Universitetet i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

Study programme/specialisation: Petroleum Geoscience Engineering	Autumn semester 2019
Author: Ali Rahimi	(signature of author) Ali Rahimi
Programme coordinator Supervisor(s): Professor Reidar Brumer Bratvold	
Title of master's thesis: Optimization of Well Location in a 5-Spot Pattern	
Credits: 30	
Keywords: Optimization Optimizer's curse Overestimation Uncertainty reduction Expected disappointment	Number of pages: 76 + supplemental material/other:

In fulfilment of the master's degree at Faculty of Science and Technology

Optimization of Well Location in a 5-Spot Pattern

Ali Rahimi

Faculty of Science and Technology

University of Stavanger

December 2019

Left Blank Intentionally

Acknowledgement

The completion of this complex thesis was not easy without participation and assistance of the people who in a way contributed to this thesis. It was exciting and challenging to work on such a new subject.

I am grateful and would like to express my appreciation and especial thanks to my thesis advisor professor Reidar B. Bratvold for giving me this opportunity to gain an insight and educate in such a different and interesting topic. I hope my research could have contributions to the oil industry and the other industries who would like to decide for the future plans.

I would also like to thank my friend Peyman Kor for all help.

Finally, I must express my very profound gratitude to my wife for providing me with great support and unceasing encouragement throughout my education and during researching for my thesis. This accomplishment would not have been possible without her.

Abstract

Deciding to choose the best alternative from a set of courses of action is a repeatable activity in most organizations (Ahti Salo et al., 2011). Companies, for example, commence some research and development projects for future investment in a business that can generate potential profit like late-stage development investments in a pharmaceutical company or taking a decision about constructing different facilities for educational services. Although these issues look different, the decision-maker should choose the portfolio of actions base on some limitations such as the feasibility of the subject, availability of resources, etc. And choose the alternatives which can contribute more to the outcome and maximize value creation. In all of them, the decision-maker should select, arrange, and allocate resources. in other words, managing the portfolio (Summers, 2019).

Usually, the normal rule for a decision-maker in a decision situation for choosing between different alternatives and maximize the value is to select the one with the maximum estimated value. These estimates are coupled with uncertainty (prediction errors). Having errors in value estimates and selecting the highest estimated value among them, induces a systematic bias that guarantees, over repeated decisions, less than the estimated expected value will be realized (Begg and Bratvold, 2008). And the value estimate for the recommended action to be biased high. This biased is called “optimizer’s curse”.

In the optimization of the well location to maximize value creation, as measured by NPV, injection and production wells must be in the optimal locations. The process of optimizing well locations include a range of uncertain factors and requires a robust (stochastic) optimization approach. With these uncertain factors, NPV values are subjected to error. By coupling such estimated measures of values with the optimization-based selection process the alternatives that values which have been overestimated most, are more likely to be selected. It is not because of any bias in the estimates themselves, but the optimization process, which simply ranks the value estimates and selects the highest estimated value among them.

In this work, we will study the optimizer’s curse in the context of optimizing a 5-spot pattern and indicate how the process of optimization leads to select the alternative with overestimated value. Further on we will develop a Bayesian model to correct these value estimates.

Contents

Left blank intentionally

Chapter 1. Introduction

1.1 Motivation

Many projects in Norwegian Continental Shelf (NCS) end-up producing less than what was forecasted. Generally, in oil industry, it is common to overestimate the production volume and underestimate the uncertainties involving in geological aspects of a reservoir. It is partly due to an optimistic view and overconfidence in uncertainty quantification, and strategies of the oil companies which can support their decision policy. A part of this overestimation comes from optimization process itself which leads to select the alternatives with higher probability of being over estimated.

Over the past decades, despite of introducing new technologies, collecting more data and improving the data quality, the ability to forecast the production has not improved so much (Mohus, 2018). As an example, Nandurdikar & Wallace used data derived from offshore oil and gas production database, which was developed and maintained by Independent Project Analysis Inc. (IPA) to indicate the deficiencies in production rate. In 1995, they showed that oil and gas projects delivered almost 94 percent of the estimated production, but by the time they published the result, it was only delivered 75 barrels instead of 100 barrels promised at the sanction time. They showed that optimistic subsurface evaluation leaded to overestimated predictions and consequently less production rate. (Nandurdikar and Wallace, 2011). Furthermore, there was a huge gap between the historical experience and predictions which was skewed towards overestimation (Figure 3.1).

Usually, in decision analysis process, alternatives are ranked by their value estimates and it recommends selecting alternatives with highest estimated values. Any decision-making attempt to optimize these estimates which consistently selecting alternatives based on the estimated values, leads the value estimate of the recommended alternative to be biased high. This phenomenon is called optimizer's curse. (Smith and Winkler, 2006) This gap between the estimated value and the actual outcome may be substantial. It has a huge effect on the portfolio's true outcome. It can be half of the value estimated when we forecast in the usual manner (Schuyler and Nieman, 2007). The optimizer's curse has gained more attention recently and it seems to be less known and underappreciated. As the optimizer's curse reduces the value added by decision analysis, it should be more considered in the process of decision making.

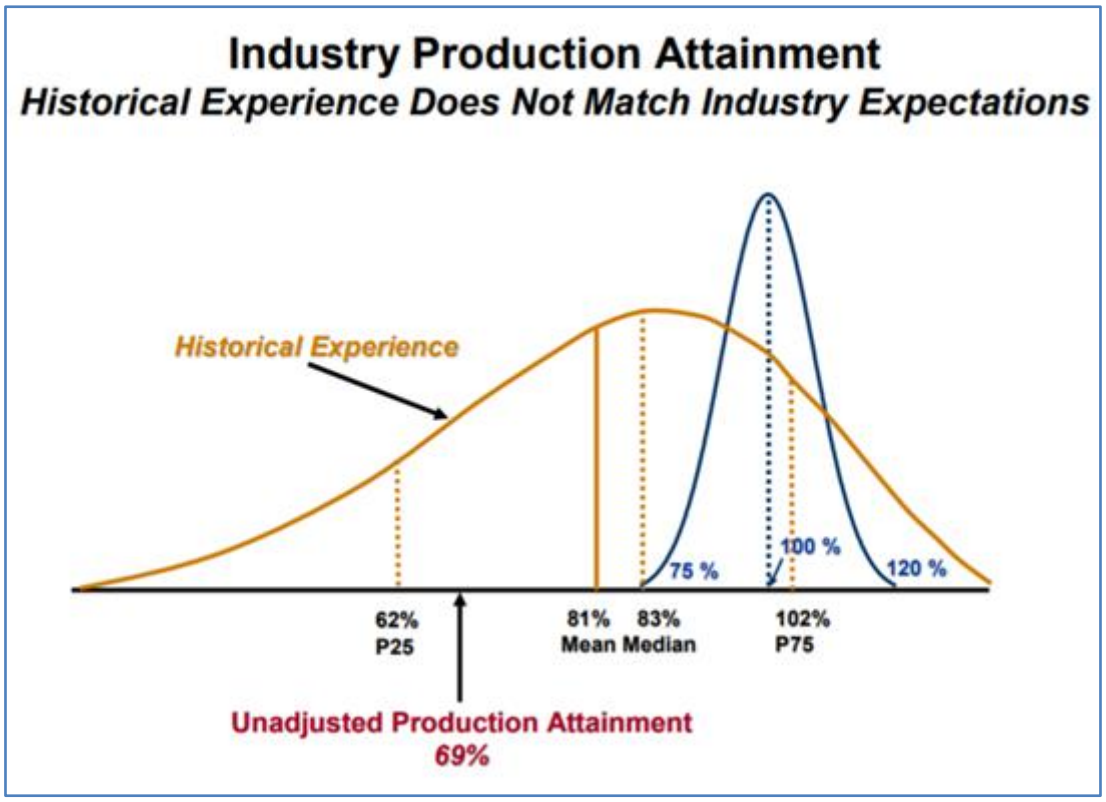

Figure 1.1: Historical and actual production attainment (Nandurdikar & Wallace, 2011)

1.2 Aim of this thesis

In this work we show how selecting and ranking among estimated measures of NPV for different well locations in the optimization-based selection process, leads to have an overestimated NPV value and then correcting these estimations by using a Bayesian model.

Objective 1: Optimization of well location and injection rate

The goal is to Optimize the well location and injection rate in a 5-spot pattern recovery method, with one injection well in the middle and four production wells scatter around. Finding the best coordinates for production wells and the best injection rate based on the highest values of NPV is the target. To obtain this goal two steps has been done:

- Generating multiple reservoir realization

The key reservoir parameter which is considered as an uncertain element in the simulation process is permeability. 100 realizations of permeability distribution are generated by using the Sequential Gaussian Simulation (SGS).

- Optimizing well location and injection rate

The optimization procedure is done by using an algorithm which is developed in R programming. The genetic algorithm, which is used in this coding, helps to generate new locations for simulation and find a better production and injection rate in each stage or iteration. In the end, the best location and injection rate for the 5-spot pattern model will be identified. The optimal location will be shown by its coordinates. Optimal injection rate will be identified by its two parameters of gamma rate and starting value.

The algorithm also lets us simulate random locations and generate the distribution of NPV for any location.

Objective 2: Study the optimizer's curse in the process of optimization

Because of uncertain parameter in the reservoir model the NPV value estimates resulted from simulations, has some errors. The goal is to indicate that optimization algorithm which ranks and selects among these value estimates leads to choose the alternatives with higher probability to be overestimated.

Objective 3: Develop a Bayesian model

To overcome the optimizer's curse, a Bayesian model is developed. The model is used to check the optimization results for the possible expected disappointment. This model can be used either to check the data from previous fields or to provide a coefficient or corrective measure to reform and adjust the production estimates for the future and ongoing projects.

1.3 Scope and limitations

Optimizer's curse is an old phenomenon which drew attention not long ago. It was not subjected to so many studies and there are only a few numbers of the resources available. Most of the work is done by Smith and Winkler and the other papers are mostly referred to them.

To My knowledge, it is the first time, that this phenomenon is studied on the process of optimization of the well location. It must be mentioned that it was a challenge to work with real data and reservoir model due to confidentiality.

As discussed with Norwegian Petroleum Direktorate (NPD), there is a lack of data for 5-spot pattern in offshore Norway to have a comparison between original estimations and actual outcome.

Chapter 2. Optimizer's curse

2.1 Introduction

Usually, a decision is made in the process of predicting the future outcome of available alternatives (Harrison and March, 1984). In the process of decision making, the normal rule of selecting among alternatives is to estimate their values and apply a decision criterion such as selecting the alternatives which has the maximum estimated values (Begg and Bratvold, 2008). Depending on the subject, different methods like net present value or expected utility can be used to generate the estimates for the alternatives.

After ranking and selecting among these value estimates based on the decision criteria, by implementing the selected alternatives, the outcome will be experienced. Sometimes the realized value of the selected alternatives is more than forecasted value and sometimes it is less. The differences between the estimated value of the selected alternative and its realized value, determined by executing that and experiencing the consequences (Harrison and March, 1984)**.** Depending on the result, decision-maker will experience post-decision positive surprise or disappointment. The decision maker who consistently selects the alternatives which have the highest estimated values, due to error in these value estimates, on average will gain a lower value than estimated, and will experience post-decision disappointment.

Even if a team of decision analysists does a complex calculation and analysis with unbiased judgments, an unpleasant bias will affect the estimations in a project portfolio assessment. The problem arises in the process of choosing the optimal alternative. Then by applying the optimal alternative, it tends to deliver less than the distribution's mean value which is forecasted in the beginning.

Even if the value estimates are unbiased, because of the uncertainty in the estimates which is coupled with optimization-based selection process, those alternatives with the highest estimated values are most likely to be selected and accepted alternatives tend to be those where random evaluation errors are optimistic. That leads the value estimate for the recommended action to be biased high. This biased is called "optimizer's curse" (Smith and Winkler, 2006). It is happened just because of optimization process not because of any bias in estimation. Optimizer's curse is a production of a statistical process. It is happened merely just by choosing the best of a set of

uncertain prospects (Marks, 2008). It describes how the simple process of screening and ranking projects introduces a bias.

2.2 Optimizer's curse interpretation

Kritzman(2011) states that the optimizing with mean-variance, taking in more errors in the optimization process. Optimization result is biased in favour of the alternatives with positive errors in the mean of the overestimated alternatives, and negative errors in the risk evaluation. So, it overstates the portfolio expected returns and leads to choose the wrong portfolio of action (Kritzman, 2011).

The estimated values that are derived from limited data are never precisely known and always tangled with the errors. In other words, fundamental input for the model optimization process has errors inside. By taking in the uncertain estimates in the optimization process, errors will be transferred to the optimization results. When the input data is bad, the result will be poor. Garbage in garbage out.

Daniel Kuhn explained how the errors in estimations lead to gain post-decision disappointment. Considering 10 different alternatives from A to J (figure 2.1), like ten different NPV's which are calculated from 10 different locations for injection and production wells, with a limited amount of budget and time, just five of them can be selected and executed. The green bars indicate the expected net present value of each alternative. (Kuhn, 2018)

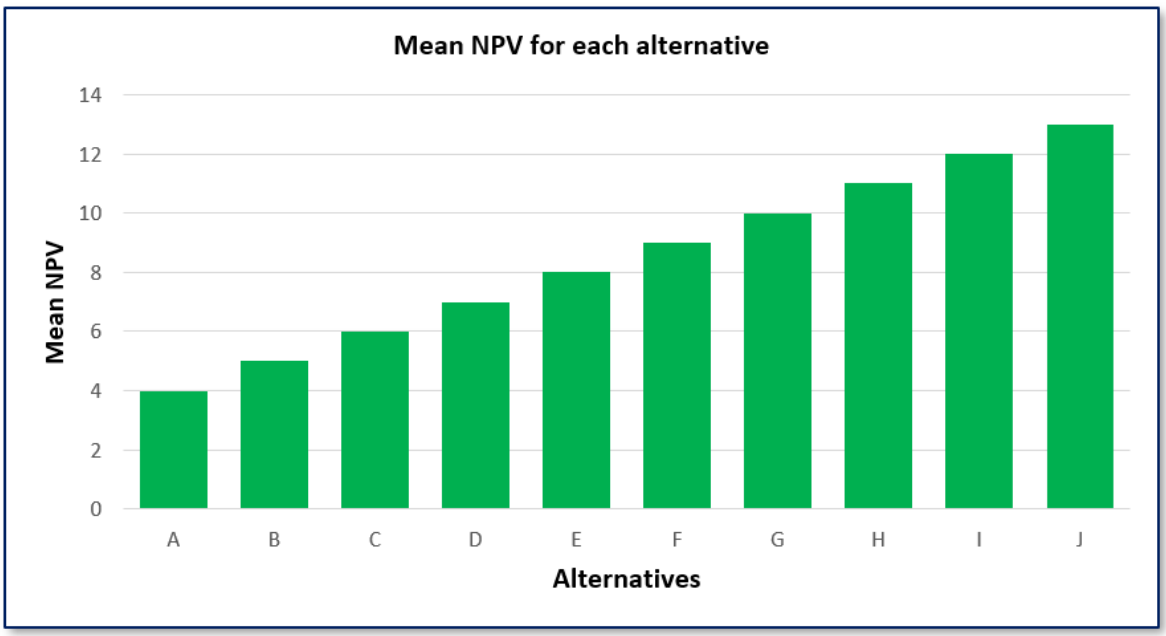

Figure 2.1. Ten different alternatives

Given this information, alternatives with the highest expected net present value will be chosen to maximize the average of the expected net present value of the selected alternatives. As it is indicated in the figure below alternatives from F to J will be picked and the mean of the expected net present value is indicated by the blue line.

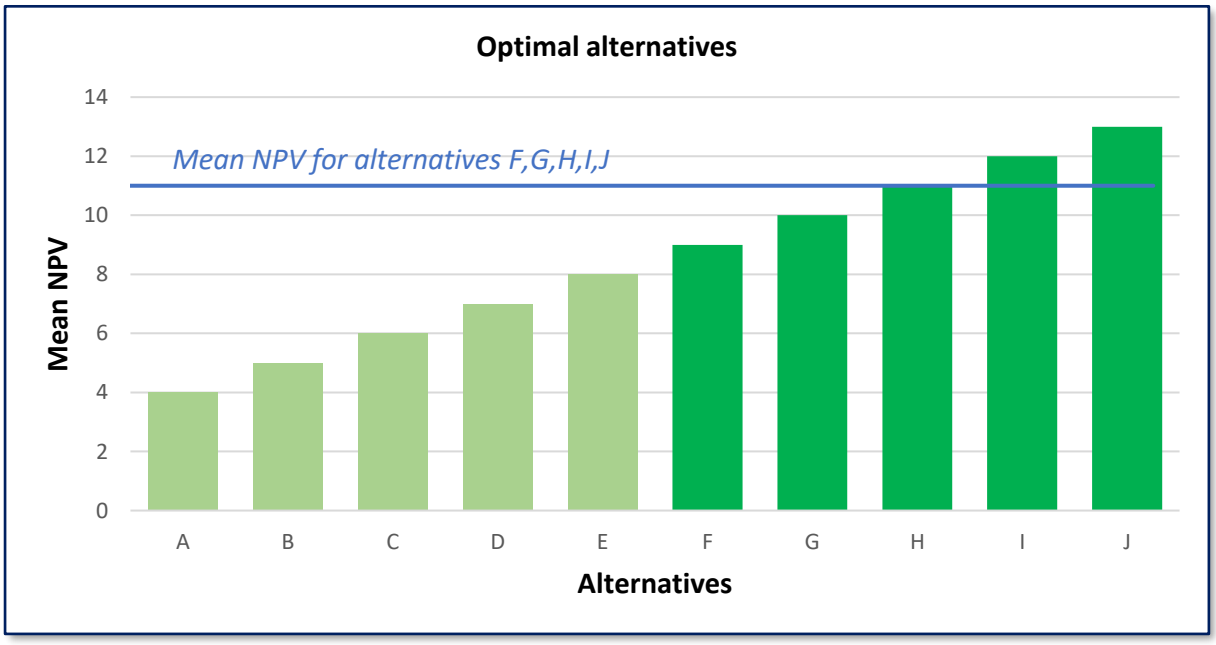

Figure 2.2. Five optimal alternatives

It must be mentioned that these expected net present values for each alternative are estimations and they are not known precisely. These estimations are uncertain and estimated values has errors. Value estimate errors can be indicated by the error bars in the figure below.

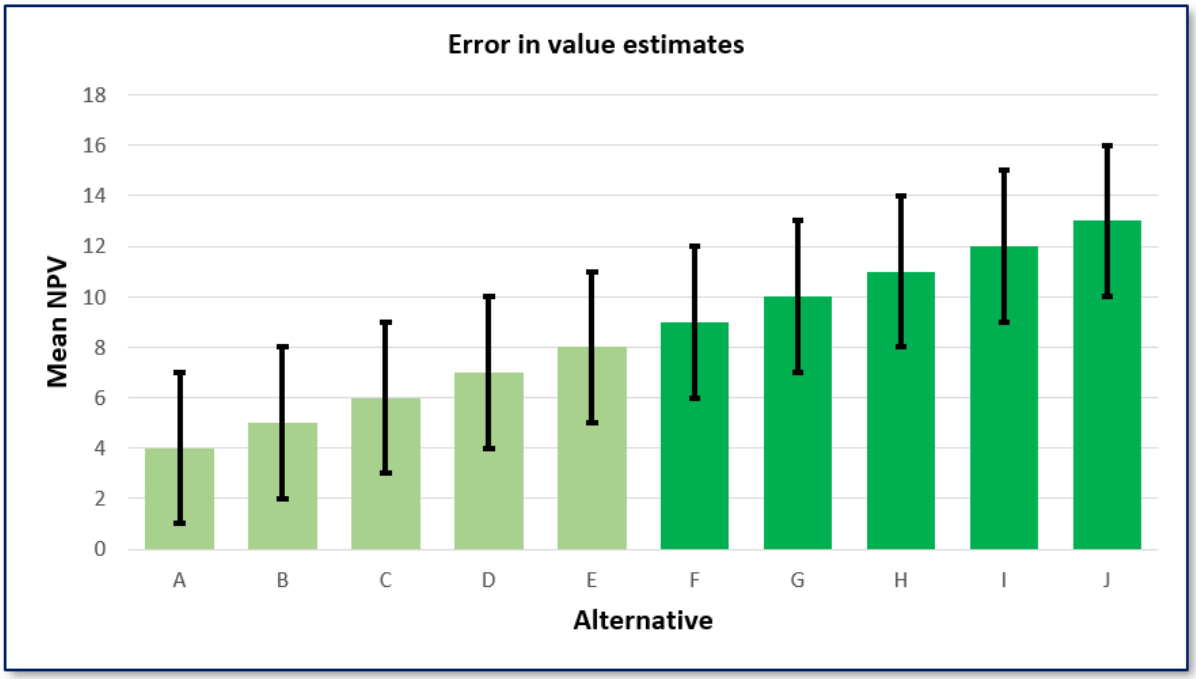

Figure 2.3. Error bars

The errors on average are equal to zero and they cancel each other out. Therefore, some of the alternatives are underestimated and some of them are overestimated. But on average, errors in value estimates should be equal to zero. The number of times that they are overestimated are equal to the number of the times that are underestimated, and the expected net present value of estimated alternatives is equal to the true net present value of alternatives.

In the figure 2.4 alternatives in red colour are overestimated, and yellow bars indicate the underestimated alternatives. The amount of error which is added to the alternatives on average is zero.

