Spring 2015

Bayesian global optimization approach to the oil well placement problem with quantified uncertainties

Zengyi Dou
Purdue University

Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Zengyi Dou

Entitled

BAYESIAN GLOBAL OPTIMIZATION APPROACH TO THE OIL WELL PLACEMENT PROBLEM WITH QUANTIFIED UNCERTAINTIES

For the degree of Master of Science in Mechanical Engineering

Is approved by the final examining committee:

Ilias Bilionis
Chair

Jitesh Panchal

Marisol Koslowski

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Ilias Bilionis

Approved by: Ganesh Subbarayan 4/28/2015

Head of the Departmental Graduate Program Date

BAYESIAN GLOBAL OPTIMIZATION APPROACH TO THE OIL WELL PLACEMENT PROBLEM WITH QUANTIFIED UNCERTAINTIES

A Thesis

Submitted to the Faculty

of

Purdue University

by

Zengyi Dou

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Mechanical Engineering

May 2015

Purdue University

West Lafayette, Indiana

For my parents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my academic adviser Professor Ilias Bilionis for his persistent support and encouragement. His advices have been directing me throughout my study. Without his guidance this thesis would not have been completed. Also, I am extremely grateful to my graduate committee, Prof. Jitesh Panchal and Marisol Koslowski. The computational aspect of this work relied on resources gracefully offered by the School of Mechanical Engineering, Purdue University. Last but not least, I would like to thank my parents and my friends who always stand by my side and offer me help throughout my life.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

NPV	Net Present Value
BGO	Bayesian Global Optimization
MC	Monte Carlo
GPR	Gaussian Process Regression
SE	Squared Exponential
FVM	Finite Volume Method
FDM	Finite Difference Method
HGA	Hybrid Genetic Algorithm
EI	Expected Improvement
PDE	Partial Differential Equation
GP	Gaussian Process
MLE	Maximum Likelihood Estimation
MAP	Maximum a Posterior Probability

ABSTRACT

Dou Zengyi. M.S.M.E., Purdue University, May 2015. Bayesian Global Optimization Approach to the Oil Well Placement Problem with Quantified Uncertainties. Major Professor: Ilias Bilionis, School of Mechanical Engineering.

The oil well placement problem is vital part of secondary oil production. Since the calculation of the net present value (NPV) of an investment depends on the solution of expensive partial differential equations that require tremendous computational resources, traditional methods are doomed to fail. The problem becomes exceedingly more difficult when we take into account the uncertainties in the oil price as well as in the ground permeability. In this study, we formulate the oil well placement problem as a global optimization problem that depends on the output of a finite volume solver for the two-phase immiscible flow (water-oil). Then, we employ the machinery of Bayesian global optimization (BGO) to solve it using a limited simulation budget. BGO uses Gaussian process regression (GPR) to represent our state of knowledge about the objective as captured by a finite number of simulations and adaptively selects novel simulations via the expected improvement (EI) criterion. Finally, we develop an extension of the EI criterion to the case of noisy objectives enabling us to solve the oil well placement problem while taking into account uncertainties in the oil price and the ground permeability. We demonstrate numerically the efficacy of the proposed methods and find valuable computational savings.

CHAPTER 1. INTRODUCTION

1.1 Motivation

During secondary oil production water (potentially enhanced with chemicals) or gas is injected to the reservoir through an *injection* well. The injected fluid pushes the oil out of the *production* well. The *oil well placement problem* involves the specification of the number and location of the injection and production wells, the operating pressures, the production schedule, etc., that maximize the net present value (NPV) of the investment. This problem is of extreme importance for the oil industry and an active area of research.

Several sources of uncertainty influence the NPV. The most important of these sources are the time evolution of the oil price (aleatoric uncertainty) and our uncertainty about the underground geophysical parameters (epistemic uncertainty). For convenience, we will be referring to these uncertain parameters as *stochastic inputs*. Taking these uncertainties into account, we see that the oil well placement problem constitutes a *design optimization problem under uncertainty*. In this thesis, we consider the risk-neutral approach of maximizing the expected NPV of the investment. Our developments are easily extendable to the risk-averse case.

