35.3506	25.2902
Standard Deviation	30447.9615	4.6549	52.5729	23.0412	3.3534
Median	10347.5908	24.7608	62.1676	29.0401	24.7071
Coefficient of Assymetry	0.9398	2.6766	1.2588	5.5596	1.8054
Minimum Value	68.7248	18.5743	20.6426	19.5094	18.7483
Lower Quartile	4115.4503	22.7625	36.9437	25.3418	23.0090
Upper Quartile	24230.8218	27.4383	106.1218	35.4387	26.8048
Maximum Value	727702.0	50.9052	291.00	325.517	169.7134









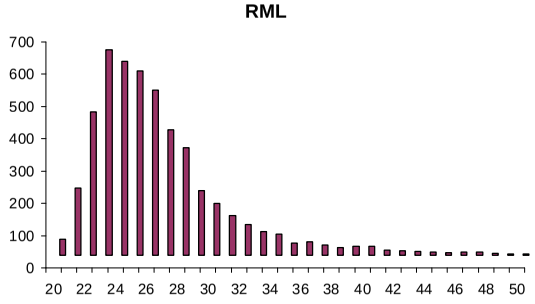


Figure 20. Histograms for Data Mismatch. Case 2.