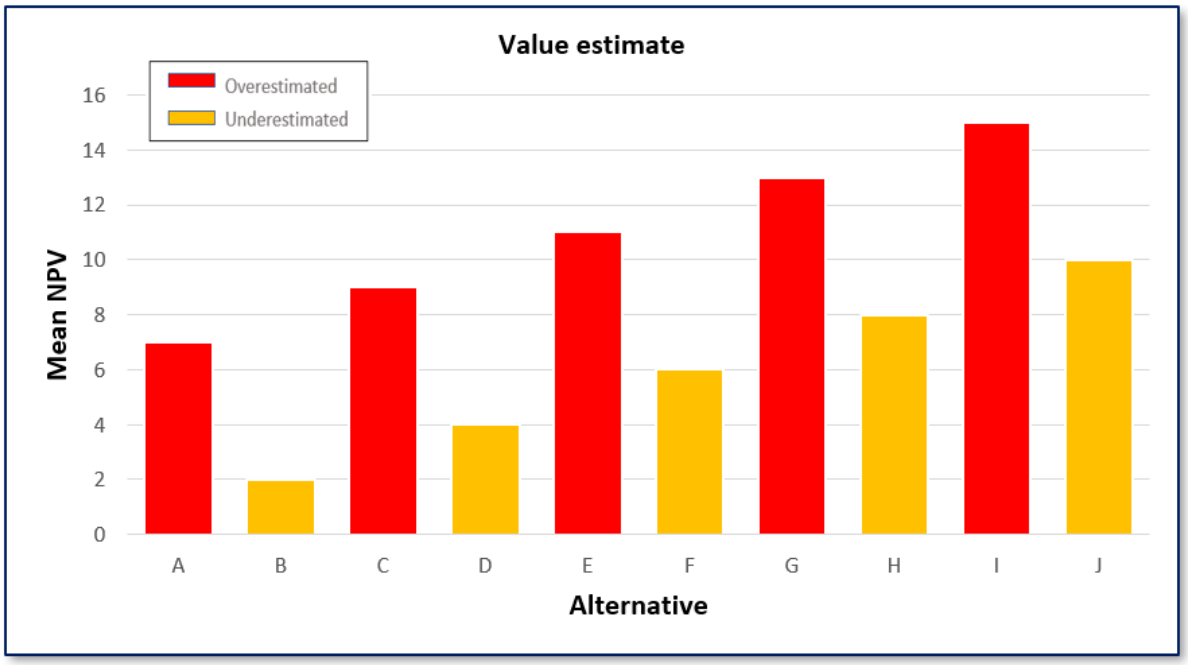

Figure 2.4. Estimated values

In figure 2.5 alternatives are ranked based on estimated values. By choosing alternatives with the highest estimated value among these estimated values, it is more likely to choose overestimated ones rather than underestimated alternatives. The probability of choosing the overestimated alternatives are higher for a decision-maker who selects the alternatives base on the estimated values and not on true values.

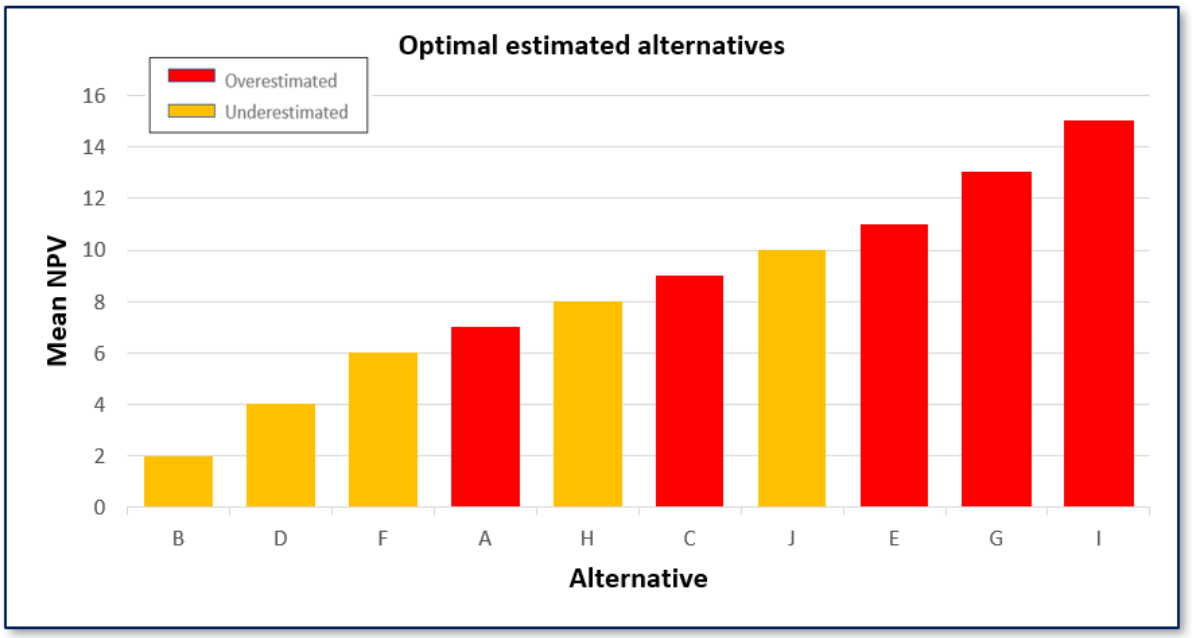

Figure 2.5. Optimal estimated alternatives

Figure 2.6 illustrates the average of the estimated expected net present value of the selected alternatives. It indicates that the NPV average of the estimated optimal alternatives is higher than the average of the true expected net present value of the previously selected alternatives. It is obvious that the estimation errors of the selected alternatives do not cancel out and they are positive.

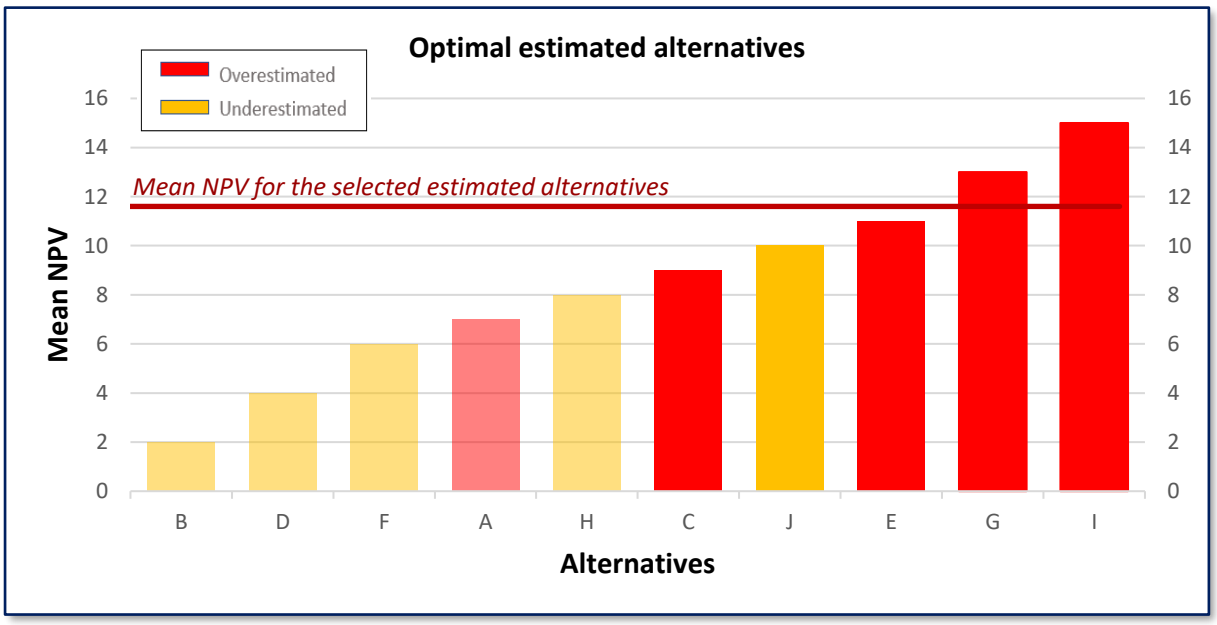

Figure 2.6.

In figure 2.7 the blue line indicates the average of the true expected net present value of the selected alternatives, whereas the red line is the average of the estimated expected net present value of the selected ones, and the green line is the average of the true expected net present value if selected alternatives are implemented.

By comparing the true expected net present value of each selected alternative which has shown by green bars, with the estimated ones (red and yellow bars), it illustrates that how much the decision-maker was optimistic about what he can achieve by implementing the selected alternatives. In fact, because of the estimation errors, the average of the estimated net present value of the selected alternatives is not achievable.

The blue line is the best that could be possible if we had full knowledge.

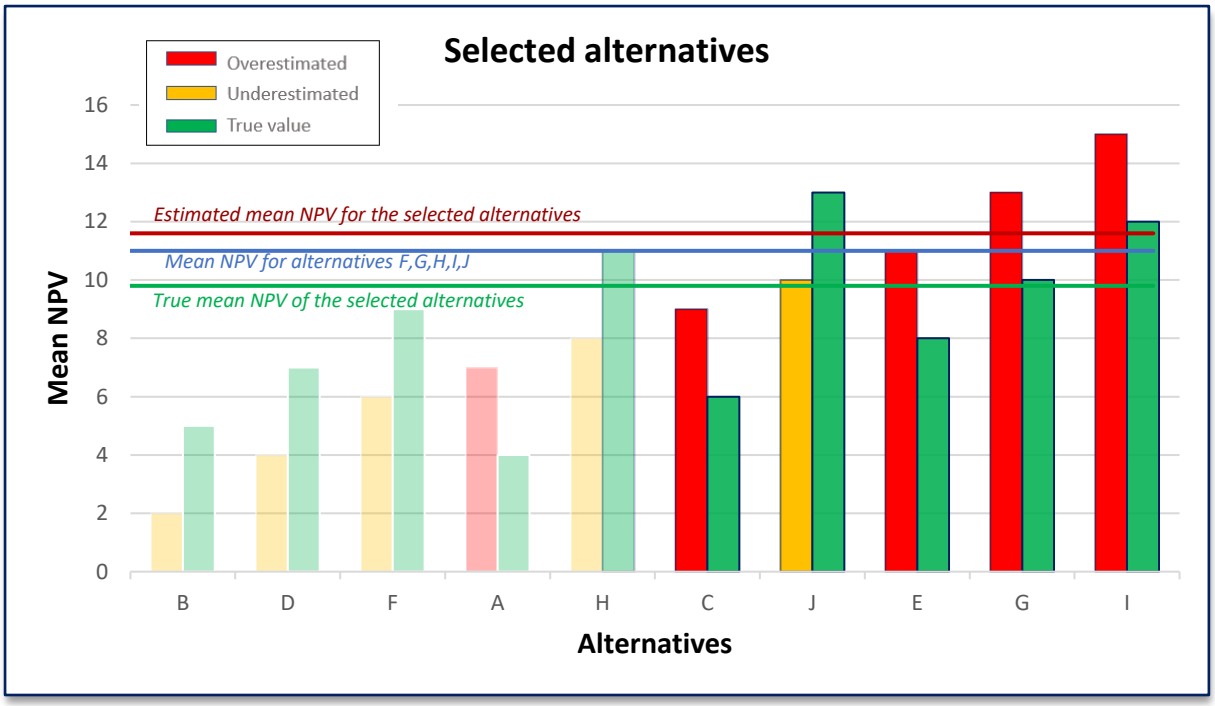

Figure 2.7.

In the example above I have demonstrated even if the errors in value estimates cancel each other out and on average is equal to zero (unbiased), the optimization result is biased and the result will be optimistic for what is achievable, and the decision-maker will be disappointed by executing the result of optimization process.

2.3 Optimizers' curse scenarios

Identical alternatives

An example by Smith & Winkler illustrates the optimizer's curse when the alternatives are identical. Consider three different alternatives with true values (μ_i) equal to 0. The value of each alternative is estimated. These value estimates (v_i) are independent and normally distributed. The mean of each distribution is equal to the true value of 0. The standard deviation of each distribution is equal to one and the estimations are conditionally unbiased. Now, choosing the highest value estimates among these three estimated alternatives and plot them to see the distribution of these selected values, the distribution of these optimal values has a mean equal to 0.85. It illustrates the mean of maximum values is positively moved away from the mean distribution. It shows that by implementing the optimal alternative, the resulted outcome is on average 0.85 percent less than the estimated value and decision-maker will experience post-

decision disappointment. the expected disappointment is 0.85 percent of the standard deviation. (Figure 2.8)

Following:

$E[v_i - \mu_i]$ = 0.85

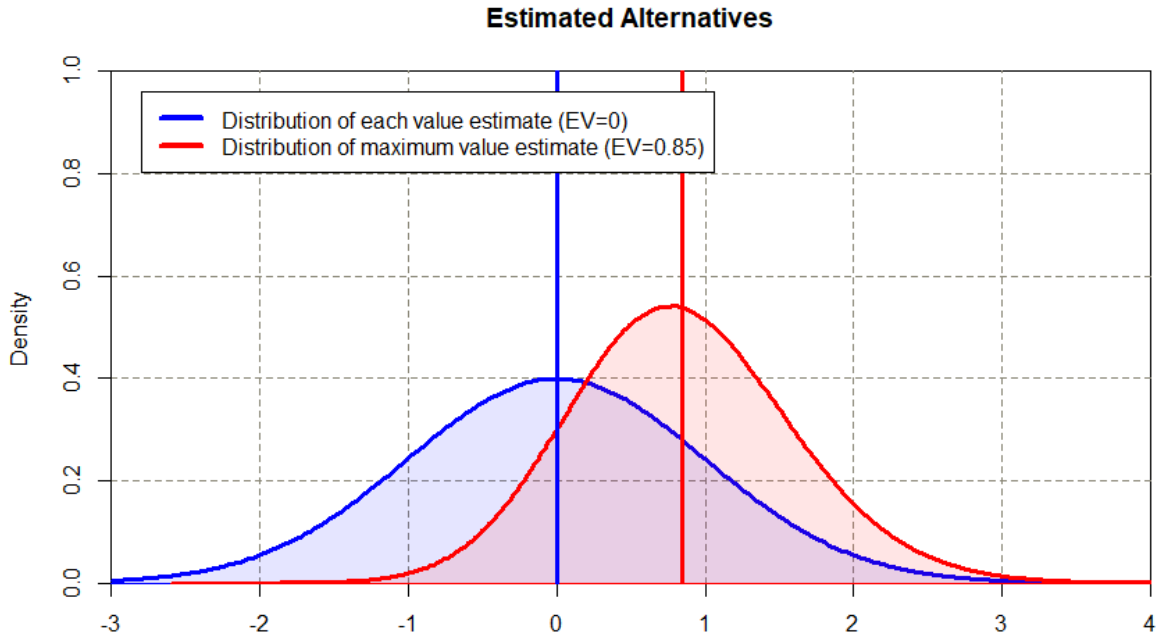

Figure 2.8. Three identical alternatives

Now consider a situation with 'n' identical alternatives (figure 2.9). Having the same distribution assumption, magnitude of the expected disappointment increases by increasing the number of the alternatives. Figure 2.9 shows that the distribution of maximum values is moved positively by increasing the number of the alternatives. It indicates that when the true values of the alternatives are equal, they cannot be distinguished even with the perfect estimates and the expected disappointment is higher (Smith and Winkler, 2006).

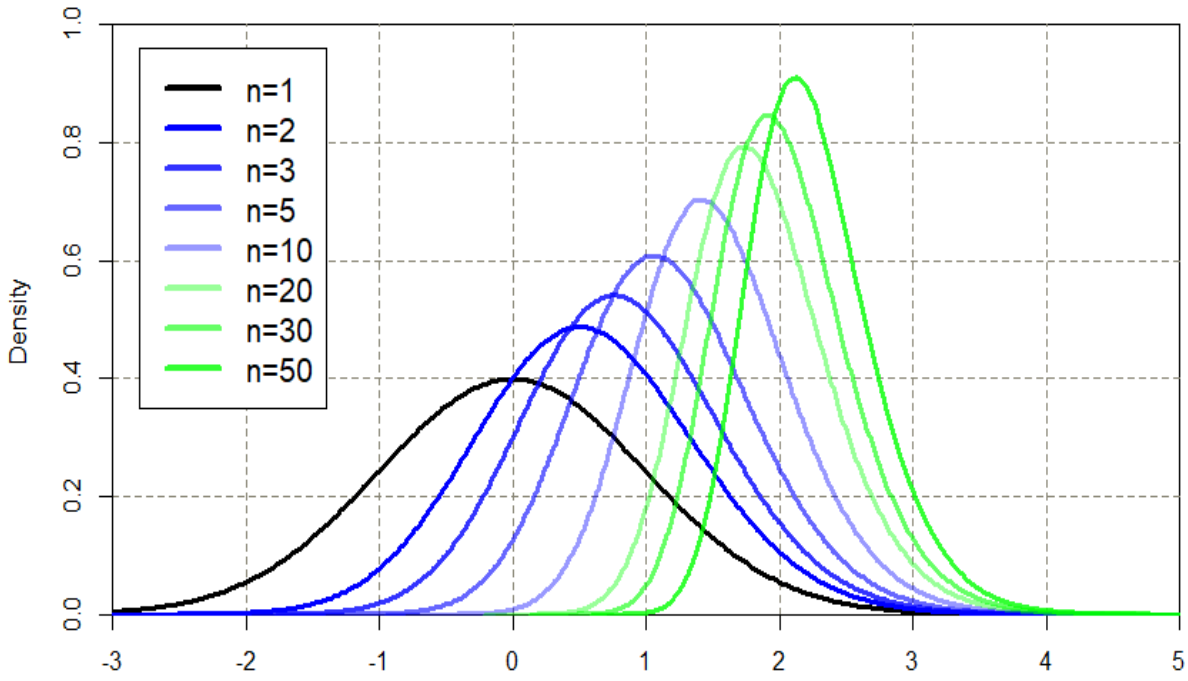

Figure 2.9. n identical alternatives

Different alternatives

If the true values of alternatives were separated by Δ: $\mu_i = -\Delta, 0, \& \Delta$ and the estimate of the mean values are unbiased, and standard deviation is equal to one. The magnitude of disappointment will be reduced by increasing the degree of the separation between alternatives. Consider three alternatives which the value estimates of these alternatives normally distributed with the mean equal to -0.5, 0 and 0.5 (figure 2.10). As the degree of the separation increases, the difference between the mean of the optimal alternative which has the highest true value among all three, and the mean of the maximum value estimates decreases. In other words, magnitude of the post-decision disappointment decreases by increasing the degree of the separation among alternatives.

$E[V_i - \mu_i]$ = 0.94 - (0+0.5) = 0.44

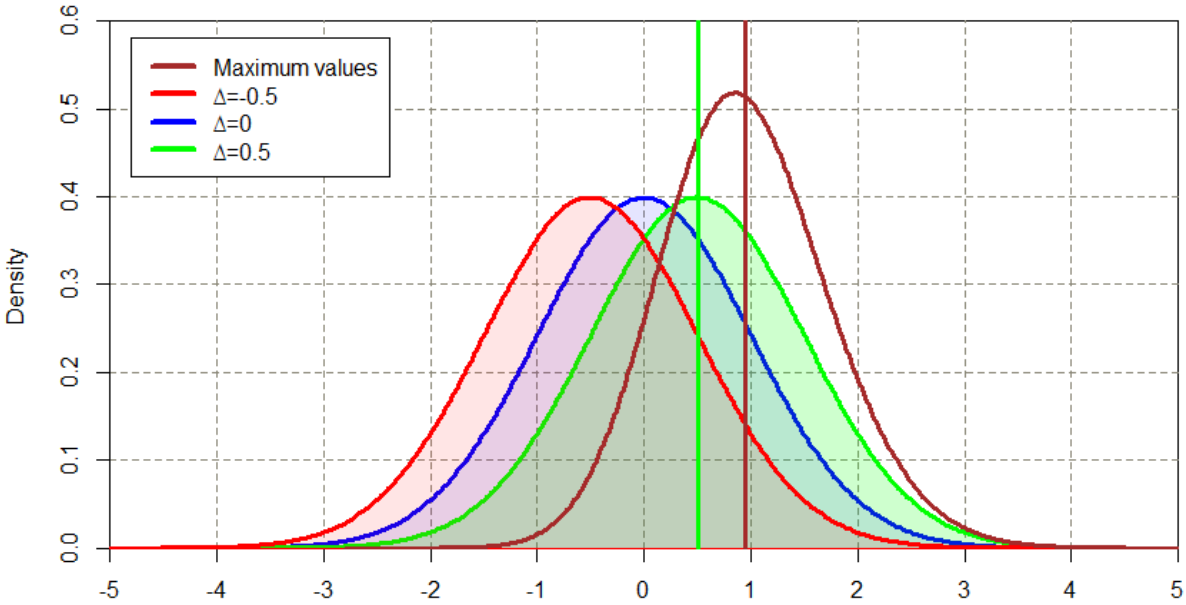

Figure 2.10. Three different alternatives

Table 2.1 indicates the magnitude of the expected disappointment as the fraction of the degree of the separation between alternatives.

Number of the alternatives = 3	
Δ	Expected disappointment
0	0.85
0.2	0.66
0.5	0.44
0.8	0.3
1	0.22

Table 2.1. Expected disappointment magnitude

Figure 2.11 indicates when the alternatives are identical but with a bigger standard deviation. In this case the expected disappointment is higher than when they have lower standard deviation (figure 2.8). In this case standard deviation is equal to two.

$E[V_i - \mu_i]$ = 1.69

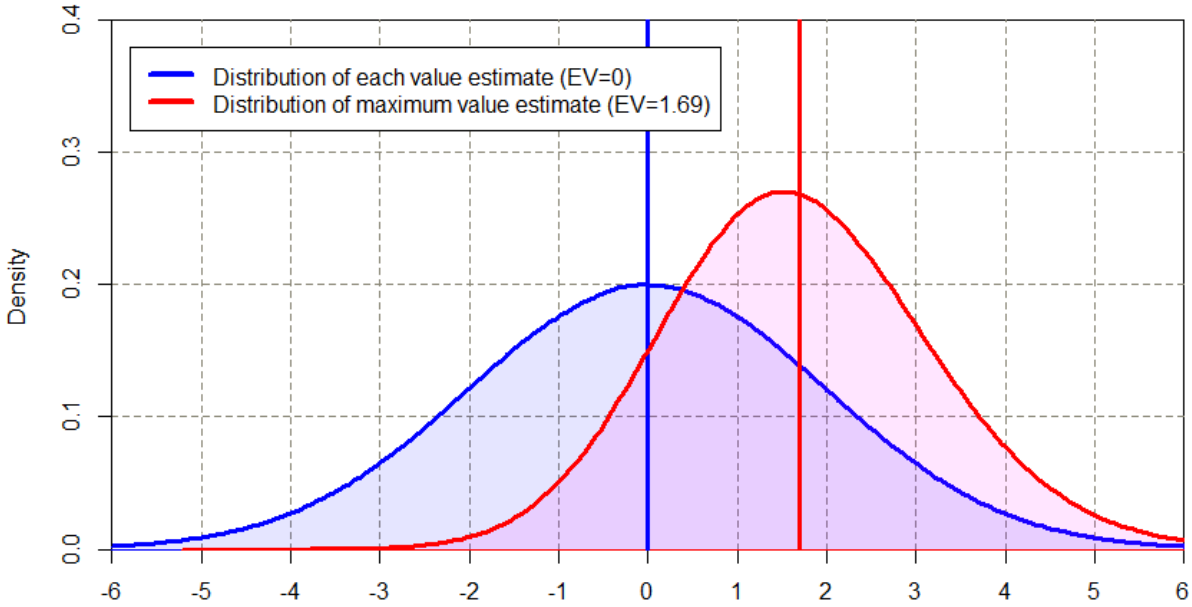

Figure 2.11. Identical alternatives with bigger standard deviation

Distribution of alternatives

Consider a distribution of NPVs for different alternatives like NPVs for different well locations with mean equal to 10 and a standard deviation of 1 (Figure 2.12). Green curve is an assumption of realized values of NPVs.