Given a set of design parameters as well as a realization of the stochastic inputs, the computation of the NPV involves the solution of a coupled system of partial differential equations (PDEs) describing the two-phase flow through the oil reservoir. In real scenario

involving complex large-scale reservoirs and/or multi-scale/physics modeling, simulation times are exceedingly long making it impossible to solve the problem using standard global optimization methods. Thus, there is an urgent need for global optimization techniques that can work with a limited data acquisition budget.

There are various other difficulties associated with oil well placement problem. First, note that the any optimization algorithm we employ needs to be gradient-free. Even though the calculation of the gradient of the NPV with respect to the well locations is theoretically possible using the method of adjoints, to the best of our knowledge, it has not been done yet. Second, the high and potentially discontinuous variability of the geophysical properties induces several local maxima in the NPV and necessitates a global approach. Third, the computation of the expectation of NPV requires the computation of high-dimensional integral (integration over all stochastic inputs). In a typical problem the number of stochastic dimensions can well be in the order of hundreds of thousands. This fact necessitates the use of a sampling or Monte Carlo (MC) approach. Fourth, the characterization of the uncertainty of the geophysical properties requires the solution of an inversion problem that fuses information from seismic surveys, data collected during the primary production phase, analysis of ground specimens, and expert knowledge. Fifth, the specification of the oil price model is a hard open problem in time series analysis.

The latter two difficulties constitute part of the formulation of the problem and are not examined in detail. The novel contributions of this thesis focus on the solution of stochastic optimization problems under uncertainty that face the following issues: expensive data acquisition, lack of gradient information, and high-dimensional

uncertainties. We tackle these issues by employing and extending the method of Bayesian global optimization (BGO).

BGO uses Gaussian process regression (GPR) to quantify our state of knowledge about the objective function as captured by a finite number of evaluations. GPR is a Bayesian meta-model. In contrast to classic meta-model approaches, it defines a probability measure on the space of meta-models. This probability measure corresponds to our uncertainty about the objective function as induced by the limited number of observations we have at our disposal. BGO exploits this uncertainty to compute the expected improvement (EI) of a new evaluation. In other words, it assesses the merit of performing a new simulation by computing the EI it would bring to the current best solution of the problem. By iteratively performing the simulations of maximum EI, BGO gradually converges to the global optimum. The novel contribution of this thesis is the extension of BGO to treat design optimization problems under uncertainty.

1.2 Literature Review

In the area of oil production, the fundamental research problems rise from the multi-scale variability of the geophysical parameters such as the porosity and the permeability of the ground.

Typically, the computational mesh is used to represent the reservoir in the sense that the geophysical parameters are assumed to be constant within its computational cell. Given the fact that this approach does not scale well as the reservoir becomes larger, researchers focused on up-scaling maps to fill the gap between geological and simulation grid size.[5, 20, 21, 41] Most recently, multi-scale techniques were proposed to solve the

large-scale problem using the fine scale geophysical information. This method is most suitable for PDEs whose solutions display a multiple scale structure.[2, 3, 18, 28]. Since the focus of this study is not on the modeling the oil reservoir, we will develop a simple two-phase immiscible flow solver using the standard finite volume method (FVM). This solver will constitute our oil reservoir simulator.

In the past, numerous of algorithms have been proposed for solution of optimization and inverse problems in oil reservoir modeling.[36, 38] The selection of the optimal *oil well scheduling and placement problem* using optimization algorithm was first studied ten years ago.[6] Since then, wide arrays of methods have been proposed. For example, Centilmen A et al. solves the problem by building a meta-model represented by artificial neural networks[15, 25], Elamvazuthi et al. proposes a hybrid optimization technique using fuzzy inference[23], Bittencourt develops a hybrid genetic algorithm (HGA) optimization method that combines the polytope algorithm with neural network.[13] Other researchers focus on more complicated aspects related to factors such as the well type, the number of wells, and orientations or production characteristic parameters such as porosity, permeability and so on. Although the research interest varies hugely, from computational algorithm stand of point, they can be divided into two main catalogs: the stochastic and the other is heuristic approach.[4] However, the existing literature has barely touched the data acquisition problem as well as the high-dimensional aleatoric and epistemic uncertainties.