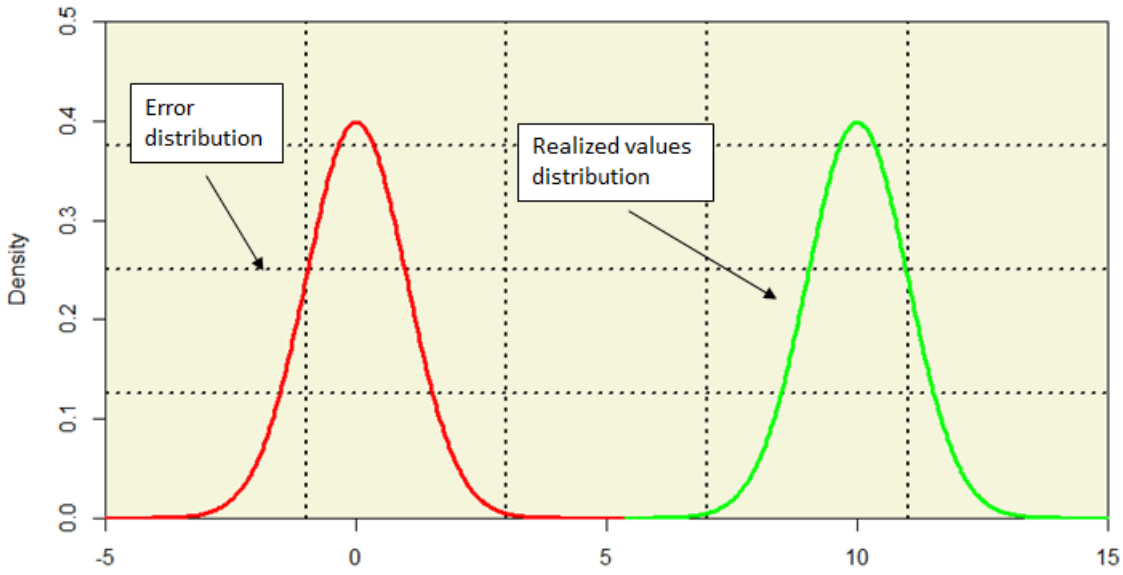

Figure 2.12. Distribution of NPV

Red curve in figure 2.13 indicates unbiased errors. By adding the distribution of errors to the distribution of realized values we can generate the distribution of the estimated values. Blue curve shows the distribution of the errors.

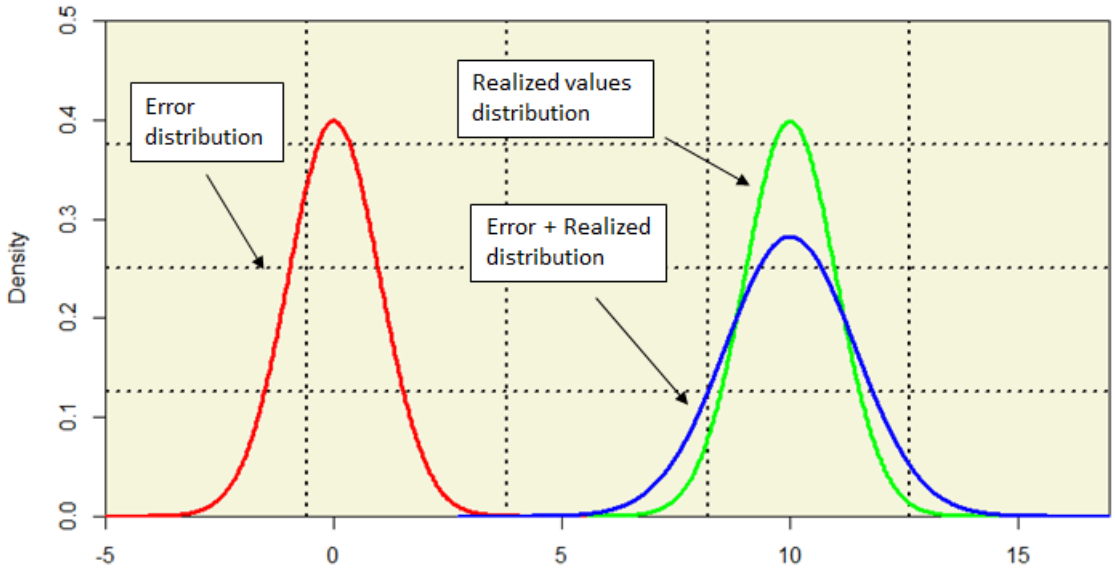

Figure 2.13. distribution of the estimated NPVs

Now by comparing the mean of the distribution of the optimal estimated values and the optimal realized values, the magnitude of post-decision disappointment will be identified (figure 2.14). The graph shows, even if the input values are unbiased, the error inside the estimated values leads to having an overestimated expected net present value for the optimal alternatives.

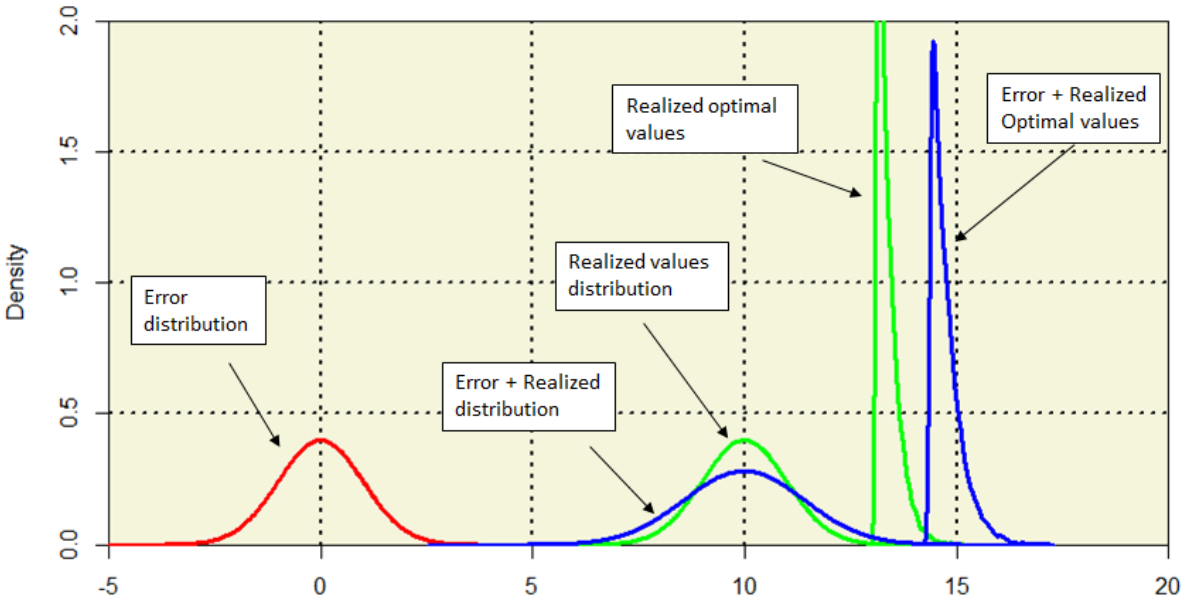

Figure 2.14. Distribution of optimal NPVs

Mean of the true optimal values = 12.66592

Mean of the estimated optimal values = 13.76905

Expected disappointment = 13.76905 - 12.66592 = 1.10313

As shown, the gap between the mean of the estimated and true optimal values is almost 1.1. It indicates that by executing the optimal alternatives on average, we will gain less than the expected estimated value.

2.4 Effect of correlation among estimated values

Consider two alternatives with the same expected net present value, which are evaluated accurately or unbiased but imprecisely, and each estimation has a 50 percent chance of being overestimated and 50 percent chance being underestimated, possible outcomes will be:

- 25 % chance of having both projects underestimated
- 25 % chance of having both projects overestimated
- 50 % chance of having one overestimated and one underestimated

If the decision-maker selects the alternatives based on the highest estimated value, the chance of selecting overestimated alternative is 75 percent (Summers, 2019). The value estimate of the selected alternative will thus overestimate the true value on average.

In the example above, value estimates are independent. However, the value estimates can be dependent and correlated. They may share a common element. For example, in selecting different strategies to develop an oil field, value estimates may share common probability of oil in place. In this example, if the two value estimates were perfectly positively correlated, then there was a 50 % chance of having both values overestimated and a 50 % chance of having both underestimated. In this case, to have an estimate for the selected alternative is equal to the true value on average, is expected. It indicates that a positive correlation among estimated values decreases the magnitude of the expected disappointment and negative correlation among them, in contrast, increases the degree of the expected disappointment. In practice, negative correlation is less likely to be considered (Smith and Winkler, 2006). In this example, true values are assumed to be fixed and value estimates are considered either independent or correlated, but in practice true values are uncertain and they might be positively correlated. For example,

when there is uncertainty about a probability of technical success resulted in true values of alternatives which are dependent on this probability, might be positively correlated. The positive correlation among true values decreases the degree of the separation among them which in return increases the magnitude of the expected disappointment and by contrast, a positive correlation among value estimates downsizes the magnitude of the optimizer's curse.

2.5 Errors and Biases

Prediction errors

In the previous section we showed that how a simple procedure of selecting and ranking among estimated values introduces a bias and leads to have a difference between predicted values that are used to make the decision and actual values that realized after the decision has been implemented. Some other reasons for this difference which Begge & Bratvold (2008) mentioned, are listed below:

Uncertainty:

Under uncertainty, predictions are estimates of expected value or expected utility when the uncertainty has been modelled. If the uncertainty has been ignored, predictions are considered deterministic. Consequently, estimated value highly unlikely to be realized on any single decision.

Biased in inputs:

Value estimates can be assessed directly, or they can be derived indirectly from assessed inputs upon which they depend. Both direct and indirect assessment are subjected to a variety of biases (Welsh et al., 2005). These input biases contribute to the gap between prediction and actual in any one decision.

Mistakes and errors:

One of the mistakes is measurement errors. It can be caused by limitation of the measurement devices or simply their misuse. Another error is simple mistakes in data entry or computation.

Use of models:

Most of the estimates are resulted from models. As the model is not reality, we cannot expect model outputs to match reality. The famous quote by George Box about the models is “All models are wrong; some models are useful”. (Begg and Bratvold, 2008)

Evaluation Biases

Common evaluation biases are:

Underestimating prior information

In task prediction, behavioural studies show that there is a tendency to underestimate prior information about the base rate of the event which is being predicted and taking decision-based on most recent evidence. It leads to predict extreme realizations. (Kahneman and Tversky, 1977, Kahneman and Tversky, 1973)

Judgment Biases

An analyse is objective when subjective input judgments are objective (Kahneman et al., 1982). Tendency to rely on some information or on certain direction, either in favour or against something, will contribute to the errors.

- **Overconfidence and optimism**

  Overconfidence is the best-known cognitive bias in oil industry, is affecting the judgments of decision-makers both on their general knowledge (Capen, 1976), and oil related questions (Welsh et al., 2005). Overstating the confidence in our knowledge leads to decrease the range of possible outcomes. For example, in evaluating geological aspects of a reservoir like average porosity and thickness, the range of confidence which usually interpreters use is about 80%. Data from the industry (Hawkins et al., 2002) and other fields (Morgan et al., 1990) has shown that the range of the overconfidence in such parameters when the actual value is included, is less than 50% of the time rather than 80% as the range of confidence should indicate.

  Overestimation which we refer as optimism, is one of the forms of overconfidence. Optimism is overestimation of one’s actual ability, performance, level of control, or chance of success. Consider a student who believes that he answered five questions correct in a 10-item quiz. He got the result and he

answered three correctly. He has overestimated his score. (Moore and Healy, 2008)

- **Anchoring**

  Cognitive bias like anchoring, leads us to rely too much on initial information in our decision-making process. When the numerical values are estimated based on poorly adjusted initial information or value (anchor), the resulting answer will be biased (Montibeller and von Winterfeldt, 2015).

- **Farming**

  The way that the information is presented, can lead people to make different judgments. Presenting some piece of information may have positive or negative effect on people (Hodgkinson et al., 1999).

Bias in decision policy

These Biases are deliberately assigned in the company's decision policy. Some sort of decision policy which can be the source for the bias are as following:

- Using high rate present value discount rate
- Alternative with lower values are accepted because of the lower risk.
- Instead of considering maximizing long-term shareholders value, put more weight on ethic toward corporate social responsibility. (Schuyler and Nieman, 2007)

Summery

In this chapter the optimizer's curse is explained. We described how we should expect to be disappointed by taking decision based on the simple process of ranking and selecting among the estimated alternatives. We showed that even if the value estimates for alternatives are unbiased, uncertainty in the estimations coupled with optimization process leads to experience post decision disappointment. Different factors which affect the magnitude of this curse are explained.

- Optimizer's curse is worse when alternatives have the same expected value
- Optimizer's curse is increasing with the number of the alternatives
- Having high degree of uncertainty increases the magnitude of post-decision disappointment
- Correlation between estimated values decreases the magnitude of this curse

- Separation among true values decreases the expected disappointment

We also introduce common evaluation errors and biases in addition to the optimizer's curse which affects our prediction of future outcome.

Chapter 3. Optimizer's curse reduction

3.1 Introduction

The solution to overcome the optimizer's curse is conceptually quite simple and straight forward. "Model the uncertainty in the value estimates explicitly and use Bayesian method to interpret these value estimates"(Smith and Winkler, 2006). Bayesian modelling helps to reduce the amount of expected disappointment that the decision maker expects to experience. It also serves to increase the number of the alternatives which belong to the optimal portfolio (Vilkkumaa et al., 2014). We will develop a Bayesian model later in this chapter.

3.2 Simple actions that helps to reduce the effect of the optimizer's curse

Smith and Winkler described when the true values are more analogous and the uncertainty in value estimations is high, the optimizer's curse effect will be higher. In other words, the likelihood of making a wrong decision is high when there is significant uncertainty in the ability to evaluate values and small differences between alternatives. Having a high degree of uncertainty in the value estimates, not only makes it difficult to distinguish the alternatives with the highest values but also makes it more likely to choose alternatives that are systematically overestimated.

Collecting more information

In chapter 2 we described how random evaluation errors which is combined with optimization-based selection process leads to post decision disappointment. Errors, which come with alternative evaluation, can be reduced by collecting more information and mitigating the uncertainty in value estimates. Generally, in making a choice between two alternatives, the value of information is highest when the decision-maker is indifferent between two alternatives and this value is lower when there is a preference in favour of one alternative over another (Delquié, 2008). In this case, one suggestion is implementing sensitivity analysis and value of information assessment to find out whether to collect additional information. This additional information must provide enough insight to distinguish between alternatives to justify the cost of the new information collection. This new information can reduce the uncertainty in the value estimates. Consequently, decreases the magnitude of the optimizer's curse.

Reducing estimation errors

In single project evaluation, defining the reduction magnitude of the optimizer's curse is not easy. While the uncertainty is assessed in the estimation of the alternatives value, in the optimizer's curse estimation errors are more important. By collecting new information to reduce the degree of uncertainty in the project evaluation, it should be expected to achieve some reduction in the estimation errors as well. (Schuyler and Nieman, 2007)

Reducing the errors in value estimates effects the optimizer's curse in two ways:

1- Expected disappointment which described as the difference between the estimated value and true value will be reduced
2- The probability of selecting the best alternative will increase

As the example by smith and Winkler in section 2 demonstrates, three alternatives each having true values separated by Δ, 0 and -Δ respectively. Value estimates are considered unbiased and normally distributed with a mean equal to the true value and a standard deviation of 1. To indicate the impact of reducing the errors in value estimates and changing in the degree of the separation between alternatives on expected disappointment, table 3.1 is created.

The table indicates that modest degradation in value estimates uncertainty (standard deviation) results in a reduction in the magnitude of the optimizer's curse. As the separation between true values becomes wider the magnitude of expected disappointment shows more reduction.

	Standard Deviation of the value estimate				
Δ	***1.00***	***0.75***	***0.50***	***0.25***	*Expected Disappointment*
0.0	*0.85*	*0.63*	*0.43*	*0.21*	
0.2	*0.66*	*0.45*	*0.25*	*0.07*	
0.4	*0.51*	*0.32*	*0.15*	*0.02*	
0.6	*0.39*	*0.21*	*0.08*	*0.00*	
0.8	*0.30*	*0.15*	*0.04*	*0.00*	
1.0	*0.22*	*0.10*	*0.02*	*0.00*	
1.2	*0.17*	*0.07*	*0.01*	*0.00*	
1.4	*0.12*	*0.04*	*0.01*	*0.00*	
1.6	*0.10*	*0.03*	*0.00*	*0.00*	
1.8	*0.07*	*0.01*	*0.00*	*0.00*	
2.0	*0.05*	*0.01*	*0.00*	*0.00*	
2.2	*0.03*	*0.00*	*0.00*	*0.00*	
2.4	*0.02*	*0.00*	*0.00*	*0.00*	
2.6	*0.01*	*0.00*	*0.00*	*0.00*	
2.8	*0.01*	*0.00*	*0.00*	*0.00*	
3.0	*0.00*	*0.00*	*0.00*	*0.00*	

*Table 3.1. expected disappointment as a fraction of **Δ** and std. deviation*

3.3 Effect of uncertainty reduction on optimal alternative

While uncertainty reduction reduces the magnitude of the optimizer's curse, it doesn't have a significant effect on the probability of selecting the alternative with the highest true value (Schuyler and Nieman, 2007). It might help to select more optimal alternatives.

Schuyler and Neiman calculated the probability of selecting the best alternative for different levels of uncertainty (table 3.2) on the example proposed by Smith & Winkler. The table 3.2 indicates another dimension of the previous example. It shows the probability of selecting the best alternative as a function of uncertainty changing in estimations and the magnitude of the separation between true values. Specifically, when the three alternatives true values are separated by $\Delta = \pm 1$ and the standard deviation for the estimated values is equal to 0.5 ($\sigma = 0.5$), there is a 92 percent chance to select the correct alternative. It is obvious that when the separation between true values are $\Delta = \pm 3$, the probability of selecting the optimal alternative is 100 %. It indicates that the probability of picking the right choice or the best alternative is not

changed considerably by reducing the uncertainty in value estimates as it reduces the expected disappointment. (Schuyler and Nieman, 2007)

	Standard Deviation of Value Estimate				
Δ	***1.00***	***0.75***	***0.50***	***0.25***	*Probability of selecting the best alternative*
0.0	*0.33*	*0.33*	*0.33*	*0.33*	
0.2	*0.42*	*0.45*	*0.51*	*0.66*	
0.4	*0.51*	*0.56*	*0.66*	*0.87*	
0.6	*0.59*	*0.67*	*0.78*	*0.96*	
0.8	*0.66*	*0.75*	*0.87*	*0.99*	
1.0	*0.73*	*0.82*	*0.92*	*1.00*	
1.2	*0.78*	*0.87*	*0.95*	*1.00*	
1.4	*0.83*	*0.91*	*0.98*	*1.00*	
1.6	*0.87*	*0.93*	*0.99*	*1.00*	
1.8	*0.90*	*0.96*	*0.99*	*1.00*	
2.0	*0.92*	*0.97*	*1.00*	*1.00*	
2.2	*0.94*	*0.98*	*1.00*	*1.00*	
2.4	*0.96*	*0.99*	*1.00*	*1.00*	
2.6	*0.97*	*0.99*	*1.00*	*1.00*	
2.8	*0.98*	*1.00*	*1,00*	*1.00*	
3.0	*0.98*	*1.00*	*1.00*	*1.00*	

Table 3.2. probability of success (Schuyler and Nieman, 2007)

3.4 Develop a Bayesian model

Post-decision disappointment can be modified by revising value estimates with Bayesian methods (Gelman et al., 2013). By implementing Bayesian modelling of estimation uncertainties, and instead of ranking the alternatives based on estimated values, selecting among these resulting revised estimates helps to

1. Select alternatives which can have higher expected future value
2. Increase the number of optimal alternatives
3. Decrease the magnitude of post-decision disappointment
4. Investigate how resources should be spent in order to reduce the uncertainty in estimations. Re-evaluating a small number of alternatives instead of spending a

pile of money on re-evaluating all possible choices can increase the expected value of the expected portfolio value. (Vilkkumaa et al., 2014)

Estimating the parameters needed for Bayesian modelling is difficult in practice. Harrison and March suggest that obtaining more accurate value estimates helps to reduce the post-decision disappointment. (Harrison and March, 1984)

In addition, in portfolio selection systematic prioritization among projects or alternatives estimated values (as opposed to the random selection), tends to add more value than trying to reduce the uncertainty and achieve more accurate estimated projects. (Keisler, 2004)

Selecting the optimal alternatives

Selecting the optimal alternatives among a set of projects is defined by the decision variable $z = [z_1, \ldots, z_m]$, which is a binary decision with $z_i = 1$ only if the project i is selected (Vilkkumaa et al., 2014). The set of alternatives which are qualified and within the constraints are defined by z. In fact, if the true values v were recognized, optimal alternatives can be determined by an equation for the optimization problem such that:

$$z(v) = \arg\max_{z \in Z} zv$$

But the true values are unknown, there are just estimated values v^E. For selecting the optimal alternatives from these value estimates equation below must be solved:

$$z(v^E) = \arg\max_{z \in Z} zv^E$$

Bayesian modelling of uncertainty

The discussion below is borrowed from VilkKumaa's work (2014). Consider v is the indicator of the true values and v^E shows the estimated one. If the decision-maker selects the alternatives based on the value estimates without considering uncertainties which are coupled with prior distribution $f(v)$ and likelihood distribution $f(v^E|v)$, overestimated alternatives are more probable to be selected and it leads to gain post-decision disappointment. Revising value estimates allows to mitigate this value overestimation by correcting the initial estimation for Bayes. By using Bayes rules and having a prior distribution $f(v)$ and the likelihood distribution $f(v^E|v)$, the posterior distribution $f(v|v^E)$ for the alternatives given the estimates, can be achieved.

$f(v|v^E) \propto f(v)\, f(v|v^E)$

Then the mean of the posterior distribution $f(v_i|v^E)$, will give the Bayesian estimate for the selected alternative. Given the value estimates, posterior distribution can be used to calculate the expected value for the alternative or the probability of having the alternative in the set of the optimal values.

If the value of alternatives based on the Bayes estimates are v^B = $[v_1^B, \ldots, v_m^B]$ then optimal alternatives can be obtained by

$z(v^B) = \arg\max_{z \in Z} z v^B$

Consider a set of alternatives in a portfolio i = 1, . . . ,m which if executed will gained values v = $[v_1, \ldots, v_m]$. These values are modelled as realizations of random variables V = $[V_1, \ldots, V_m]$ $\sim f(v)$. Which $f(v)$ or joint distribution function assumed to be known. Estimated values v^E = $[v_1^E, \ldots, v_m^E]$ have errors which is normally distributed with a mean of 0 and standard deviation equal to τ . If the realized values of the alternatives are the random variables which are independent and identically distributed, then:

$V_i = \mu_i + E_i$

When E_i is a normal distribution with a mean of zero and variance of σ^2 and μ_i is the mean of the realized values. Value estimates can be obtained by:

$(V_i^E | V_i = v_i) = v_i + \Delta_i$

When, Δ_i is a normal distribution of estimation errors with a mean equal to 0 and variance equal to τ^2 . By considering the same probability distribution family for both true values and estimated values, now the Bayes estimates of alternatives can be obtained by:

$v_i^B = \alpha_i\, v_i^E + (1 - \alpha_i)\, \mu_i$

α_i can be acquired by having the standard deviation of realized values and the estimation errors:

$\alpha_i = \left(1 + \frac{\tau_i^2}{\sigma_i^2}\right)^{-1}$

By having a weighted average of the prior mean and the observed estimates, the Bayes estimates for a normal distribution can be obtained. The variance ratio $\frac{\tau_i^2}{\sigma_i^2}$ provides the weighting. Now if

the standard deviation of the errors is larger than the prior standard deviation then, the weight has a positive effect on the prior expectation μ_i and reduces the v_i^E. In contrast when the estimations error standard deviation is smaller than the prior standard deviation, then α_i value is closer to 1. It indicates that the estimated value v_i^E will provide a more realistic number which could be closer to the realized value.

Adjusting the estimated values by Bayesian method not only reduces the degree of overestimation but also the alternatives whose values are underestimated will be adjusted.

In the equation above, if the expected true value for the alternatives in a portfolio was more than the estimated value, the Bayesian adjustment corrects the estimated value for under estimation. (Vilkkumaa et al., 2014)

Consider a set of alternatives in a portfolio which the realized values V_i of these alternatives in the portfolio are independent and identically distributed following:

$V_i = \mu_i + E_i$

The mean and the standard deviation of this distribution are given:

$\mu_i = 15$

$E_i \sim (0,3) \longrightarrow \sigma \sim N(0,3)$

Two alternatives A & B from the set of alternatives are considered. Value estimates for these alternative modelled as:

$(V_i^E | V_i = v_i) = v_i + \Delta_i$

$\Delta_A \sim N(0,3)$

$\Delta_B \sim N(0,2.8)$

Figure 3.1 illustrates the estimated value and value for each alternative. Red and Green circles are indicator of estimated value for A & B, respectively.

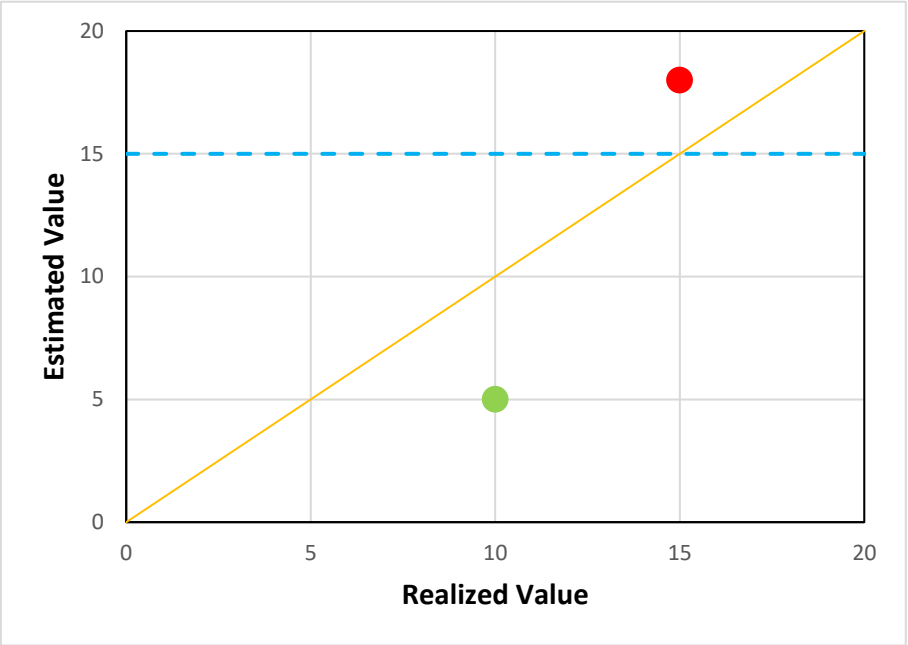

Figure 3.1. Values vs. estimates

By adjusting these estimated values for Bayes, resulted Bayes estimates are:

$$v_i^B = \alpha_i \, v_i^E + (1 - \alpha_i) \, \mu_i$$

$$\alpha_i = \left(1 + \frac{\tau_i^2}{\sigma_i^2} \right)^{-1}$$

$$v_A^B = 16.50$$

$$v_B^B = 9.65$$

In figure 3.2 the pink and the light green circles are the Bayes estimates for A and B respectively. As the graph shows these Bayes corrected estimates are closer to the realized values for each alternative.

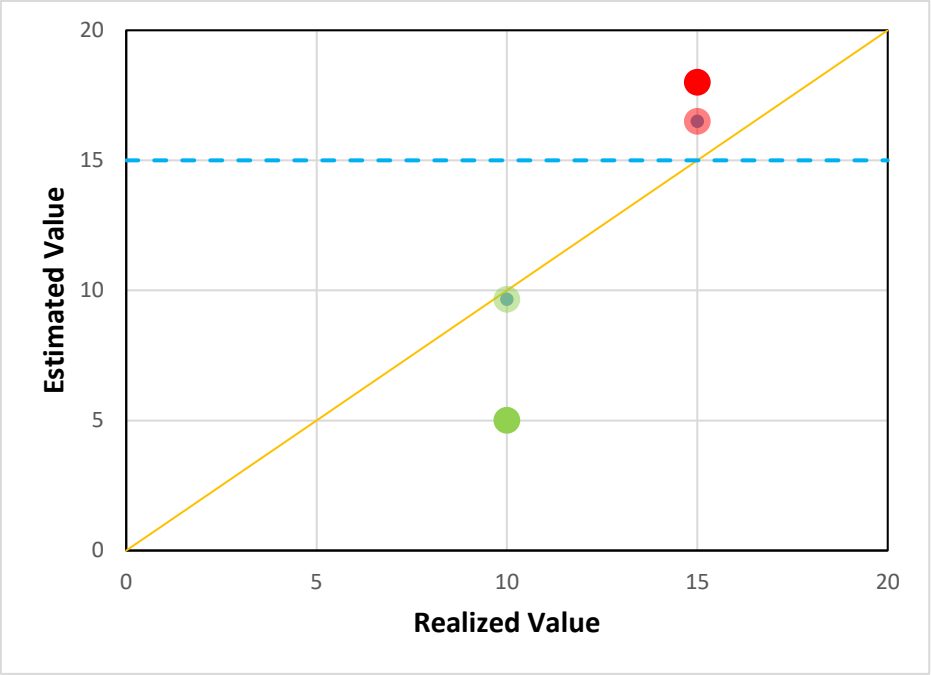

Figure 3.2. Values vs. Bayes estimates

For alternative A this estimate is corrected for overestimation and for alternative B it is corrected for underestimation.

Summery

In this chapter we explained how to reduce the optimizer's curse. simple actions like collecting more information and reducing estimation uncertainty are introduced. We also explained how reducing the errors in value estimates can help to decrease the magnitude of the optimizer's curse by:

1- Reducing the difference between the estimated value and true value
2- Increasing the probability of selecting the best alternatives

It is also demonstrated although uncertainty reduction decreases the amount of post-decision disappointment and helps to increase the number of the optimal values in a portfolio, it doesn't have a huge effect on the probability of selecting the alternative with the highest value. To reduce the magnitude of the optimizer's curse, a Bayesian model based on the Vilkkumaa's paper is developed. It helps to correct the original estimates of the alternatives by eliminating the gap between the estimated values and realized values and increases the expected future value of the selected portfolio.

Chapter 4. Uncertainty capturing and model development

4.1 Introduction

Usually, the geology of the underground cannot be determined directly, using indirect techniques to estimate the recoverable oil and gas volume of reserves is very uncertain. Although new technologies have made it more precise, but it still has a significant amount of uncertainty inside the prescription of underground (Morehouse, 1997).

In reservoir simulation, data from subsurface can be divided in two groups, static and dynamic. Static subsurface data can be derived from assimilating well logs data and seismic information. It is a combination of measured data which is collected directly from wells, with indirect measurements, like seismic data. A three-dimensional static description of the subsurface was then built by integrating well data and seismic information by using geostatistical methods (Ramirez et al., 2017).

In the history matching approach, dynamic subsurface descriptions integrate existing dynamic measurements, like well pressure and saturation. Quantification of uncertainty in subsurface descriptions is possible by applying Bayesian methods like Markov chain Monte Carlo, or by using deterministic methods. These approaches consider the uncertainty by ensembles of models which is referred to as samples from the posterior distribution of the uncertain parameters as they incorporate the dynamic information.

4.2 Geostatistical methods

With Geostatistical methods, uncertainty in the reservoir properties can be quantified as they move away from the location of the measured data. Static subsurface descriptions help to quantify uncertainties in the field. As these static subsurface descriptions do not include dynamic information, they can be referred to as samples from the prior distribution of the uncertain parameters (Ramirez et al., 2017). To make a continuous surface a geostatistical analyser uses sample points taken from different locations. These sample points can be the value or measurements of some phenomenon (Johnston et al., 2001). It can be some data that are derived directly from wells. These data from wells are measurements of some phenomena like porosity or permeability etc. These measurements will then be used to predict values for each location in

the reservoir. By using geostatistical methods these data can be interpolated to the whole reservoir. Interpolation can be done by deterministic and geostatistical techniques. Both two methods rely on similarity of nearby sample points (data points) to assign values to the whole volume. Deterministic method interpolation is done by mathematical functions while the geostatistical method uses both statistical and mathematical functions for interpolation and assesses the uncertainty of the predictions (Johnston et al., 2001).

4.3 Sequential Gaussian simulation

Quantify the uncertainties, both in production forecasting and optimization process is an important task. It may take hundreds of thousands of simulations to find out what is the uncertain parameter space.

The Gaussian distribution is chosen because it can initiate conditional distributions that all have the normal or Gaussian shape with the mean and variance which are provided by kriging.

A Gaussian field like a permeability ensemble is generated by a Sequential Gaussian Simulation method. Mean value, variance value and kriging method will be used in the procedure of the Sequential Gaussian Simulation (SGS). To calculate a value at an unstimulated grid cell, Gaussian procedure uses input data and simulated data. It generates realizations with equal probability which can be processed in a subsequent step to evaluate the uncertainty. In the Kriging interpolation method permeability mean value and standard deviation measures are provided for each cell in the grid. Then the Gaussian simulation draws samples from Monte Carlo simulation of the permeability distribution to make multiple realizations of permeability. It generates many equally probable realizations of a property to capture uncertainty in the permeability. The result provides a better representation of the natural variability of the property and delivers the mean value to quantify the uncertainty. Usually, Sequential Gaussian Simulation procedure is used to generate the values for continuous variables for geostatistical simulations in reservoir modelling (Gao, 2019).

Basic following steps in Sequential Gaussian Simulation is:

1. Original well data will be transformed into normal-score data with mean value equal to zero and unit variance. Any normal distribution can be identified by mean value (μ) and standard deviation (σ). Variable X from original data can be transformed to a standard normal variable like Y by the formula below:

$Y=(X-\mu)/\sigma$

If the mean and standard deviation of a variable X is equal to μ and σ when it transformed to variable Y, the mean would be equal to 0 and the standard deviation would be 1 (trek, 2019).

2. Establish a grid network and coordinate system
3. Assign transformed data to the simulation grid and nearest grid node
4. Create a random path through the grid nodes
5. Find the closest data and the grid nodes which is simulated before
6. Using kriging in the random path to build up a conditional distribution to estimate the mean and standard deviation at that node based on surrounding data and variogram. A local conditional probability distribution will be generated
7. Randomly choosing a value from the local conditional probability distribution and consider it as the node value including the newly simulated value as a part of the conditioning data
8. Check the results, if it honours the data, variogram and geological concept and repeating previous steps until all grid nodes have a value
9. Transform the realizations back to the original space

4.4 Generating permeability realizations

In the Geological model the uncertain Petrophysical parameter is permeability. To capture the uncertainty in the model, 100 realizations are generated. Sequential Gaussian Simulation (Pyrcz and Deutsch, 2014) is used to generate the different realizations of the permeability in the model. Within the model which is adopted from Peyman Kor (2019) and later on developed specifically for this work by me, the production strategy is considered as a 5-spot pattern.

Parameters	***value***
Nugget effect	Sill/2, md^2
Type	Spherical
Range	20(grid cell)
Anisotropy Ratio	1
Azimuth	0-degree (North)

Table 4.1. Semi variogram parameters (Kor, 2019)

As the anisotropy ratio is shown in the table above, horizontal permeability considered to be equal to the vertical permeability Kh = Kv. In this work, 100 realizations of the permeability are generated for each training observation. A grid of values that are randomly selected from a standard normal distribution is built by the Sequential Gaussian Simulation procedure. By defining the semivariogram in the kriging layer, the semivariogram/covariance model will be applied to ensure that the raster values conform to the spatial coordinates found in the input data set. Developed raster provides an unconditional realization. Gaussian distribution generates more realizations by utilizing different rasters of random values.

SequentialGaussian Simulation steps are as follow:

1. Using the Q-Q plot, log 10 of permeability is transferred to Gaussian values
2. The random path is generated by calculating the distance between the data and unknown location
3. Spherical variogram model is used for the input model of spatial continuity
4. Using isotropic variogram model to calculating the variogram matrix
5. Calculating the auto-covariance matrix by subtracting the variogram from the variance
6. Calculating simple kriging weight by multiplying the left-hand side inverted covariance matrix to the right-hand side matrix
7. Calculating the kriging estimates and variance

8. Applying Monte Carlo simulation to provide the simulated realizations in the random path(Kor, 2019)

In Figures below four different randomly chose permeability distributions which are generated by the Sequential Gaussian Simulation are indicated. The standard deviation and mean of each one is written underneath each ensemble. The red line is the indicator of the mean.

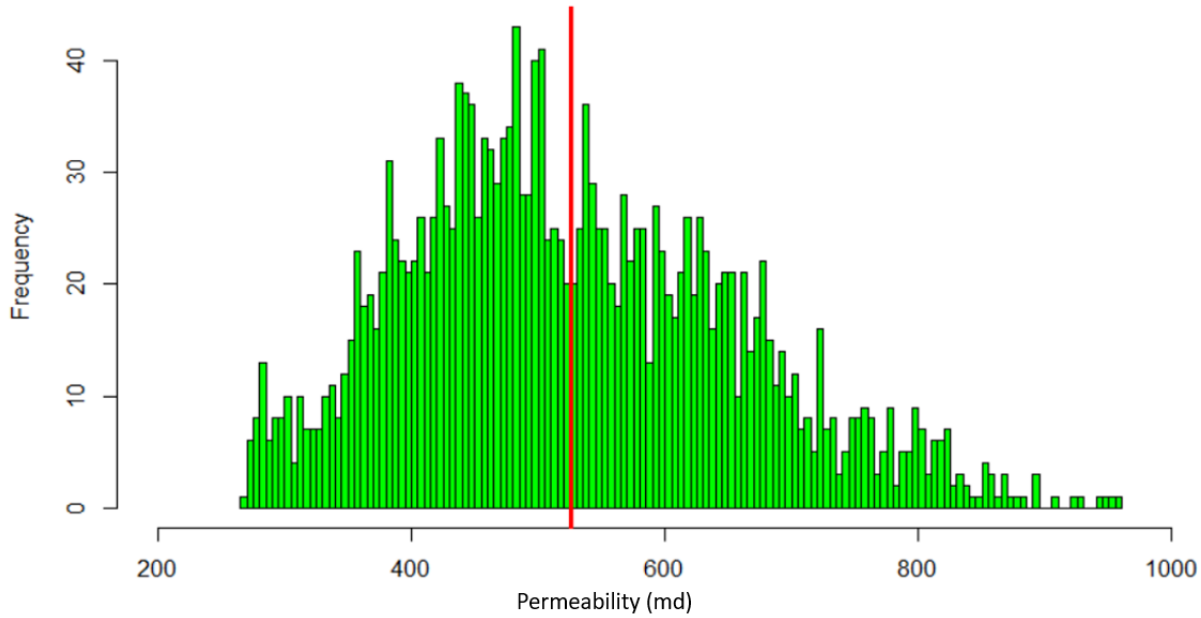

Figure 4.1. Permeability distribution realization#1

Mean = 525.7165

Standard deviation = 130.3313

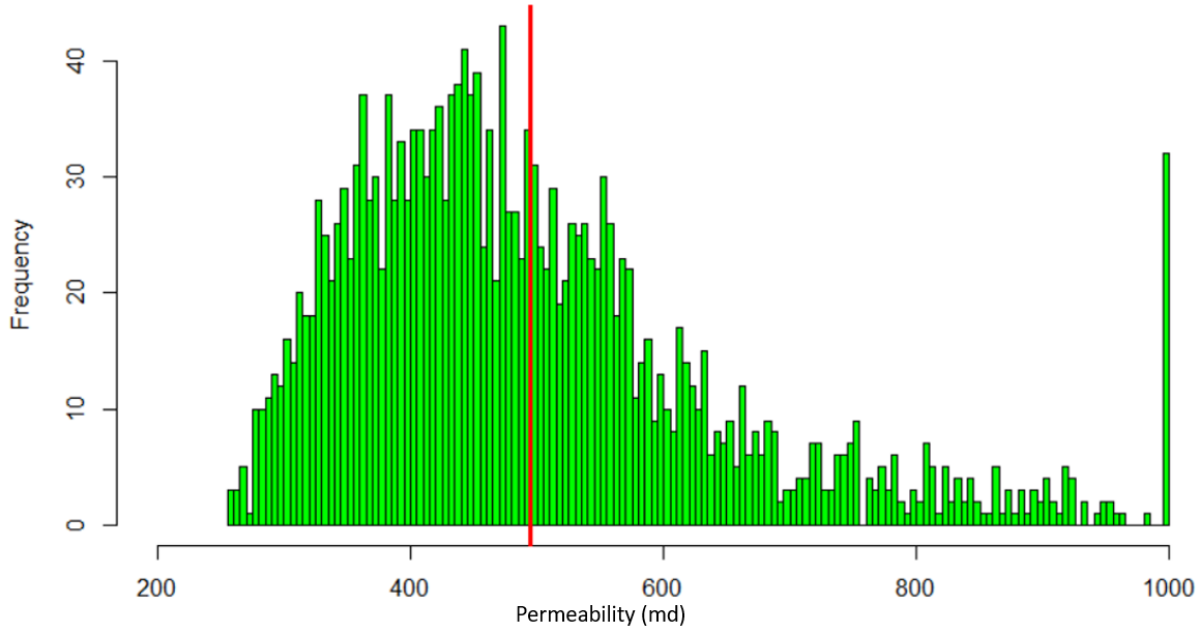

Figure 4.2. Permeability distribution realization#37

Mean = 494.2892

Standard deviation = 150.3611

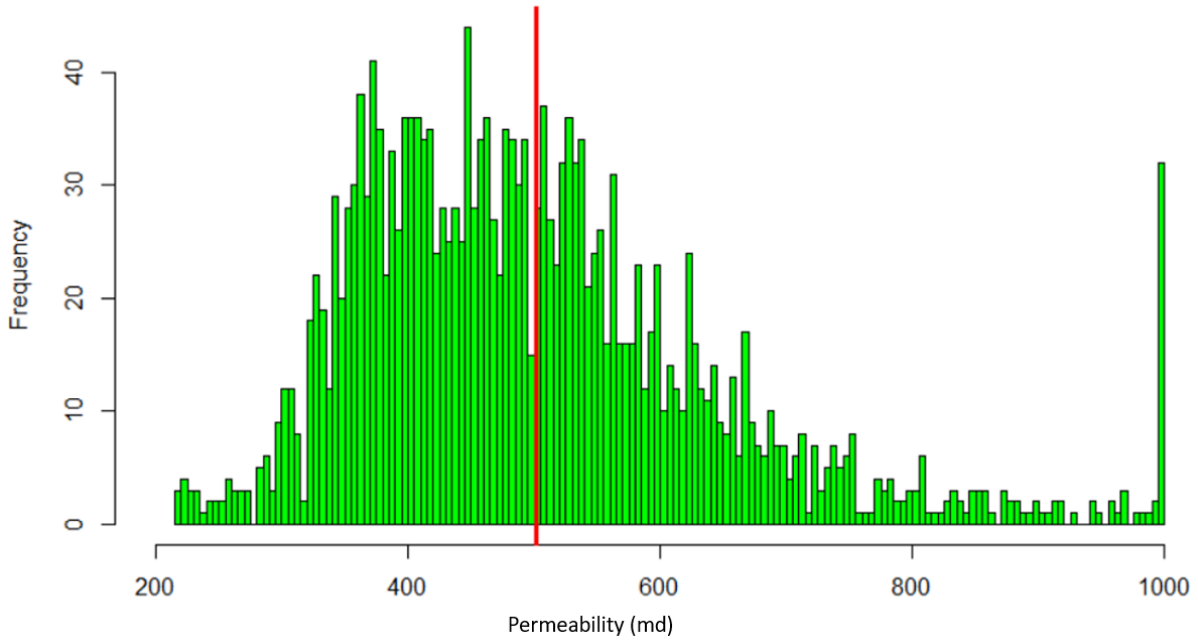

Figure 4.3. Permeability distribution realization#22

Mean = 501.0345

Standard deviation = 145.6655

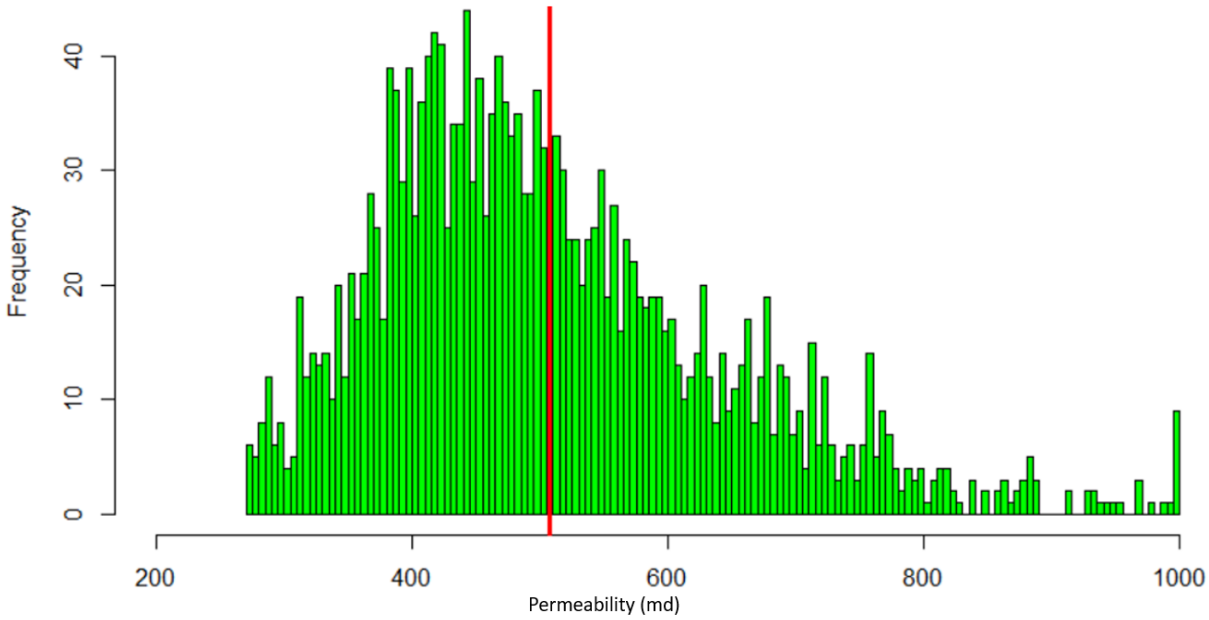

Figure 4.4. Permeability distribution realization#85

Mean = 507.3936

Standard deviation = 134.9212

As it is shown in the permeability distribution for four randomly chosen realizations, permeability values vary from almost 200 to 1000 md. The table below indicates the range of permeability for each possible situation. Permeability values illustrate that the reservoir rock considered for the model is semi-previous, consolidated rock consists of oil reservoir rocks.