In this study, we proposed a new approach for design optimization under uncertainty with a limited data-budget and applied it to the well-placement problem in oil-reservoir modeling. We developed an oil reservoir simulator based on the FVM that solves the

coupled PDEs describing the two-phase immiscible flow occurring during secondary oil production. Based on limited observations, we construct a GPR to quantify our state of knowledge about the NPV of the investment, and then use the EI data acquisition criterion to actively select the most informative simulations. The remainder of this study is organized as follows. Section 2 introduces a general background of physics and mechanics of oil reservoir modeling, as well the numerical method we use to solve the coupled PDEs. Section 3 proposes the target economical function, NPV, and discusses in detail each quantity that affects it. Section 4 describes the GPR for construction of meta-models. Section 5 discusses data-acquisition criteria for design optimization. Section 6 shows the results for several different testing scenarios and conclusions are given.

CHAPTER 2. PHYSICS AND MATHEMATICS MODEL FOR OIL RESERVOIR

In this section, we introduce all critical physical parameters to mathematical model describing an oil reservoir during secondary oil production. The physics involve a two-phase immiscible flow which can be described via a coupled PDE system. We conclude by discussing the details of the numerical solution of the coupled PDEs.

2.1 Porosity

In geoscience, the holes in sandstones are called *porosity*. Porosity is a measure of space not occupied by the solid rock, also defined as volume fraction of total rock occupied by other phase such as water, oil or gas.[1] Usually rock porosity is denoted by ϕ, and it is a number between 0 and 1. Strictly speaking, oil and gas are not stored as a continuous empty space. Instead, they are in the space between grains and sandstones. Rock is treated as a compressible material like a sponge although in reality it is hard and appears solid. Mathematically, the *compressibility* of rock is defined as:

$$c_r = \frac{1}{\phi}\frac{d\phi}{dp}, \tag{2.1}$$

where p refers to overall pressure in the reservoir. In order to simplify the model, in this study we assume there is no compressibility, which means porosity only depends on spatial coordinate.

2.2 Permeability

Permeability is a tensor quantity that measures the ease with which a fluid (in this study, either water or oil) can move through a porous rock. Permeability is denoted as $\mathbf{K}$, and typically has unit of darcy (D) or millidarcy (mD). In petroleum production, millidarcy is used more often, for 1 darcy is a relatively huge number. A more accurate definition is that 1 darcy is equal to 1 cp fluid flow through a homogeneous material at 1cm/s speed under 1 atm/cm pressure gradient.[1]

Absolute permeability is the measurement of soil or rock's ability to transmit a single phase or fluid. If there is more than one phase, the effective permeability of a specific phase depends also on the other phases existing at the same location. This is, of course, a coarse model since the two phases do not mix with each other from a microscropic perspective. This effective permeability, a nonlinear function of absolute *permeability* $\mathbf{K}$ and *saturation* s, will be called *relative* permeability. Relative permeability is the ratio of effective permeability of a particular fluid at a particular saturation to the absolute permeability of that fluid at total saturation:

$$\mathbf{K}_i = \mathbf{K} k_{ri}(s_{i+1}), \tag{2.2}$$

where i stands for the phase or liquid we are interested and $i+1$ stand for the other phase. Generally speaking, relative permeability is determined by pore-size, the viscosity of the two phases, and the forces between the fluids.[22] These factors are easily taken into account. The challenging part in secondary oil production is the significant, non-trivial temperature dependence of the permeability.[34]

The measurement of permeability is costly and difficult.[27] Even using the latest techniques and instruments, there is still a lot of residual uncertainty that makes the result

of the numerical simulator untrustworthy. Thus, when making the well developing plan, it is better to consider the uncertainty of the permeability field.