Permeability	Pervious				Semi-pervious				Impervious				
Unconsolidated sand and gravel	Well sorted gravel		Well sorted sand or sand and gravel		Very fine sand, silt, loess, loam								
Unconsolidated clay and organic					Peat		Layered clay		Unweathered clay				
Consolidated rocks	Highly fractured rocks				Oil reservoir rocks			Fresh sandstone	Fresh limestone, dolomite		Fresh granite		
k (cm^2)	0.001	0.0001	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-11}	10^{-12}	10^{-13}	10^{-14}	10^{-15}
k (millidarcy)	10^{+8}	10^{+7}	10^{+6}	10^{+5}	10,000	1,000	100	10	1	0.1	0.01	0.001	0.0001

Table 4.2. Ranges of common intrinsic permeability (Bear, 2013)

12 out of 100 permeability distributions in the model has been visualized in figure below

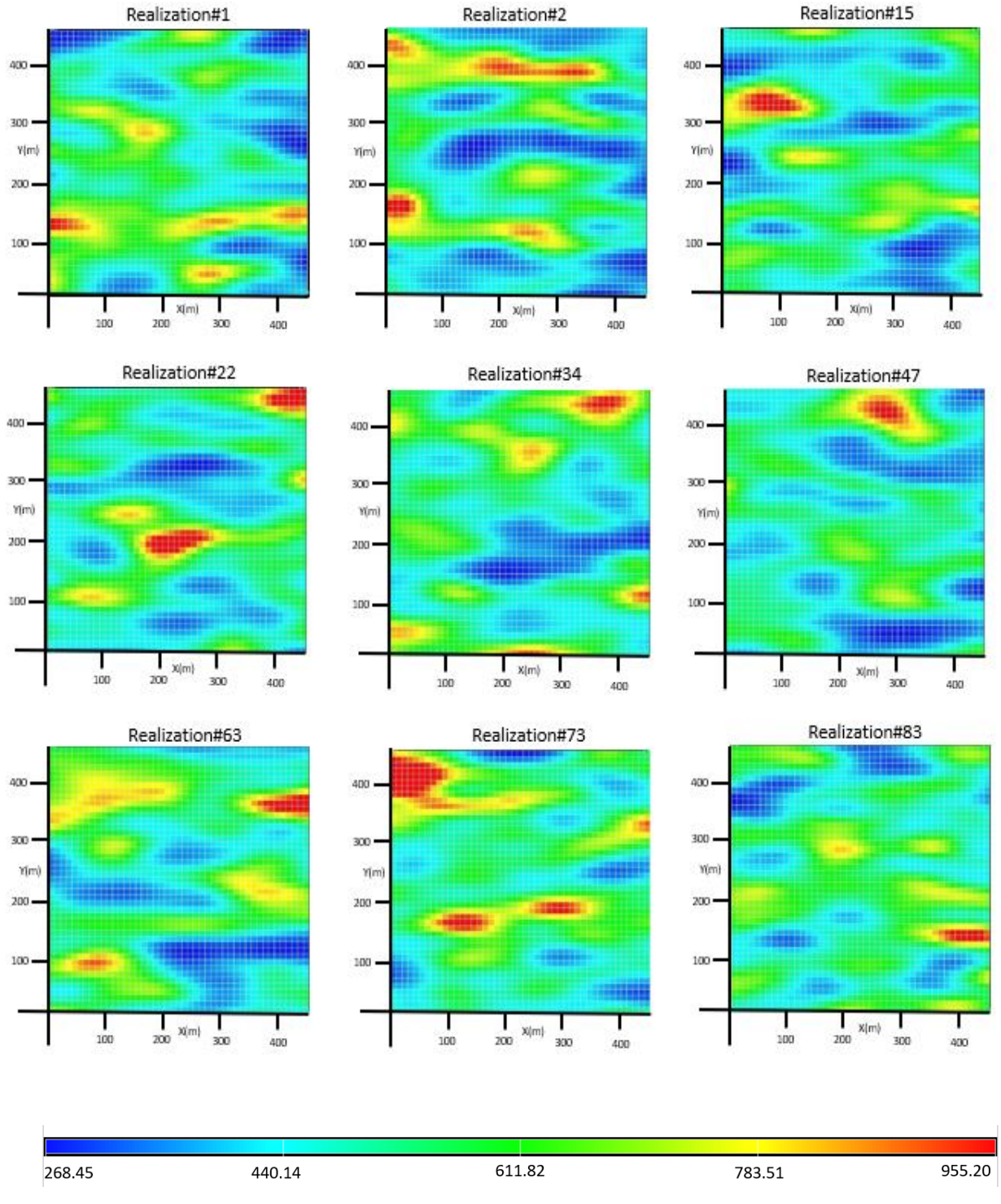

Figure 4.5. Realizations of Permeability in the Geological Model

4.5 Geological model

To study the flow a two-dimensional model in rectangular format of the reservoir is considered. It is a normal 5-spot pattern with 1 injection well in the middle and 4 production wells scatter around. The geometry has 45 cells in each direction. There are no fault transmissibility multipliers. Each cell dimension is 10 x 10 x 10 and the grid measurement is 450m on each side with 10m thickness. The thickness of the model is consisting of one cell. As the Water/Oil is two phases are considered for the model, the eclipse 100 is used for this work. The connection point of the wells considered on the top of the cell.

Model properties are shown in the table below.

Parameters	***value***
Dimension	45*45
size	10m*10m*10m
Initial pressure	234 psi
Compressibility	10^-5 Psi^-1
Porosity	21 %
Injection well coordinate	23 - 23
Water saturation	0.6

Table 4.3. Model properties

Relative permeability and wet ability in the model

It has long been recognized that the wettability of the rock is an important factor to determine the efficiency of the production and recovery in a waterflood system (Jerauld and Rathmell, 1997). The reservoir which is more oil-wet has a lower oil recovery rate in comparison with a water-wet reservoir. While there is a compromise on the least oil recovery in the waterflooding method for the oil-wet reservoirs, there is no consensus on whether the wet condition leads to maximum oil recovery. The only agreement is when the reservoir is at some intermediate-wetting state, not strongly oil-wet and not strongly water-wet. In this condition, the best oil recovery will be achieved. (wiki, 2016)

Relative permeability in the model is shown by the curves in the figure below. The oil and water relative permeability curves cross each other when Sw > 0.6. it indicates that the model is more water wet.

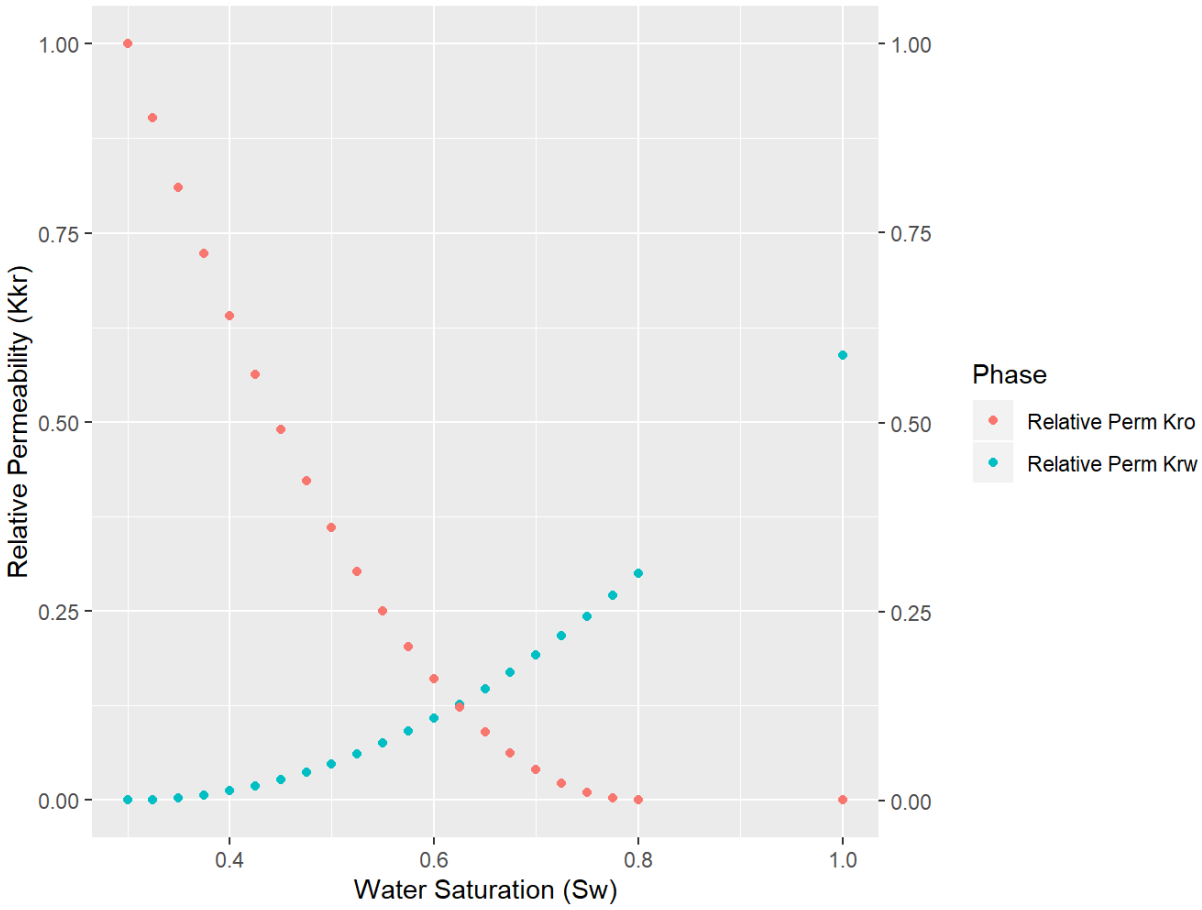

Figure 4.6. Relative Permeability Curves, water/oil System (Kor, 2019)

Summary

In this chapter we explained that permeability is the only uncertain petrophysical parameter in the model. To capture the uncertainty in permeability Sequential Gaussian Simulation (SGS) is used. There is a brief explanation of how Sequential Gaussian Simulation (SGS) generates the realizations of the permeability. Rock and fluid characteristics of the model is explained. Relative permeability in the model is also described. Geological model characteristic is indicated and relative permeability and wettability is introduced.

Chapter 5. Optimization

5.1 Introduction

Well Location Optimization is of the important factors for maximizing recovery from the reservoir. It can be identified by subsurface data and flow simulation. But information from the subsurface is never perfect and it is very uncertain and complex. This high degree of uncertainty can cause errors in estimation and leads to money loss. This situation has motivated the development of optimization processes that can help subsurface specialists to choose improved well locations (Ramirez et al., 2017).

In this thesis the optimization method which is used to find the best location for the production wells is a combination of a stochastic search algorithm and a robust field development optimization.

5.2 Robust Field Development

In this chapter, the optimization algorithm which is used for field development will be explained. The algorithm for optimization is based on Payman Kor's (2019) algorithm which later on I developed, changed and added some more codes for the purpose of my thesis. A robust and multidimensional optimization which optimizes the locations for 4 production wells and injection rate simultaneously will be used. The objective of this optimization is to optimize the expected value of the NPV function over the all geological realizations for each location by finding a control vector both for production wells coordinates and injection rate. In the process of optimization, to capture the geological uncertainties, a set of 100 realizations is used. The permeability is the only geological uncertainty considered in the simulation of the reservoir. The geostatistical method which is used to generate the realizations as described in chapter 4 is a Sequential Gaussian Simulation (SGS). Since the algorithm uses the average of all realizations to find the NPV for each location, the workflow is considered robust to geological uncertainty (Hong et al., 2017). It must be mentioned that the injection well is considered as a fixed location in the centre of the model.

In each step of the optimization process, the new generation of different locations is created by the genetic stochastic algorithm. For each location, 100 realizations will be implemented, and 100 outcomes will be produced. An average of 100 outcomes for each location will then be used to rank the locations. In the end, the algorithm reaches the stopping criteria. The best location will

be picked based on the highest estimated expected net present value and the coordinates will be identified.

Changing the setting for the optimization algorithm and applying bigger numbers for the population and iterations needs more computational capacity and time which due to limited period was not possible.

5.3 Optimization algorithm

In the process of reservoir simulation, well locations usually considered as discrete variables. Therefore the optimization algorithm which is normally used in the reservoir simulation is the Genetic algorithm. (Wang et al., 2007)

- Genetic algorithm

Genetic algorithm is used to solve the problem and developing the optimization strategy which can optimize the well locations and injection rates simultaneously.

The Genetic algorithm which is one of the sub approaches of the evolutionary algorithms is a stochastic search algorithm (Holland, 1975). It is inspired by the basic principles of biological evolution and natural selection (Affenzeller et al., 2017). The Genetic algorithm can solve both continuous and discrete optimization problems (Scrucca, 2013). It can have constraints on the parameters space as well (Yu and Gen, 2010). It can solve optimization problems by providing an exact or an approximate answer (Goldberg, 1989, Sivanandam et al., 2007).

Following (Spall, 2004), to solve the problem of finding the optimal location and injection rate, the Genetic stochastic algorithm will be implemented. This stochastic search algorithm can develop an optimization strategy to solve the optimization problem. The problem of maximizing a string number objective function f : S → R like optimization can be solved by solving equation below:

$$\Theta^* \equiv \underset{\theta \in \Theta}{argmax}\, f(\theta) = \{\theta^* \in \Theta : f(\theta^*) \geq f(\theta), \forall \theta \in \Theta\}$$

The search space is defined as $S \subseteq R^p$, when $\Theta \subseteq S$, i.e., $\theta=(\theta 1, \theta 2,..,\theta p)$ is the domain of parameters where each θi varies between lower and upper bound. The optimization problem can be considered as the equation 1, where the NPV search space is S, where the extent of the

optimization problem parameter is bounded by water injection and well location specifications. (Scrucca, 2013)

The Genetic algorithm is a procedure to resolve both constrained and unconstrained optimization problems based on a natural selection process that imitates the biological evolution. The algorithm adjusts a population of individual solutions repeatedly. At each step Genetic algorithm picks randomly individuals from the current population and uses them as parents to create the children for the next generation. In a consecutive generation, the population rises towards an optimal solution. Genetic algorithm is applicable to problems that are not suitable for standard optimization algorithms including a problem in which the objective function is stochastic, discontinuous, nondifferentiable or highly nonlinear. Genetic algorithm varies from a classical, derivative-based optimization algorithm in two main ways, which is summarized in the table below. (Mathworks, 2019)

Algorithm	
Classical	*Genetic*
A single point is generated at each iteration. The sequence of points approaches an optimal solution	A population of points is generated at each iteration. The best point in the population approaches an optimal solution
next point in the sequence will be selected by a deterministic calculation	Next population will be selected by calculation which uses random number generators

Table 5.1. genetic algorithm type (Mathworks, 2019)

For more details about Genetic algorithm, see Appendix 9.1.

5.4 Genetic algorithm setup

The Genetic algorithm package is used in a code written in R language programming to apply the optimization for the model. The model is simulated for each realization in eclipse to run the flow.

The setting for the genetic algorithm operators are shown in table below:

Parameters	***value***
Population size	10
Crossover function	0,8
Mutation function	0,1
Iteration	100
Number of unimprowed solutions	10

Table 5.2. Setup for optimization algorithm

5.5 Injection optimization

Different scenarios for injection rates are calculated as follow:

$$A * exp(-\gamma * t)$$

In this work 5 different ranges of the A value considered as the table below

100	150	200	250	300

The number of the γ values are 11. they vary from 0.0005 up to 0.002. Total number of values which is considered for the injection scenarios are 55. It is shown in the graphs below. Each injection outline is identified by it's A and γ values.

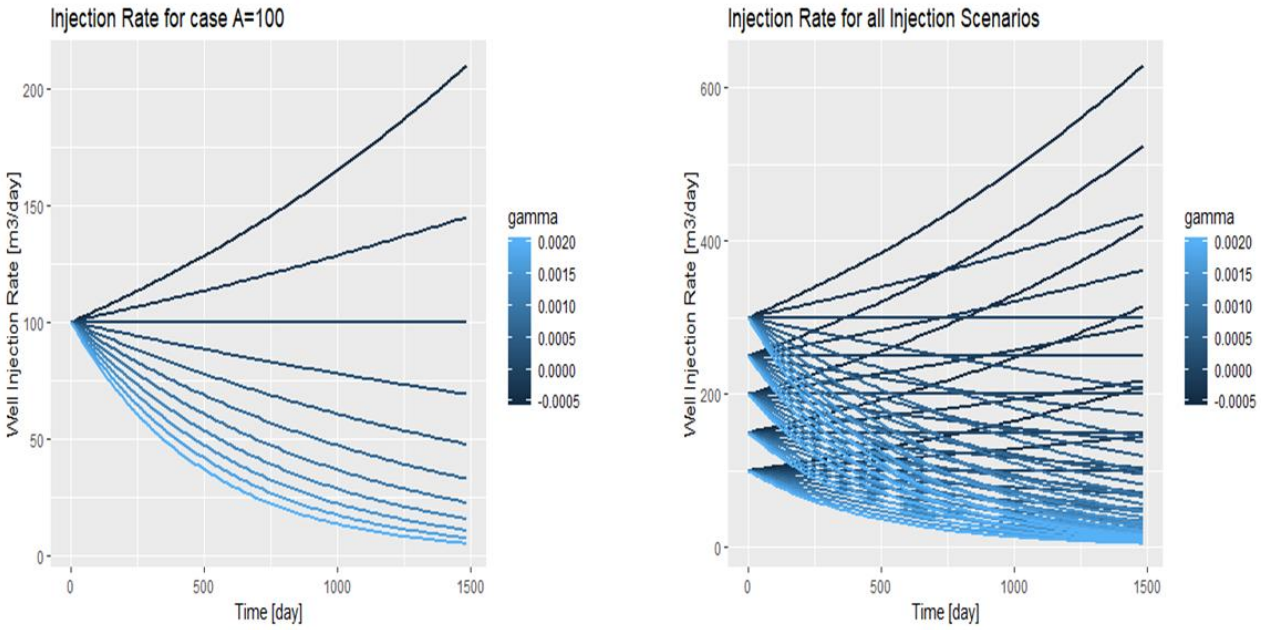

Figure 6.3. Injection scenarios (kor 2019)

5.6 NPV Calculation

Formula to calculate the NPV value is:

$$NPV = \sum_{k=1}^{n_T} \frac{[{}^{k}_{q_o}P_o - {}^{k}_{q_w}P_{wp} - I^k P_{wi}]\,\Delta_{tk}}{(1+b)^{t_k/D}}$$

${}^{k}_{q_o}$: Oil production rate at time k

${}^{k}_{q_w}$: Water production rate at time k

I^k : Water injection rate

P_o : Oil price

P_{wp} : Water production cost

P_{wi} : Water injection cost

B : Discount factor

D : Reference time for discounting (if the cash flow was discounted daily and b was expressed as fraction per year then D=365)

t_k : Cumulative time for discounting

In each iteration and for any location, an average of net present value over 100 realizations will be calculated. Algorithm then takes an average of mean NPV over all locations in each stage. Following:

$$EV(NPV_X) = \sum_{i=1}^{n_r=50} \frac{NPV_i(X)}{N}$$

Where the production well coordinates are indicated by X

Table below indicates the parameters which are considered to calculate the NPV after the flow simulation is done and the numerical result of the simulation is ready.

Prameters	***value***
Water injection cost/barrel	5
Water production cost/barrel	18
Discount factor	8 %
Oil price/barrel	60 $

Table 5.3. Setup for NPV calculator

Summary

In this chapter the optimization algorithm which helps to find the optimal coordinates for location of the production wells and optimal injection rate is explained. A robust and multidimensional optimization which optimizes the locations for production wells and injection rate simultaneously is demonstrated.

The objective of this optimization which is to optimize the expected value of the NPV function over the all geological realizations for each location, by finding a control vector both for production wells' coordinates and injection rate is explained. Parameters and setup for the optimization algorithm are introduced. Different scenarios for injection rate which is used in this optimization is indicated. Formula used to calculate the NPV value for each location and setup for the NPV calculation is described.

Chapter 6. Optimization result

Result 1

The optimization algorithm ran for several times over the different parameters and settings. After several efforts for simulating the model, with the best computational capacity available in the lab, it took several days to complete the simulation for each setting.

The first setting for the optimization algorithm consists of 10 members of the population (coordinates) and 100 iterations. Each stage or iteration of the algorithm consists of 10 locations. For each location 100 realizations of permeability are considered. The algorithm generates the location coordinates and the Eclipse simulates the location for each realizations. The output of the algorithm is an expected net present value over all members of the population in that stage. It also gives the highest expected net present value for each step. In the end if the highest expected net present value repeated for 10 times, algorithm reaches the stopping criteria and stops the optimization. It gives the coordinates for the location of the wells which has the highest net present value as the optimal location.

The algorithm transferred the location for the highest expected net present value to the next step. Therefore, if the NPV of the last set of the locations does not exceed the previous step, the best location for the previous step is considered for the next one.

In here, after 47 iterations optimization algorithm reached the stopping criteria. By considering 100 realizations for each location, and 10 locations in each step, the number of simulations becomes 47000. Optimization improvement is shown in the figure 6.1.

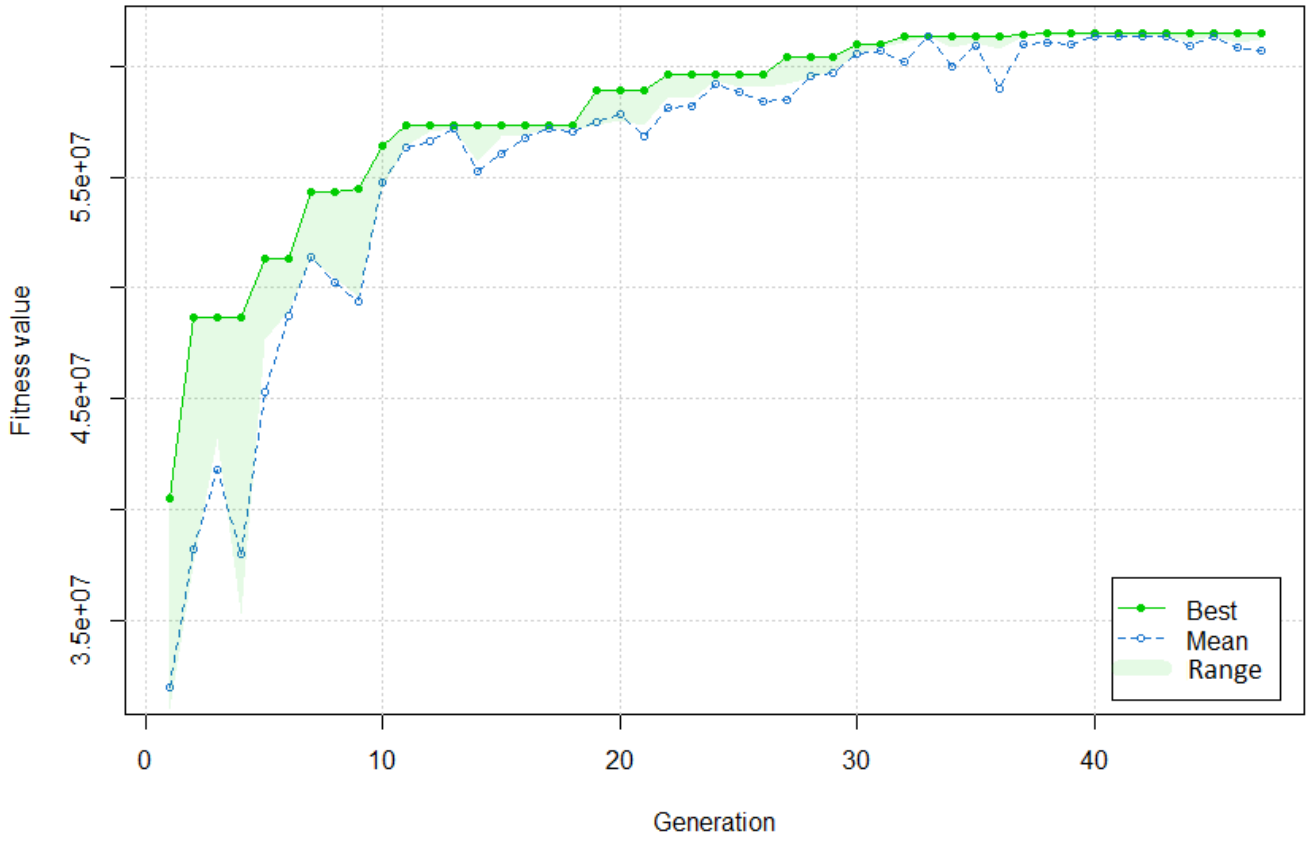

Figure 6.1. Optimization development

In the graph, the green dots are indicators of the highest expected net present value for each step or iteration. In each step average of the expected net present value for all locations is shown by blue dots. As the suggested location coordinates are not considered to start the simulation, and the algorithm starts by selecting random locations from the beginning, starting values of NPV are different in all trial simulations over the model.

Table below indicates the NPV values for the best location over the 100 realizations. After 47 iterations the maximum expected value which is derived from average NPV of all 100 realizations at the best location is 61.4MM.

NPV Values for 100 Realizations Over the Best Location									
6052594	6360644	6234746	6224026	6111391	6002485	6140261	6096553	6089916	6181755
6055478	6103638	6255931	6214642	6071156	6118823	5845982	6257287	6197697	6103030
6202594	6246279	6191577	6004031	6098463	6275570	6044005	5978986	6120684	6150643
6250245	6194010	6062658	6149987	6044417	6137133	6110605	6020554	6052248	6324517
6147739	6223873	6167189	6084365	6151637	6032706	6094955	5968341	6120157	6144134
6029591	6134562	5993018	6184775	6148892	6162333	6157993	6234270	6047851	6010849
6180075	6033519	6294169	6176299	6120503	6240170	6101513	6206358	6247901	6258179
6131064	6128331	6063279	6141844	6114115	6240157	6208162	6209481	6117667	6165362
6235686	6268356	5886739	6281071	6114949	6369377	6120683	6372977	6198354	6059929
5998548	6064644	6138962	6147940	6116330	6123210	6100771	6220915	6154400	6069532

Table 6.1. NPV values for the optimal location

The figure below is the histogram over the 100 realizations for the optimal location. Mean NPV is indicated by the red line. The expected net present value and standard deviation value are written below the figure.

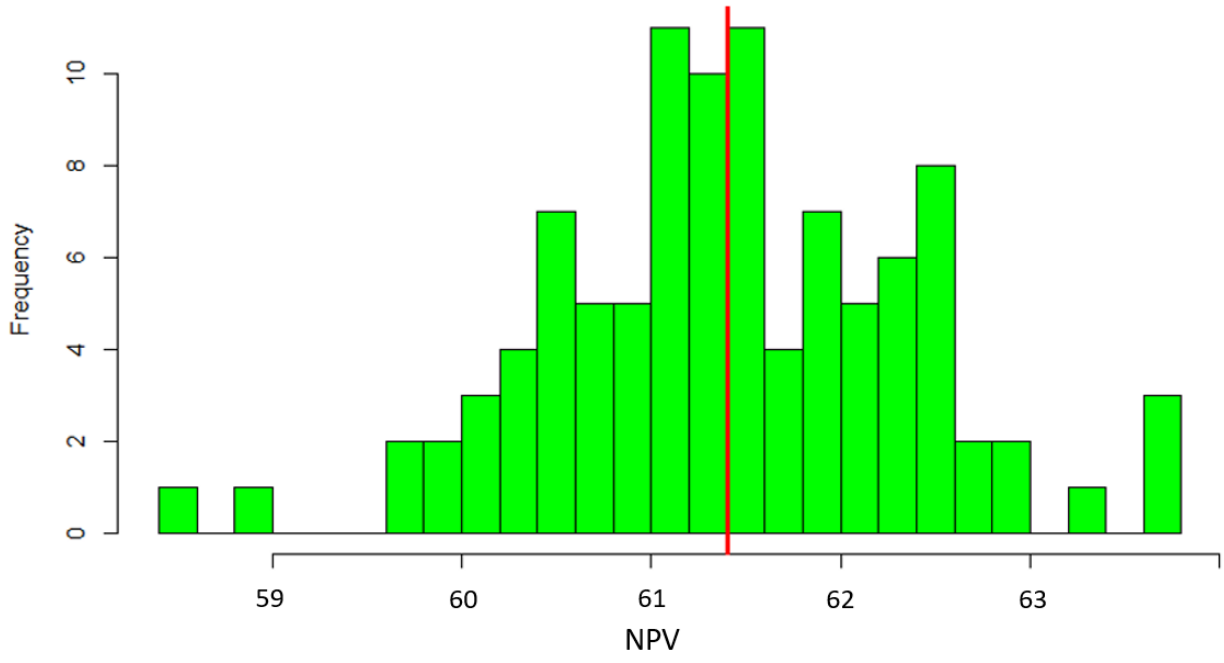

Figure 6.2. Histogram of the NPV values for the optimal location

Mean of best location NPV = 61,4$MM

Standard deviation of best location NPV = 956511.17

Table 6.2 is expected net present value for each step in the optimization process.

NPV Mean Values for 47 Stages of the Simulation									
3197704	4876127	5629838	5674269	5686544	5839377	6072585	5900175	6137177	6086820
3823489	5137071	5664325	5717558	5813438	5845449	6022738	6099645	6137177	6070073
4180553	5027727	5720542	5705600	5819633	5954816	6132419	6106932	6136288	
3801033	4941065	5524947	5750306	5918926	5970157	5999783	6096168	6092933	
4533131	5477841	5601072	5780817	5884089	6055076	6092184	6137177	6134900	

Table 6.2. Mean NPV values for each step in optimization

Mean NPV over the all iterations = 56.1$MM

Standard deviation evaluated for all steps = 56.1MM

The graph for the cumulative net present values for the optimal location is shown in figure 6.3

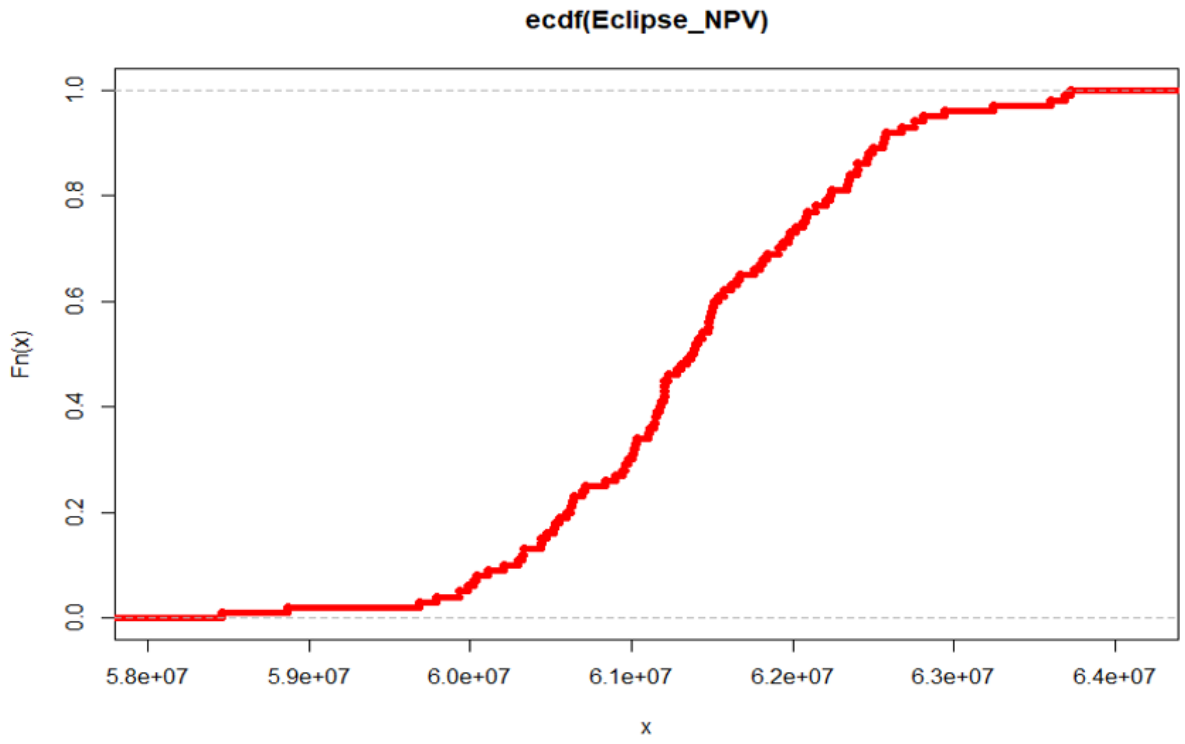

Figure 6.3. cumulative distribution plot of NPV values for optimal location

In table 6.3 the best location coordinates are shown. It also indicates the optimal injection Gamma rate and starting injecting value.

Prameters	***value***
coordinate for injection well	23 – 23
coordinate for production No: 1	7.419 – 43.519
coordinate for production No: 2	41.054 – 14.173
coordinate for production No: 3	4.594 – 9.894
coordinate for production No: 4	39.724 – 42.924
Injection Gamma rate	0,00103
Injection Starting rate	264.53

Table 6.3. Best location coordinates & injection rate

Result 2

This second result is gained by setting the parameters for the optimization in table 6.4.

Parameters	***value***
Population size	100
Iteration	1
Number of unimprowed solutions	-

Table 6.4. Setup for the optimization algorithm

Because it has just one step, the output is consisted of just one average over 100 expected net present values and one optimal location. therefore there is no plot for the optimization improvement as it was for the previous one.

The net present values over the 100 realization for optimal location are shown in table 6.5.

NPV Values for 100 Realizations Over the Best Location									
5296270	5300354	5338802	5288182	5310145	5433785	5516410	5368212	5330215	5235959
5323057	5255080	5380658	5318128	5384469	5384571	5283340	5130347	5352586	5186576
5244602	5358805	5276768	5036706	5351140	5256291	5232307	5203303	5346559	5166658
5203502	5289474	5169787	5346778	5162720	5326781	5105324	5303148	5238124	5282897
4928254	5426267	5284647	5323005	5347337	5403382	5375683	5303980	5179196	5262219
5279337	5421428	5195119	5327573	5239606	5121630	5260979	5244419	5173485	5248367
5109043	5244120	5323900	5267151	5434761	5278262	5231623	5218752	5387233	5158611
5228058	5246136	5284653	5417103	5484466	5277862	5389905	5360848	5396378	5257155
5445296	5336834	5365773	5422899	5323220	5425483	5196356	5194271	5324688	5280383
5357997	5284241	5357528	5453660	5292016	5395171	5271080	5255322	5303282	5361545

Table 6.5. Mean NPV values for each location

The histogram of the net present value distribution for the best location is shown in the figure 6.4. The red line indicates the estimated expected net present value or the mean for the distribution:

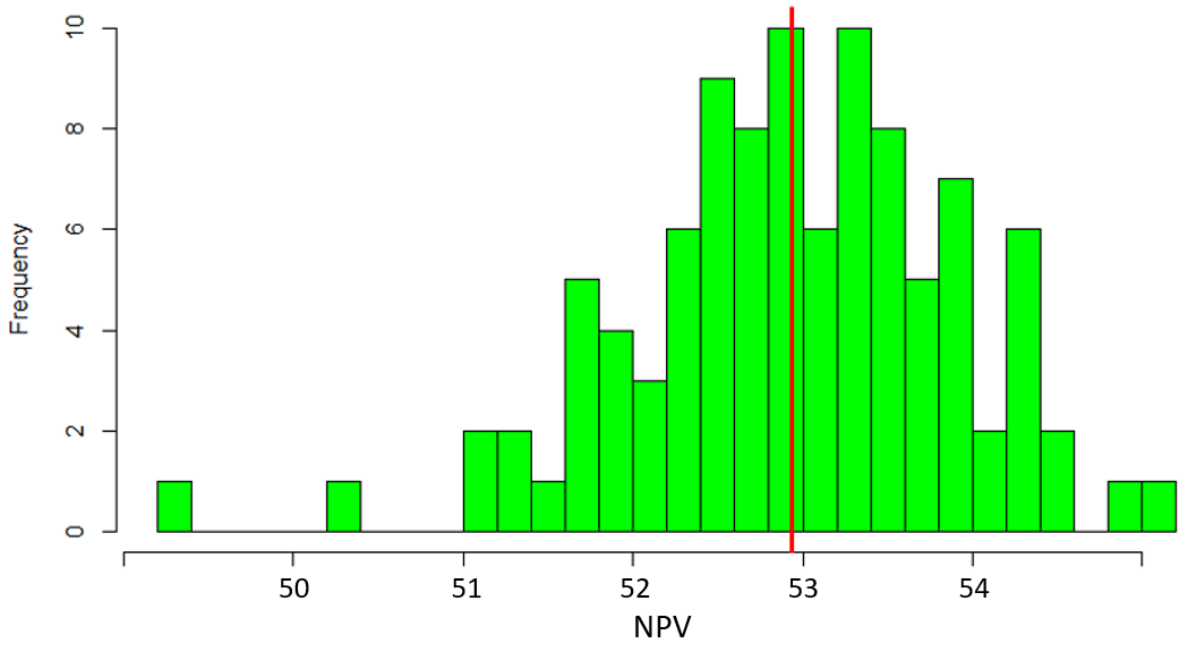

Figure 6.4. Histogram of the NPV values for the optimal location

Mean of best location NPV = 52930785

Standard deviation of best location NPV = 966900.2

The cumulative Net present value for the optimal location is shown in figure 6.5.

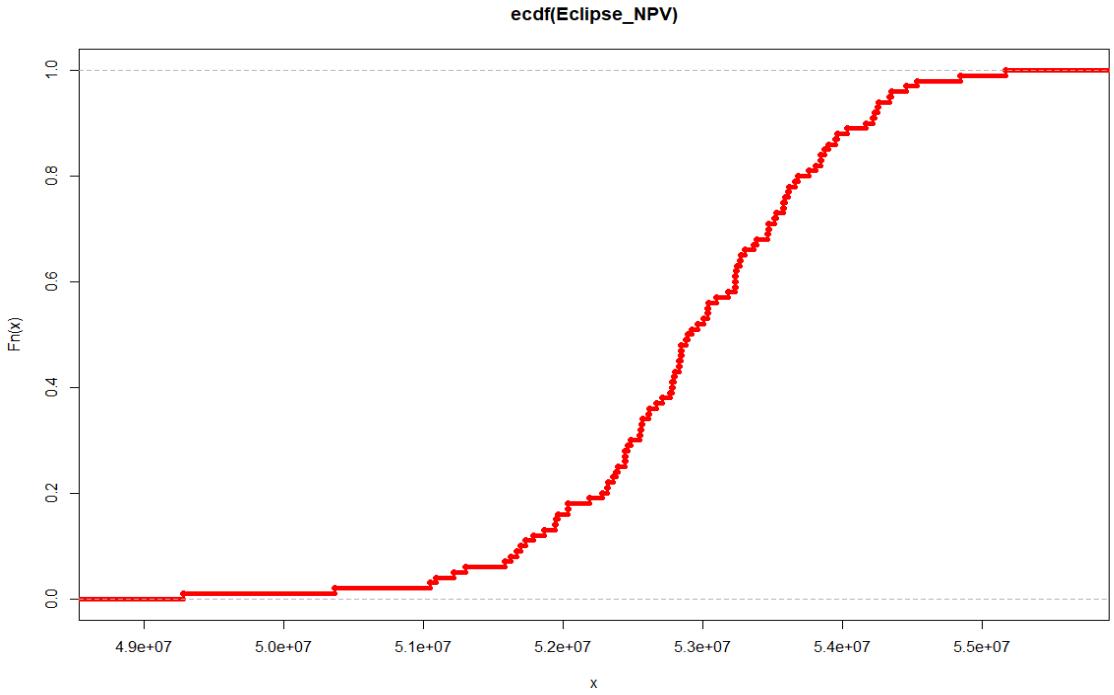

Figure 6.5. Cumulative Net present value

The expected net present value over 100 different locations = 32758210

The result for the best location coordinate and injection rate based on gamma and starting value is shown in table 6.6.

Prameters	***value***
coordinate for injection well	23 – 23
coordinate for production No: 1	37.089 – 5.247
coordinate for production No: 2	7.207 – 14.059
coordinate for production No: 3	2.383 – 7.053
coordinate for production No: 4	40.277 – 37.831
Injection Gamma rate	0,00124
Injection Starting rate	223.441

Table 6.6. Best location coordinates

Chapter 7. Discussion

Introduction

The focus of this work is on optimizer's curse or the expected disappointment in the optimization of well location in a 5-spot pattern. The optimization process which is used in this work is done by simulating the locations for different geological realizations and taking an average over all values for NPV. Then it selects the optimal location based on the highest mean NPV. This procedure repeated for each stage until it reaches the stopping criteria. In the end, the highest expected net present value is the indicator of the optimal location for the wells and the best injection rate is identified for this location. The process of optimization in this work simply selects and ranks the alternatives which have the highest estimated expected net present value. these estimated values are combined with errors. The error might come from the model or the uncertainties.

Having estimated values for optimal location and for all locations makes it possible to check the model for expected disappointment. Then the estimated values can be corrected by a Bayesian method which is explained in chapter 3.

Discussion I

We can check the magnitude of post-decision disappointment by using the Bayesian model developed in chapter 3. Standard procedure in robust optimization draws random values from the prior and designate it as the true value following:

$V_i = \mu_i + E_i$

By using equation from chapter 3 expected net present value estimated for optimal location, can be adjusted for Bayes. It should be mentioned that as the model is simple and the uncertainty is covered by 100 realizations for permeability, the possible gap between estimated value and the corrected value for Bayes is expected to be low. It indicates that the magnitude of expected disappointment should be low.

Figure 7.1 illustrates the distributions of the net present values estimated for the optimal location and distribution of the expected net present values for all simulations.