Permeability data can be collected in three ways. First, permeability information can be extracted from seismic surveys.[40] Second, permeability can be derived from through Bayesian model inversion using data collected in primary production stage.[10, 12] Third, small scale permeability measurements can also be gathered through drilling and then used to construct an upscaled version of the permeability field.[44] No matter which method is used to get permeability data, there will always be some residual uncertainty. Since permeability uncertainty directly affects production rate, it should not be ignored.

In this study, we construct a synthetic, albeit realistic, stochastic permeability model based on the data included in the SPE Comparative Solution Project.[21] Namely, we model the logarithm of the permeability as a Gaussian process.[24] Specifically:

$$\begin{aligned} \ln K(\mathbf{x}_s) &= \ln K_0(\mathbf{x}_s) + g(\mathbf{x}_s), \\ \text{where} \quad g(\mathbf{x}_s) &= \mathrm{GP}(g(\cdot) \mid 0, k(\cdot,\cdot)), \end{aligned} \tag{2.3}$$

where $K(\mathbf{x}_s)$ is the permeability at the spatial location $\mathbf{x}_s = (x_{s1}, x_{s2})$, and $g(\mathbf{x}_s)$ is a Gaussian distribution with an exponential covariance function:

$$k(\mathbf{x}, \mathbf{x}') = v_s^{\,2} \exp\left\{ -\sum_{i=1}^{d} \frac{\left| x_i - x_i' \right|}{l_{si}} \right\}. \tag{2.4}$$

To make matters worse, permeability may change during the production process, as well as when the environment temperature and pressure change. Permeability also depends on the direction and fluid type, i.e. permeability is not isotropic and constant property for different fluids. So it could have different values when different fluids

passing through the same rock sample. In this study, however, we assume the rock is an isotropic material and permeability is a constant real number.

2.3 Fluid Properties

In secondary oil production, the empty space in the rock is filled with either aqueous phase or oleic phase. The concept of *saturation* is introduced to describe the volume fraction occupied by different phases. For multiple phases, following relationship always holds:

$$\sum_{allphases} s_i = 1, \tag{2.5}$$

In this study, we consider a two-phase flow. One phase is the aqueous phase which mainly contains water and the other one is oleic phase mainly contains hydrocarbon. In two phase flow model, there is no need to distinguish whether the oleic phase consists of gas or oil. The conservation of mass gives rise to the following relationship:

$$\sum_{i=1}^{N} c_{io} = \sum_{i=1}^{N} c_{iw} = 1, \tag{2.6}$$

where the mass fraction of i in phase j is written as c_{ij}. For the viscosity μ_i and density ρ_i, we have:

$$\mu_i = \mu_i(p_i, c_{1i}, ...c_{Ni}), \tag{2.7}$$

$$\rho_i = \rho_i(p_i, c_{1i}, ...c_{Ni}). \tag{2.8}$$

Due to interfacial tension, different phases have different phase pressure p_i. The *capillary pressure* is defined to be:

$$p_{cij} = p_i - p_j, \tag{2.9}$$

where $i, j = o, w$. Although the capillary pressure depends on many properties, we assume that it is the function of saturation only.

2.4 Mathematical Model of an Oil Reservoir

Fluid motion can be described by the conservation of mass, momentum, and energy and thermodynamics. Oil production is not an exception. The momentum equation is included in the Darcy's law, which is an empirical equation that describes the relationship between fluid velocity and the pressure gradient.

In this study, we use the immiscible two-phase oil model, based on the assumption that the capillary effect is neglected. For more precise model considering the capillary effect more information can be found in [19]. As mentioned before, we consider an aqueous and oleic phase, assuming only water is in the aqueous phase and that the oleic phase contains hydrocarbon mixture. The following PDEs are used to describe the two-phase flow through the oil reservoir. According to Darcy's law the primary driven force pushes water and oil through porous rock is gravity and the pressure gradient. Each phase has each own continuity equation:

$$\frac{\partial(\phi \rho_i s_i)}{\partial t} + \nabla \cdot (\rho_i v_i) = q_i, \tag{2.10}$$

where i stands for either aqueous phase or the oleic phase. Expanding the Equation (2.10), we obtain a more detailed form of continuity equation:

$$\frac{\partial \phi}{\partial t} s_i + \phi \frac{\partial s_i}{\partial t} + \phi \frac{s_i}{\rho_i} \frac{\partial \rho_i}{\partial t} + \nabla \cdot v_i + \frac{v_i \cdot \nabla \rho_i}{\rho_i} = \frac{q_i}{\rho_i}. \tag{2.11}$$

Before plugging the pressure equation into the continuity equation, let us introduce the concept of *phase mobility* defined as $\lambda_i = k_{ri} / \mu_i$. According to Equation (2.5),

saturation of oil and water has the relationship: $s_w + s_o = 1$, then sum continuity equation of aqueous phase or oleic phase and define $q = q_w / \rho_w + q_o / \rho_o$, we have:

$$\nabla \cdot (v_w + v_o) + \frac{\partial \phi}{\partial t} + \phi \frac{s_w}{\rho_w} \frac{\partial \rho_w}{\partial t} + \phi \frac{s_o}{\rho_o} \frac{\partial \rho_o}{\partial t} + \frac{v_w \cdot \nabla \rho_w}{\rho_w} + \frac{v_o \cdot \nabla \rho_o}{\rho_o} = q. \qquad (2.12)$$

The oil wells either produce oil and water mixture or inject water. The source term q on the right hand side in the Equation (2.12) is non-positive at the location of a production well and non-negative at the location of an injection well. In real production process, both injection and production wells are subject to control and optimization.

Darcy discovered an empirical equation that connects the volumetric flow density v to the pressure and the gravity force. The multiphase version of Darcy's Law is stated as:

$$v_i = -\mathbf{K} \frac{k_{ri}}{\mu_i} (\nabla p_i + \rho_i g \nabla z). \qquad (2.13)$$

As mentioned before, since the purpose of this study is not to set up an extremely precise oil simulator, assumption that rock and two fluid phases are incompressible is reasonably proposed to simplify the problem. By inserting Darcy's law into Equation (2.12), it reduced to:

$$\begin{gathered} \nabla \cdot v = q, \\ v = -[\mathbf{K}\lambda_w (\nabla p_w + \rho_w g \nabla z) + \mathbf{K}\lambda_o (\nabla p_o + \rho_o g \nabla z)]. \end{gathered} \qquad (2.14)$$

In Equation (2.14), there are two unknown phase pressure p_o and p_w. The normal solution is introducing the capillary pressure $p_{cow} = p_o - p_w$, which is regarded as a function of water saturation s_w. However, this makes the saturation and pressure equations strongly coupled. So another approach, using *global pressure* p is introduced. Also, we assume that the capillary pressure p_{cow} is a monotone function of the water

saturation s_w .Global pressure is defined as: $p = p_o - p_c$, where the complementary pressure p_c is defined by:

$$p_c(s_w) = \int_1^{s_w} f_w(\xi) \frac{\partial p_{cow}}{\partial s_w}(\xi)\, \mathrm{d}\xi, \tag{2.15}$$

where $f_w = \lambda_w / (\lambda_w + \lambda_o)$ is factional-flow function that measures the water fraction of the total flow. Finally, introducing the total mobility $\lambda = \lambda_w + \lambda_o$, we obtain the following elliptic equation for the global pressure p :

$$\nabla \cdot [\mathbf{K}\ \nabla p \quad \mathbf{K}(\ _w\ _w + \ _o\ _o) g z] = q. \tag{2.16}$$

The pressure equation gives us the first primary unknown p . The secondary unknown s_w is introduced in the saturation equation:

$$\phi \frac{\partial s_w}{\partial t} + \nabla \cdot \left(f_w(s_w)[v + d(s_w, \nabla s_w) + g(s_w)] \right) = \frac{q_w}{\rho_w}. \tag{2.17}$$

2.5 Numerical Implementation: Finite Volume Method

The most basic method for the solution of PDEs is the finite difference method (FDM). This method is based on a finite difference approximation of the partial derivatives of the solution occurring in the PDEs. In this study, we use the finite volume method (FVM). FVM is derived from the conservation of physical quantities over cell volumes and is, therefore, consistent with them. In FVM, the unknown functions or properties are represented in terms of average values over a set of finite-volumes, over which the integrated PDE model is required to hold in an averaged sense. The hidden assumption is that the flux that leaves one small volume is equal to the flux that enters another adjacent volume, so that continuity equation is valid automatically.