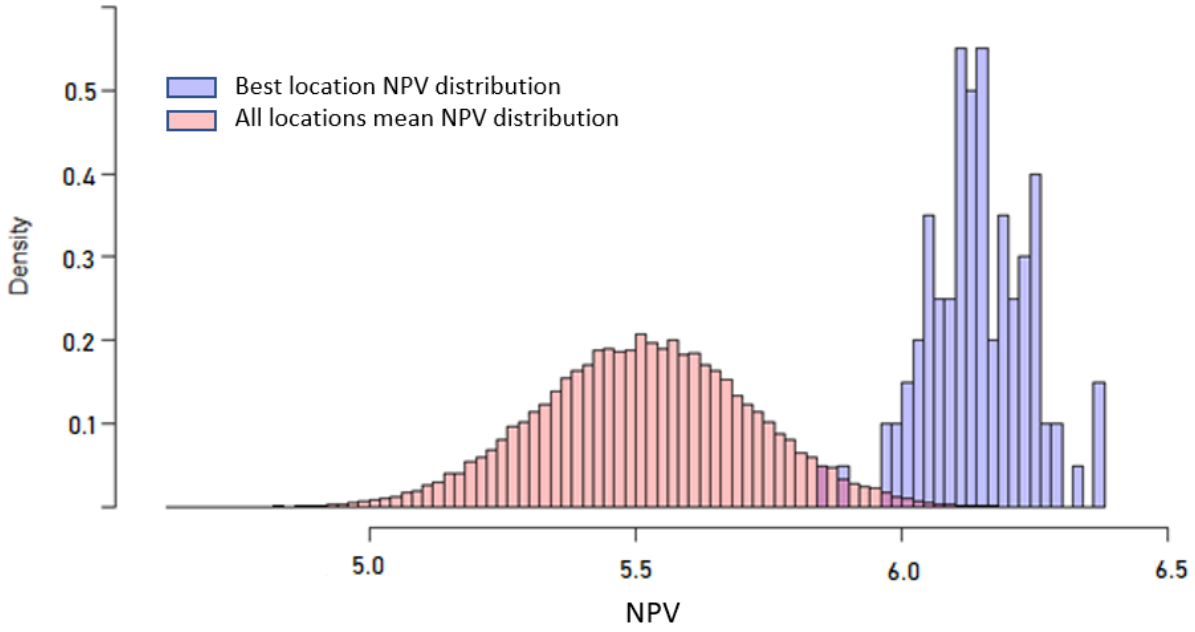

Figure 7.1. NPV distribution

In the graph, red histogram indicates the distribution of the all simulations and the blue histogram illustrates the distribution of NPV over 100 realizations for the best location.

To correct the estimated value for the optimal location following chapter 3:

$$\alpha_i = \left(1 + \frac{\tau_i^2}{\sigma_i^2}\right)^{-1}$$

$$v_i^B = \alpha_i \, v_i^E + (1 - \alpha_i) \, \mu_i$$

Having the standard deviation for the distribution of all simulations and the best location, α_i can be calculated like

$$\tau = 0.95651117MM$$

$$\sigma = 1.77854756MM$$

$$\alpha_i = \left(1 + \frac{\tau^2}{\sigma^2}\right)^{-1} = \left(1 + \frac{0.95651117^2}{1.77854756^2}\right)^{-1}$$

Having the expected net present value for the optimal location (v_i^E) and for all steps (μ_i).

$$v_i^E = 61.4\$MM$$

μ_i = $56.1\$MM$

Bayes corrected estimate v_i^B can be calculated as follow:

$v_i^B = \left(1 + \frac{0.95651117^2}{1.77854756^2}\right)^{-1} * 61.4 + (1 - \left(1 + \frac{0.95651117^2}{1.77854756^2}\right)^{-1}) * 56.1 = 60.23\$MM$

v_i^B = $60.7\$MM$

As it is shown the difference between the estimated expected net present value and Bayes correction for that is not very big as it was expected.

(I) $v_i^E - v_i^B = 61.4 - 60.2 = 1.17\$MM$

It is almost 1.97 percent of the expected net present value estimated for the best location.

In chapter 2, the effect of uncertainty reduction is explained. Here, If the distribution of all simulation values had a bigger standard deviation in compare to estimated one α_i was closer to 1. It proves that the estimated value has more weight in the Bayes equation than the realized value and the Bayes corrected estimation value would be closer to the value estimate.

By increasing the standard deviation of realized values, the separation gap between realized values increases and it leads to a decrease in the magnitude of the optimizer's curse. In fact, if there were more uncertainty in the input data for the model, it would be expected to have a bigger gap between the estimated value and corrected one for bias. For instance, transmissibility or more realizations of permeability and porosity increases the uncertainty.

If the resulted expected net present value for the optimal location was lower than the realized net present value, then the Bayes correction for that value is expected to show a bigger number. It shows that the value is underestimated, and Bayes correction value is higher than the initial estimation.

Discussion II

In chapter 5 the process of generating the new population by a Genetic algorithm is explained (for more detailed explanation see appendix 9.1). Only in the first step of the Genetic algorithm the generates population (coordinates for locations) randomly. For the next steps, the algorithm uses different methods such as combining the genes from the previous generation to produce a new population and transferring the best of the previous step to the next population. Therefore, the

algorithm in each step is closer to the optimal value. To explain how the Genetic algorithm works, we imagine a target and arrows with an aim to hit the bullseye. On the left-side figure (7.2a), arrows clustered tightly and off the bullseye to the right side. Since it is tightly clustered, it indicates aiming with a steady hand that generates just small random errors. It clustered to the right because the bow's sight might be off and affects all arrows. It shows precise but inaccurate targeting. Inaccuracy represents bias in the aiming. By contrast in the right-side figure (7.2b), targeting with an unsteady hand which produced large random errors into every shot. The scatter pattern of shots represents imprecision but as the cluster is centred on the bullseye, it indicates that the shots are accurate or unbiased (Summers, 2019).

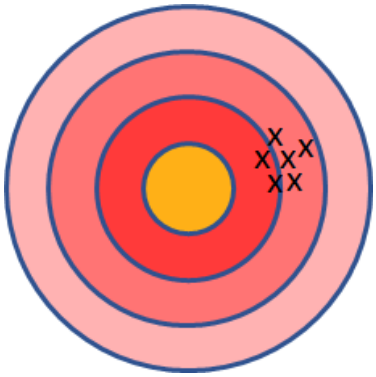

Figure 7.2a

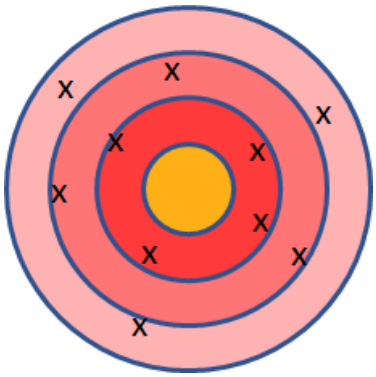

Figure 7.2b

By looking at location coordinates for discussion I and II, it appears that when the algorithm runs for more steps, the location of the wells goes towards the corners of the model (Table 6.3). In contrast when it uses only the first step, we can have the well locations spread all over the model. In this case two of the location coordinates are too close (Table 6.6).

To study the magnitude of expected disappointment when selecting the alternatives accurately but not precisely, the optimization is done with different parameters. Second attempt for the optimization is done by just one step and a population of 100.

The result for the optimal location expected net present value v_i^E and expected net present value for all locations μ_i are:

v_i^E = 52.9$\$MM$

μ_i = 32.7$\$MM$

Figure 7.3 shows the distribution of expected net present value for all 100 locations and net present values for the optimal location.

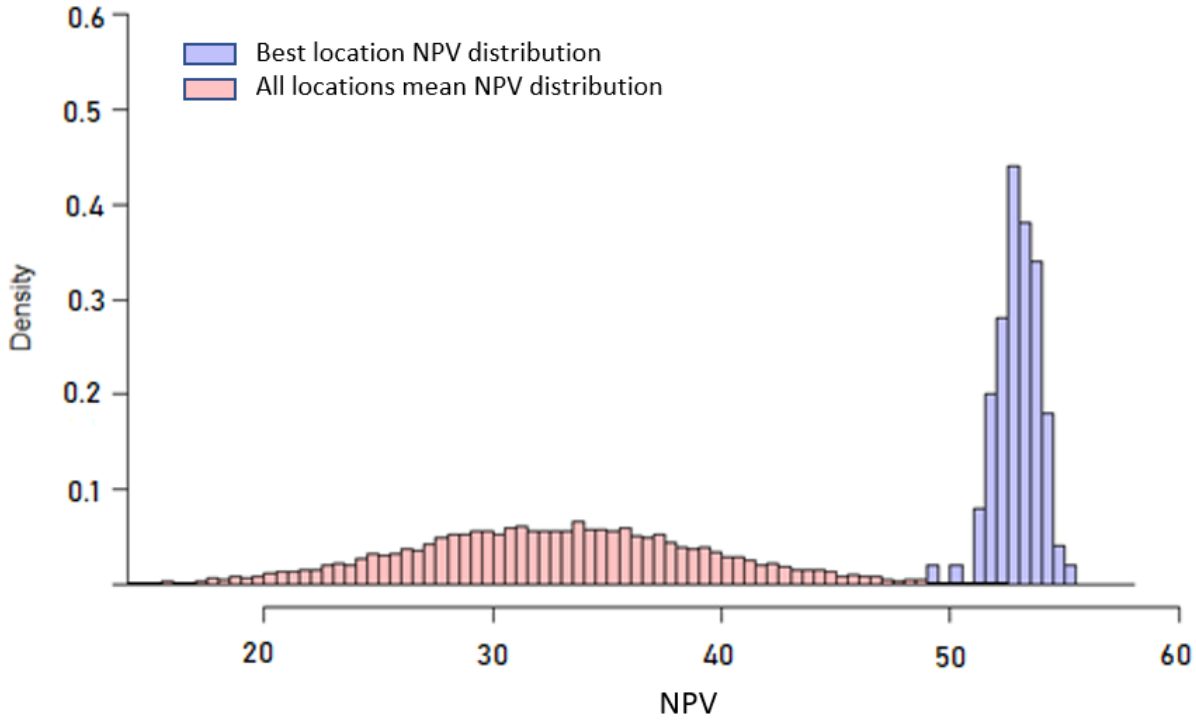

Figure 7.3. NPV distribution

Having standard deviation for the optimal location τ ,over 100 realizations and the standard deviation for the 100 expected net present values σ for all locations, the Bayes estimation would be:

$\tau = 0.9669002MM$

$\sigma = 5.467260MM$

$v_i^B = 52.3MM$

(II) $v_i^E - v_i^B = 52.9 - 52.3 = 0.65\$MM$

It is almost 1.15 percent of the value estimated for the best location, which is lower than the condition with several steps (first discussion). It can be explained by having a bigger difference or separation among alternatives. But we should consider that the NPV values for the best location and for all locations are lower than the first attempt. In other words, the resulted coordinate for the best location is not good as the first attempt.

Discussion III

In the first optimization attempt which has 47 iterations, possible post-decision disappointment is investigated by considering the μ_i over all steps or iterations. In the second attempt (section II), possible overestimation is investigated for a single step. In third attempt (III) I tried to check the post-decision disappointment by considering only the step or iteration which gives the best location.

v_i^E = $61.4\$MM$ & τ = $0.95651117MM$

μ_i = $60.99MM$ & σ = $0.9777726MM$

Value estimate corrected for Bayes is:

v_i^B = $61.00MM$

(III) v_i^E - v_i^B = 61.4 - 61.00 = $0.4\$MM$

By comparing the result for the attempt II and III, the magnitude of the optimizer's curse is reduced as the genetic algorithm generates more locations and the standard deviation over the population expected net present value becomes closer to the expected net present value of the best location. In section (II) all members of the population were selected randomly so there is a bigger standard deviation for all members of the population. But in section (III), the standard deviation for the net present values over all members of the population is smaller, and the resulted best-expected net present value estimation is closer to the Bayes corrected estimated value.

Chapter 8. Conclusion

Conclusion

The primary goal of this work is to study the optimizer's curse on the concept of optimization of well location in a 5-spot pattern and how optimizer's curse affects the result of this optimization. Selection technics which are used in optimization process leads to the creation of some artefact which is called Optimizer's curse. It has no psychological dimension (Marks, 2008) and appears in any decision-making that needs to select among alternatives in which the true values are unknown and values are just estimations. It is not caused by any inherent bias in the estimates, but the product of the selection process itself.

Optimizer's curse happens when errors inside the estimations coupled with the optimization-based process. It has been demonstrated in this work; even unbiased evaluation errors lead to results impacted by optimizer's curse. These random errors, which are included in the assessment of the alternatives, can be reduced by collecting more data to reduce uncertainty in value estimates. Examples have demonstrated that the curse can be significant in practice and the expected disappointment would be substantial when the value estimates are biased.

To overcome this curse, implementing a Bayesian method and modelling of estimation uncertainties helps to debias the estimations. Considering estimation results as uncertain and use Bayes rule to combine it with prior information helps to correct the value estimates for the bias which is built in the optimization process. (Smith and Winkler, 2006)

failure to identify and modify this curse leads to a substantial loss in a project assessment. It leads to spending more money to achieve an overestimated project portfolio. And the actual outcome is likely to be significantly lower than forecasted.

It is important to consider this curse in the project evaluation and understand how it affects the process of selecting the alternatives. Otherwise, the difference between the estimated alternative and the realized value of it can be expressed as systematic underperformance of the project or poor or deliberate misleading of the outcomes and costs.

In conclusion, it must be mentioned that if the optimizer's curse is not removed or corrected, most portfolio evaluations are overestimated, and the decision-maker may face post-decision disappointment after implementing the selected alternative.

Recommendation

To study and investigate the optimizer's curse by working with REAL reservoir model which is more complicated is recommended. Using field data to study this curse would be interesting. It is difficult to study the optimizer's curse in real cases because usually companies do not publish their value estimates which is used to make decisions. Having an opportunity to work with such data will be helpful to study the optimizer's curse and other factors which affects the predictions.

Study the optimizer's curse with conducting some experiment would be interesting. It can be done by estimating values for a complex problem with different alternatives and select the optimal one, then execute the selected alternative to study the result for possible expected disappointment.

References

AFFENZELLER, M., WINKLER, S. M., BURLACU, B., KRONBERGER, G., KOMMENDA, M. & WAGNER, S. Dynamic observation of genotypic and phenotypic diversity for different symbolic regression GP variants. Proceedings of the Genetic and Evolutionary Computation Conference Companion, 2017. ACM, 1553-1558.

AHTI SALO, JEFFREY KEISLER & ALEC MORTON 2011. An invitation to portfolio decision analysis. *In:* AHTI SALO, JEFFREY KEISLER & ALEC MORTON (eds.) *Portfolio Decision Analysis: Improved Methods for Resource Allocation.* 9781441999436: Springer-Verlag.

BEAR, J. 2013. *Dynamics of fluids in porous media*, Courier Corporation.

BEGG, S. H. & BRATVOLD, R. B. Systematic Prediction Errors in Oil and Gas Project and Portfolio Selection. SPE Annual Technical Conference and Exhibition, 2008. Society of Petroleum Engineers.

CAPEN, E. 1976. The Difficulty of Assessing Uncertainty (includes associated papers 6422 and 6423 and 6424 and 6425). *Journal of Petroleum Technology,* 28**,** 843-850.

DELQUIÉ, P. 2008. The value of information and intensity of preference. *Decision Analysis,* 5**,** 129-139.

GAO, Y. 2019. *What Is Sequential Gaussian Simulation* [Online]. Knowledgette LLC. 2225 E Bayshore Rd #200, Palo Alto, CA, 94303.: knowledgette. Available: https://www.knowledgette.com/courses/what-is-sequential-gaussian-simulation/lectures/2253852 [Accessed].

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. & RUBIN, D. B. 2013. *Bayesian data analysis*, Chapman and Hall/CRC.

GOLDBERG, D. 1989. Genetic algorithms in search, optimization, and machine learning. Boston, MA: Addison-Wesley.

HARRISON, J. R. & MARCH, J. G. 1984. Decision making and postdecision surprises. *Administrative Science Quarterly***,** 26-42.

HAWKINS, J., COOPERSMITH, E. & CUNNINGHAM, P. Improving stochastic evaluations using objective data analysis and expert interviewing techniques. SPE Annual Technical Conference and Exhibition, 2002. Society of Petroleum Engineers.

HODGKINSON, G. P., BOWN, N. J., MAULE, A. J., GLAISTER, K. W. & PEARMAN, A. D. 1999. Breaking the frame: an analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal,* 20**,** 977-985.

HOLLAND, J. 1975. Adaptation in natural and artificial systems: an introductory analysis with application to biology. *Control and artificial intelligence*.

HONG, A., BRATVOLD, R. & NÆVDAL, G. 2017. Robust production optimization with capacitance-resistance model as precursor. *Computational Geosciences,* 21**,** 1423-1442.

JERAULD, G. & RATHMELL, J. 1997. Wettability and relative permeability of Prudhoe Bay: A case study in mixed-wet reservoirs. *SPE Reservoir Engineering,* 12**,** 58-65.

JOHNSTON, K., VER HOEF, J. M., KRIVORUCHKO, K. & LUCAS, N. 2001. *Using ArcGIS geostatistical analyst*, Esri Redlands.

KAHNEMAN, D., SLOVIC, S. P., SLOVIC, P. & TVERSKY, A. 1982. *Judgment under uncertainty: Heuristics and biases*, Cambridge university press.

KAHNEMAN, D. & TVERSKY, A. 1973. On the psychology of prediction. *Psychological review,* 80**,** 237.

KAHNEMAN, D. & TVERSKY, A. 1977. Intuitive prediction: Biases and corrective procedures. Decisions and Designs Inc Mclean Va.

KEISLER, J. 2004. Value of information in portfolio decision analysis. *Decision analysis,* 1**,** 177-189.

KOR, P. 2019. *Decision-Driven Data Analytics for Well Placement Optimization in Field Development Scenario-Powered by Machine Learning.* University of Stavanger, Norway.

KRITZMAN, M. 2011. The graceful aging of mean-variance optimization. *Journal of Portfolio Management,* 37**,** 3.

KUHN, D. 2018. *Optimization in Energy Systems* [Online]. Copenhagen, Denmark: DTU CEE Lectures. Available: http://podcast.llab.dtu.dk/feeds/dtu-cee-summer-school-2018/ [Accessed].
MAKIDO, Y. K., MESSINA, J. P. & SHORTRIDGE, A. M. 2008. Exploring the impacts of pseudo-random number generators on sub-pixel classification. *GIScience & Remote Sensing,* 45**,** 471-489.
MARKS, R. E. 2008. The satisficer's curse. *Available at SSRN 1136039*.
MATHWORKS. 2019. *Find global minima for highly nonlinear problems* [Online]. united states: The MathWorks, Inc. Available: https://se.mathworks.com/discovery/genetic-algorithm.html [Accessed].
MOHUS, E. 2018. *Over Budget, Over Time, and Reduced Revenue, Over and Over Again-An Analysis of the Norwegian Petroleum Industry's Inability to Forecast Production.* University of Stavanger, Norway.
MONTIBELLER, G. & VON WINTERFELDT, D. 2015. Cognitive and Motivational Biases in Decision and Risk Analysis. *Risk Analysis,* 35**,** 1230-1251.
MOORE, D. A. & HEALY, P. J. 2008. The trouble with overconfidence. *Psychological review,* 115**,** 502.
MOREHOUSE, D. F. 1997. The intricate puzzle of oil and gas reserves growth. *Natural Gas Monthly***,** vii-xx.
MORGAN, M. G., HENRION, M. & SMALL, M. 1990. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*, Cambridge university press.
NANDURDIKAR, N. S. & WALLACE, L. 2011. Failure to Produce: An Investigation of Deficiencies in Production Attainment. *SPE Annual Technical Conference and Exhibition.* Denver, Colorado, USA: Society of Petroleum Engineers.
PEBESMA, E. J. & WESSELING, C. G. 1998. Gstat: a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences,* 24**,** 17-31.
PYRCZ, M. J. & DEUTSCH, C. V. 2014. *Geostatistical reservoir modeling*, Oxford university press.
RAMIREZ, B. A., JOOSTEN, G. J. P., KALETA, M. P. & GELDERBLOM, P. P. 2017. Model-Based Well Location Optimization – A Robust Approach. *SPE Reservoir Simulation Conference.* Montgomery, Texas, USA: Society of Petroleum Engineers.
SCHUYLER, J. R. & NIEMAN, T. Optimizer's Curse: Removing the Effect of this Bias in Portfolio Planning. Hydrocarbon Economics and Evaluation Symposium, 2007. Society of Petroleum Engineers.
SCRUCCA, L. 2013. GA: a package for genetic algorithms in R. *Journal of Statistical Software,* 53**,** 1-37.
SIVANANDAM, S., SUMATHI, S. & DEEPA, S. 2007. *Introduction to fuzzy logic using MATLAB*, Springer.
SMITH, J. E. & WINKLER, R. L. 2006. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science,* 52**,** 311-322.
SPALL, J. 2004. Stochastic Optimization, Handbook of Computational Statistics. Springer Heidelberg.
SUMMERS, G. 2019. Product Development: The Optimizer's Curse. *Pharmexec.* Advanstar Communications.
TREK, S. 2019. *Standard Normal Distribution* [Online]. Available: https://stattrek.com/probability-distributions/standard-normal.aspx [Accessed].
VILKKUMAA, E., LIESIÖ, J. & SALO, A. 2014. Optimal strategies for selecting project portfolios using uncertain value estimates. *European Journal of Operational Research,* 233**,** 772-783.
WANG, C., LI, G. & REYNOLDS, A. C. 2007. Optimal Well Placement for Production Optimization. *Eastern Regional Meeting.* Lexington, Kentucky USA: Society of Petroleum Engineers.
WELSH, M. B., BRATVOLD, R. B. & BEGG, S. H. Cognitive biases in the petroleum industry: Impact and remediation. SPE Annual Technical Conference and Exhibition, 2005. Society of Petroleum Engineers.
WICKHAM, H. & WICKHAM, M. H. 2016. Package 'plyr'. *Obtenido de https://cran. rproject. org/web/packages/dplyr/dplyr. pdf*.
WIKI, O. F. 2016. *Wettability* [Online]. Available: http://www.oilfieldwiki.com/wiki/Wettability#cite_ref-1 [Accessed].