2.6 Solution Strategies for Coupled System

Based on the derivation in previous section, the fractional-flow model for immiscible two-phase flow can be described using pressure Equation (2.16) and saturation Equation (2.17). These two equations are nonlinearly coupled through the saturation-dependent mobilities λ_i. In addition, they are also coupled through other parameters that depend on pressure or saturation. The classic way to solving coupled PDEs is to make an implicit discretization for each equation and then solve for the two unknowns together. However, due to numbers of iterations required, this way is too expensive especially for large nonlinear systems of equations. In this study, we apply another method called *sequential splitting method*. This method is designed as follows: First, before the global pressure and total velocity are solved for, the data from the previous time step is used to compute the saturation-dependent coefficient in Equation (2.16). Then, total velocity v is kept as a constant parameter in Equation (2.17), while saturation is advanced. Next, saturation-dependent coefficients in Equation (2.16) are updated, and the pressure equation is solved again and so on. Finally, the numerical scheme is developed without considering the coupling problem between the two equations.

Although this splitting method will introduce an error by decoupling the equations, its efficiency makes up for it.

CHAPTER 3. OBJECTIVE FUNCTION-NET PRESENT VALUE

3.1 Net Present Value

There are three major factors that we need to consider when computing the NPV of an oil reservoir investment. Those are: the profits we make by selling the oil we extract, the operation costs of running the production plan, and the water disposal costs.

The most important factor that affects the NPV is amount of crude oil that comes out of the production well as a function of time. Due to the complicated economic environment, crude oil price fluctuates constantly and, thus, needs to be modeled as a stochastic process. To further complicate matters, the producer may decide to sell the oil right away or store it in anticipation of better prices. In this study, we will assume that the producer sells the oil at the current market price. In our numerical examples, we will investigate two different cases of price models: a time independent constant price model and a log-normal random walk price model.

The second NPV factor is the operation cost. This includes the cost of pumping water into the injection well, as well as the cost of the equipment and labor. Since the cost of equipment and labor remains constant, if we do not optimize over the number of wells, we ignore it in our analysis. Furthermore, we assume that the cost of water injection is proportional to the rate of injection and constant over time.

The third and final cost we consider is the cost associated with the disposal of waste. Water coming out of the production well is highly contaminated and has to be disposed according to the local environmental protection laws. Here, we also assume that the cost of water disposal is proportional to the rate with which waste water is produced and does not vary with time.

Putting all this considerations together and using a discount factor r, we can write the NPV of the investment over a time period T as:

$$f_T(\mathbf{x}) = \int_0^T \{ \sum_{prod.wells} [(c_o(t) q_o^-(\mathbf{x},t) - c_{w,disp} q_w^-(\mathbf{x},t))] - \sum_{inj.wells} c_{w,inj} q_w^+(\mathbf{x},t)\} / (1+r)^{t/365} dt \quad (3.1)$$

The implicit assumption here is that all the other operation expenses (lifting, separation, filtering, pumping and reinjection and so on), are independent of well location and can, thus, be excluded from our analysis. In Equation (3.1), c stands for unit cost of the quantity . In our numerical simulations, we take the cost for water injection to be 0.03$/bbl and the cost for water lifting is 0.04$/bbl. As mentioned already, we explore two cases of oil price: constant and log-normal random walk. The latter is discussed in Sec. 3.2. In Equation (3.1) q stands for flow rate of the quantity and it is a function of the well locations as well as of time. The flow rate functions are implicitly defined through the solution of the two-phase flow PDEs which we introduced in Chapter2.