WRIGHT, A. H. 1991. Genetic algorithms for real parameter optimization. *Foundations of genetic algorithms.* Elsevier.
YU, X. & GEN, M. 2010. *Introduction to evolutionary algorithms*, Springer Science & Business Media.

Chapter 9. Appendix

9.1 Genetic algorithm

- How the genetic algorithm works

The Genetic algorithm is a subcomponent of the evolutionary strategies. At each evolutionary iteration, a population is generated. This population is composed of some individuals which are called chromosomes or strings and each chromosome is made of genes that control the characters inherited from the parents. Inherited genes of a certain character are placed along the chromosome and the corresponding string positions are called Loci. Each genotype would be a possible solution to the optimization problem. In the Genetic algorithm, the decision variable or phenotypes are achieved by applying some mapping from the chromosome representation into the decision variable space, which represents a potential solution to the optimization problem. A proper decoding function may be required for mapping chromosomes onto phenotypes. (Scrucca, 2013)

The algorithm first creates a random initial population which can be either a suggestion for the first location or takes the first location randomly and then creates several locations based on the population size in the algorithm.

In the next stage the algorithm generates a sequence of new populations. At each step, algorithm uses past locations to generate the new population. Genetic algorithm process is described as below.

- The fitness value of each member of the population is Computed and scored
- Based on the fitness value, probabilities will be assigned to each member of the population. These scaled values are called expectation values
- Selecting members of the population or parents based on their expectations
- Some of them which are chosen as elite will be passed to the next generation
- Generating new population from the parents
- New population is produced either by mutation or by crossover. In other words, by doing either some random changes to a single parent or by mixing the vector entries in a pair of parents
- Next generation is produced by replacing the children with the current population

- When the new population is not significantly different from the previous generation, the algorithm will be terminated.

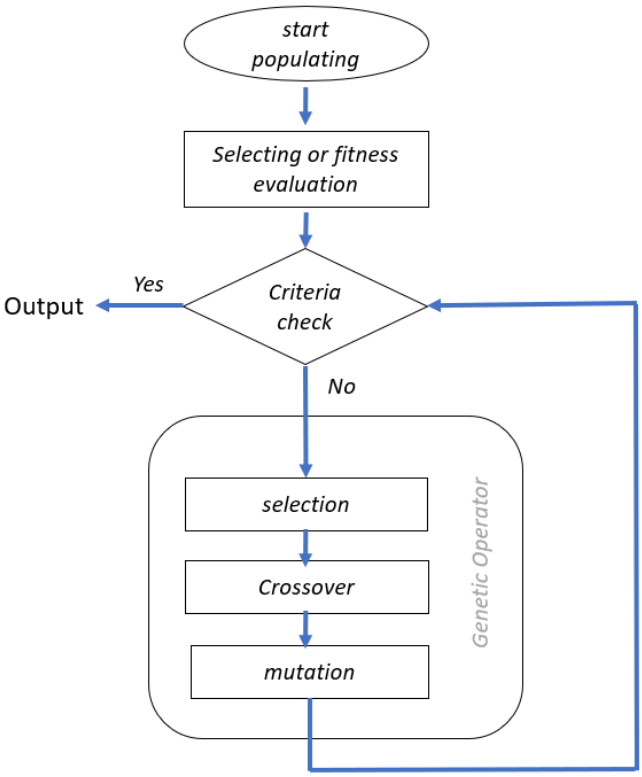

figure 6.1. Genetic algorithm flowchart

- Genetic algorithm production methods

1) Selection

In the selection process of parents for the next generation, the probabilities assigned to the parents by fitness function is used to choose the best parents. Fitness scaling process scales the values and selection process chooses parents for the next generation based on these scaled values. If a parent contributes its genes to more than one child, it can be chosen several times.

2) Crossover

In crossover reproduction technic, chromosomes of two parents are used to generate two child chromosomes. The new population or children will be generated by mixing pairs of parents in the current population. One of the methods which is widely used for the crossover is the one-point crossover. This method works with splitting the chromosomes of both parents into left

and right subchromosomes, where each pair of two left and two right subchromosomes have the same length. Chromosomes in each child is generated by getting left subchromosomes from one parent and right subchromosomes from other parent. The generation of the new population in the default crossover function happens as a random weighted average of parents. (Wright, 1991) the figure below shows the crossover function.

3) Mutation Children

Another common reproduction operator in the Genetic algorithm is the mutation. In the mutation process, a new population is generated by randomly changing the genes of the parents in the current population. When a chromosome is selected to generate a new child with the mutation process, some of its genes of the chromosome, are randomly selected and modified to create a new population. (Wright, 1991) Figure 6.2 illustrates different generation processes.

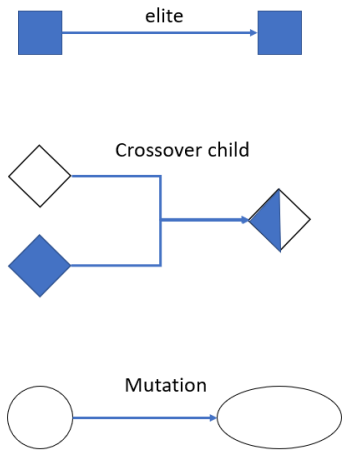

Figure 6.2. Genetic algorithm population generation types

Although the genetic algorithm which is used in this work to help the optimization process, is a stochastic search algorithm, in each step, it selects the alternatives which contribute more to the outcome. And It generates a new population by transferring selected genes. By considering this fact that a model is not perfect, the algorithm produces locations that tend to be more precise than to be accurate. It can be said that it is not an absolute stochastic search algorithm. It is only the first stage of the algorithm which generates the location coordinates completely random.

9.2 Optimizer's Curse I

```
N1 <- rnorm(10000000,0.5,1)

N2 <- rnorm(10000000,0,1)

N3 <- rnorm(10000000,-0.5,1)

P1 <-plot(density(N1), xlim=c(-3,3), ylim=c(0,0.6), xaxs="i", yaxs="i" , lwd="3" , col="green",

      main="Estimated Alternatives" , xlab ="" , ylab = "Density ")

polygon(density(N1), col = rgb(red = 0, green = 1, blue = 0, alpha = 0.1), border="green")

P2 <-lines(density(N2) , xlim=c(-3,3), ylim=c(0,0.6), xaxs="i", yaxs="i" , lwd="3" , col="blue")

polygon(density(N2), col = rgb(red = 0, green = 0, blue = 1, alpha = 0.1), border="blue")

P3 <-lines(density(N3) , xlim=c(-3,3), ylim=c(0,0.6), xaxs="i", yaxs="i" , lwd="3" , col="red" )

polygon(density(N3), col = rgb(red = 1, green = 0, blue = 0, alpha = 0.1), border="red")

X1 <-pmax(N1,N2,N3)

#  X1

Px <-lines(density(X1), lwd= "3", col="black")

polygon(density(X1), col = rgb(red = 0.5, green = 0.4, blue = 0.1, alpha = 0.1), border="red")

L1 <-abline(v=mean(N1), col="green", lwd="3")

mean(N1)

L1 <-abline(v=mean(N2), col="blue", lwd="3")

L1 <-abline(v=mean(N3), col="red", lwd="3")

L1 <-abline(v=mean(X1), col="black", lwd="3")

mean(X1)

sd(X1)

mean(X1)-mean(N1)
```

9.3 Optimizer’s curse II

```
N <- 100000000
AA <- rnorm(N, mean = 10, sd = 1)

AD <- sort(AA, decreasing = TRUE)
AM <- AD [c(1:1000000)]

plot(density(AA), col="green", lwd="3", xlim=c(-5,20), ylim=c(0,2), xaxs="i", yaxs="i", xlab =
"Distribution P & errors")

EE <- (rnorm(N, mean = 0, sd = 2))
FF <- lines(density(EE),col="red", lwd="3", xlim=c(-5,20), ylim=c(0,2), xaxs="i", yaxs="i")
AE <- AA+EE

lines(density(AM), col="blue", lwd="3")
lines(density(AE), col="blue", lwd="3")
max(AE)
min(AE)

DC <- sort(AE, decreasing = TRUE)
DM <- DC [c(1:1000000)]

lines(density(DM),col= "green", lwd="3", xlim=c(-5,20), ylim=c(0,2), xaxs="i", yaxs="i")

L1 <-abline(v=mean(AA), col="red", lwd="3")

L1 <-abline(v=mean(EE), col="red", lwd="3")

L1 <-abline(v=mean(AE), col="blue", lwd="3")

L1 <-abline(v=mean(DM), col="black", lwd="3")

L1 <-abline(v=mean(AM), col="black", lwd="3")
```

9.4 Sequential Gaussian Simulation

```
library(gstat)                    (Pebesma and Wesseling, 1998)
library(sp)
# data frames
library(plyr)                     (Wickham and Wickham, 2016)
library(fields)
nx = 45
ny = 45
xsize = 10.0
ysize = 10.0
xmin = 0
ymin = 0
xmax = xmin + nx * xsize
ymax = ymin + ny * ysize
x<-seq(xmin,xmax,by=xsize)
y<-seq(ymin,ymax,by=ysize)
colmap = topo.colors(100)

nscore <- function(x)                    (Makido et al., 2008)
  nscore <- qqnorm(x, plot.it = FALSE)$x  # normal score
  trn.table <- data.frame(x=sort(x),nscore=sort(nscore))
  return (list(nscore=nscore, trn.table=trn.table))}
addcoord <- function(nx,xmin,xsize,ny,ymin,ysize)
  ixy = 1
  for(iy in 1:nx) {
    for(ix in 1:ny) {
      coords[ixy,1] = xmin + (ix-1)*xsize
      coords[ixy,2] = ymin + (iy-1)*ysize
      ixy = ixy + 1 }}
  coords.df = data.frame(coords)
  colnames(coords.df) <- c("X","Y")
  coordinates(coords.df) =~X+Y
  return (coords.df)}
sim2darray <- function(spdataframe,nx,ny,ireal)
  model = matrix(nrow = nx,ncol = ny)
  ixy = 1
  for(iy in 1:ny) {
    for(ix in 1:nx) {
      model[ix,iy] = spdataframe@data[ixy,ireal]
      ixy = ixy + 1 } }
  return (model)}
sim2vector <- function(spdataframe,nx,ny,ireal)

  model = rep(0,nx*ny)
  ixy = 1
  for(iy in 1:ny) {
    for(ix in 1:nx) {
      model[ixy] = spdataframe@data[ixy,ireal]
```

```
    ixy = ixy + 1 }}
  return (model)}

mydata1=read.csv('Welllocation4SGS_inj.csv')
j=1
cumdata <- matrix(0,nrow = 1500,ncol = 4)
cumdata <- as.data.frame(cumdata)
for (i in 1:1){
  sample <- matrix(data.matrix(mydata1[i,2:9]),nrow = 4,ncol = 2)
  locdata <- data.frame(rbind(matrix(c(23,23),nrow = 1,ncol = 2),sample))
  locgrid <- matrix(0,nrow=5,ncol = 2)
  for (k in 1:5) {
    locgrid[k,1] <- (locdata[k,1]-1)*xsize + 5
    locgrid[k,2] <- (locdata[k,2]-1)*ysize + 5
  }
  m <- 500
  s <- 150
  location <- log(m^2 / sqrt(s^2 + m^2))
  shape <- sqrt(log(1 + (s^2 / m^2)))
  Perm <- matrix(rlnorm(5,location,shape),nrow = 5,ncol = 1)
  logperm <- log10(Perm)
  logperm <- matrix(logperm,nrow = 5,ncol = 1)
  mydata2 <- data.frame(cbind(locgrid,Perm,logperm))
  z <- i*5
  cumdata[j:z,1:4] <- mydata2
  colnames(mydata2) <- c('X','Y','Perm','logperm')
  j=j+5


  coordinates(mydata2) = ~X+Y
  npor.trn = nscore(mydata2$logperm)
  mydata2[["NPermeability"]]<-npor.trn$nscore

  cuts = c(2.4,2.45,2.5,2.65,2.7,2.8,2.9)
  cuts.var = c(0.05,.1,.15,.20,.25,.3,.35,.4,.45,.5,.55,.6,.65,.7,.75,.8,.85,.9,.95)

  spplot(mydata2, "logperm", do.log = TRUE,
        key.space=list(x=1.05,y=0.97,corner=c(0,1)),cuts = cuts,
        scales=list(draw=T),xlab = "X (m)", ylab = "Y (m)",main ="Permeability (Log(K)),
        in md")
  coords <- addcoord(nx,xmin,xsize,ny,ymin,ysize)

  sill = var(mydata2$logperm)
  min = min(mydata2$logperm)
  max = max(mydata2$logperm)
  zlim = c(min,max)
  vm.nug1 <- vgm(psill =0.5*sill, "Sph", 200, anis = c(000, 1.0),nugget=0.5*sill)
  condsim.nug1 = krige(logperm~1, mydata2, coords, model = vm.nug1, nmax = 100, nsim =
10)
```

```
 par(mfrow=c(2,2))
 real1 <- sim2darray(condsim.nug1,nx,ny,1)
 image.plot(10^real1,x=x,y=y,xlab="X(m)",ylab="Y(m)",
        zlim = c(min(10^real1),max(10^real1)),
        col=colmap,legend.shrink = 0.6);
 mtext(line=1, side=3, "Realization #1", outer=F);box(which="plot")

 real2 <- sim2darray(condsim.nug1,nx,ny,2)
 image.plot(10^real2,x=x,y=y,xlab="X(m)",ylab="Y(m)",
        zlim =c(min(10^real2),max(10^real2)) ,
        col=colmap,legend.shrink = 0.6);
 mtext(line=1, side=3, "Realization #2", outer=F);box(which="plot")
 real3 <- sim2darray(condsim.nug1,nx,ny,3)
 image.plot(10^real3,x=x,y=y,xlab="X(m)",ylab="Y(m)",zlim =
c(min(10^real3),max(10^real3)),
        col=colmap,legend.shrink = 0.6);
 mtext(line=1, side=3, "Realization #3", outer=F);box(which="plot")

 real4 <- sim2darray(condsim.nug1,nx,ny,4)
 image.plot(10^real4,x=x,y=y,xlab="X(m)",ylab="Y(m)",
        zlim = c(min(10^real4),max(10^real4)),
        col=colmap,legend.shrink = 0.6);
 mtext(line=1, side=3, "Realization #4", outer=F);box(which="plot")}
```

Adopted from (Kor, 2019)

9.5 Optimiziation algorithm

```
install.packages('GA')
install.packages('doParallel')
install.packages('parallel')
library(GA)
library(parallel)
library('doParallel')

setwd("C:/ROOPT_Ide/Rcode")
library(readr)
X2D_1Inj_4Prod_R1 <- read_delim("2D_1Inj_4Prod_R.DATA",
                        "\t", escape_double = FALSE, na = "null",
                        trim_ws = TRUE)


Injec_up1 <- read_delim("Injec_up1.csv",
                    ";", escape_double = FALSE, col_names = FALSE,
                    trim_ws = TRUE)

Eclipse <- function(X1,Y1,X2,Y2,X3,Y3,X4,Y4,A1,gam1) {
  result <- c(rep(0,10))
  for (i in 1:10) {
    result[i] <- Eclipse1(X1,Y1,X2,Y2,X3,Y3,X4,Y4,A1,gam1,i)
  }
  mean_result <- mean(result)
  # mean_result <- mean_result*0.6
  return(mean_result)
}

run <- 'C:/ecl/macros/eclrun.exe eclipse "C:/ROOPT_Ide/En#/Ensemble.DATA"'

Eclipse1 <- function(X1,Y1,X2,Y2,X3,Y3,X4,Y4,A1,gam1,i){
  wd <- 'C:/ROOPT_Ide/En#'
  run1 <- gsub('#',i,run)
  newwd <- gsub('#',i,wd)
  setwd(newwd)
  X1 <- round(X1)
  X2 <- round(X2)
  X3 <- round(X3)
  X4 <- round(X4)
  Y1 <- round(Y1)
  Y2 <- round(Y2)
  Y3 <- round(Y3)
  Y4 <- round(Y4)
  slope <- c(-0.0005,-
0.00025,0,0.00025,0.00050,0.00075,0.00100,0.00125,0.00150,0.00175,0.00200)
  Avalue <- c(100,150,200,250,300)

  A_1 <- Avalue[findInterval(A1,Avalue)]
```

```
 Gam1 <- slope[findInterval(gam1,slope)]

 data <- X2D_4Inj_1Prod_R1
 g1 <- data[131,1]
 g2 <- data[132,1]
 g3 <- data[133,1]
 g4 <- data[134,1]
 g5 <- data[138,1]
 g6 <- data[139,1]
 g7 <- data[140,1]
 g8 <- data[141,1]
 G1 <- gsub('103',X1, g1)
 G2 <- gsub('104',Y1, G1)
 G3 <- gsub('105',X2, g2)
 G4 <- gsub('106',Y2, G3)
 G5 <- gsub('107',X3, g3)
 G6 <- gsub('108',Y3, G5)
 G7 <- gsub('109',X4, g4)
 G8 <- gsub('110',Y4, G7)
 G9 <- gsub('113',X1, g5)
 G10 <- gsub('114',Y1, G9)
 G11 <- gsub('115',X2, g6)
 G12 <- gsub('116',Y2, G11)
 G13 <- gsub('117',X3, g7)
 G14 <- gsub('118',Y3, G13)
 G15 <- gsub('119',X4, g8)
 G16 <- gsub('120',Y4, G15)
 data[131,1] <- G2
 data[132,1] <- G4
 data[133,1] <- G6
 data[134,1] <- G8
 data[138,1] <- G10
 data[139,1] <- G12
 data[140,1] <- G14
 data[141,1] <- G16

 write.table(data, file ="Ensemble.DATA", sep = "\t",quote = F,row.names = F,col.names = F)
 int <- seq(from=0,to=1470,by=30)
 InjectionRatesex <- Injec_up1
 vector <-
c(c(3,4,5),c(11,12,13),c(19,20,21),c(27,28,29),c(35,36,37),c(43,44,45),c(51,52,53),c(59,60,61),c(67,
68,69),

c(75,76,77),c(83,84,85),c(91,92,93),c(99,100,101),c(107,108,109),c(115,116,117),c(123,124,125),c
(131,132,133),

c(139,140,141),c(147,148,149),c(155,156,157),c(163,164,165),c(171,172,173),c(179,180,181),c(18
7,188,189),

c(195,196,197),c(203,204,205),c(211,212,213),c(219,220,221),c(227,228,229),c(235,236,237),c(24
3,244,245),
```

```
c(251,252,253),c(259,260,261),c(267,268,269),c(275,276,277),c(283,284,285),c(291,292,293),c(29
9,300,301),

c(307,308,309),c(315,316,317),c(323,324,325),c(331,332,333),c(339,340,341),c(347,348,349),c(35
5,356,357),
          c(363,364,365),c(371,372,373),c(379,380,381),c(387,388,389),c(395,396,397))

 InjectionRatesex <- InjectionRatesex[-vector,]
 InjectionRatesex[seq(from=2,to=250,by=5),5] <- A_1*exp(-Gam1*int)
 write.table(InjectionRatesex, 'InjectionRates.inc',quote = F,row.names = F,col.names = F, na = '')
 a1 <- A_1*exp(-Gam1*int)
 AA <- a1
 shell(run1)
 return(NPVcalculator(AA,newwd))
}


NPVcalculator <- function(AA,newwd){
 setwd(newwd)
 ens <- "ENSEMBLE.RSM"
 NPV <- read.delim(ens,  header=FALSE, comment.char="#")
 NPVcalc <- NPV[, c(2,5,6)]
 colnames(NPVcalc) <- c('Days', 'Oilrate', 'waterrate')
 NPVcalc <- na.omit(NPVcalc)
 NPVcalc <- NPVcalc[-seq(9,400,by = 2), ]
 rownames(NPVcalc) <- 1:nrow(NPVcalc)
 days <- as.numeric(levels(NPVcalc$Days))[as.integer(NPVcalc$Days)]
 days <- na.omit(days)
 l1 <- min(which(days==c(1), arr.ind = TRUE))
 l2 <- min(which(days==c(1500), arr.ind = TRUE))
 days <- days[c(l1:l2)]
 oil <- as.numeric(levels(NPVcalc$Oilrate))[as.integer(NPVcalc$Oilrate)]
 oil <- na.omit(oil)
 oil <- oil[c(l1:l2)]
 water <- as.numeric(levels(NPVcalc$waterrate))[as.integer(NPVcalc$waterrate)]
 water <- na.omit(water)
 water <- water[c(l1:l2)]
 FWCT <- numeric(length = length(water))
 FWCT <- water/(water+oil)
 t <- seq(0,days[length(which(FWCT<0.85, arr.ind = TRUE))],by = 30)
 t[1] <- 1
 cashflow <- numeric(length(t)-1)
 discashflow <- numeric(length(t)-1)
 b <- 0.08
 for (j in 1:(length(t)-1)) {
  m1 <- t[j]
  z1 <- which(days==c(m1), arr.ind = TRUE)
  m2 <- t[j+1]
  z2 <- which(days==c(m2), arr.ind = TRUE)
```

```
   cashflow[j] <- (mean(oil[z1],oil[z2])*6.29*75-mean(water[z1],water[z2])*6.29*19-
AA[j]*5*6.29)*30
   discashflow[j] <- cashflow[j]/((1+b)^(t[j+1]/360))
  }
  return(sum(discashflow))}


#sugges <- c(1,45,45,1,1,1,45,45,150,0)
GA <- ga(type = "real-valued", fitness = function(x) +
Eclipse(x[1],x[2],x[3],x[4],x[5],x[6],x[7],x[8],x[9]
                                              ,x[10])
       ,lower = c(rep(1,8),rep(100,1),rep(-0.0005,1)),
       upper = c(rep(45,8),rep(300,1),rep(0.002,1)),popSize = 10, maxiter = 100,run = 10,parallel =
T,) #suggestions = sugges
GA@solution
plot(GA)
Eclipse <- function(X1,Y1,X2,Y2,X3,Y3,X4,Y4,A1,gam1) {
  result <- c(rep(0,100))
  for (i in 1:100) {
    result[i] <- Eclipse1(X1,Y1,X2,Y2,X3,Y3,X4,Y4,A1,gam1,i)
  }
  return(result)
}



Eclipse_NPV <- Eclipse(1,45,45,1,1,1,45,45,150,0)
Eclipse_NPV

######### Histogram plot
hist(Eclipse_NPV, breaks = 10, main = "Npv Distribution Over
    Geological Realizations", probability = T, xlab="NPV ($$ MM)"
    , col="blue", border="red")
Eclipse_NPV
Adopted from (Kor, 2019)
```