The design problem we wish to solve is:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} f_T(\mathbf{x}), \quad (3.2)$$

subject to the constraint that the well locations **x** must lie within the reservoir.

3.2 Log-Normal Random Walk with Drift: Oil Price Model

The goal of this section is to construct a simple stochastic process that models the time evolution of the oil price. The model is a simple re-application of the common random walk models used to model the fluctuation of stock prices. We make no attempt to model sharp variations such as market crashes. The focus is on modeling the overall trend of the oil price as well as its volatility during "normal" periods.

Since the oil price is restricted to be a positive number, we define the model in log-space. It is:

$$\log c_o(t+1) = \log c_o(t) + \quad t + \quad \times z_t, \tag{3.3}$$

where the $c_o(t)$ stands for the oil price of day t and $c_o(t+1)$ is the oil price of following day, σ is the volatility of the oil price, z_t is standard normal distribution and μ is a small drift representing a generic trend in the price ("up" if it is positive, "down" if it is negative). This is model is commonly used in financial engineering as a the price mode of a commodity.[35] Figure 3.1 shows five time series sampled from the model. All start from an initial price $c_o(0) = 560.8$, with the same drift $\mu = 10^{-8}$ and volatility $= 10^{3}$. In our final numerical examples, we will seek to optimize the expectation of the NPV over this uncertainty.

The oil well placement problem becomes now a design optimization problem under uncertainty. Assuming a risk-neutral production company, it can be formulated as:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \mathbb{E}_{\{c_o(t)\}_{t=1}^{T}} \left[f_T(\mathbf{x}) \right], \tag{3.4}$$

where the expectation is over the stochastic process describing the evolution of the oil price.

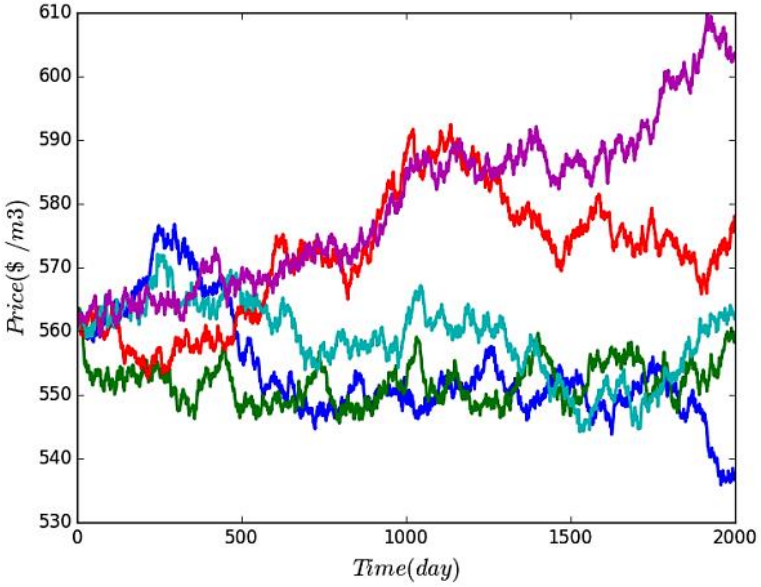

Figure 3.1. Five time series of sampled from the log-normal random walk that models the evolution of the oil price. The samples evolve over 2000 days.

CHAPTER 4. GAUSSIAN PROCESS REGRESSION

The optimization of the oil well placement problem using directly the oil reservoir simulator is computationally infeasible. Therefore, it is essential to replace the (expected) NPV function with a cheap-to-evaluate meta-model (surrogate). The meta-model we use in this work is Gaussian process regression (GPR). [28]

GPR is the Bayesian interpretation of classical Kriging regression techniques.[42] It is a powerful non-linear and non-parametric regression technique that has the added benefit of being able to quantify the epistemic uncertainty induced by the limited observations. As we show in Chapter 5, this uncertainty is the key for deriving information acquisition policies. Here, we cover the mathematical background of GPR covering its formulation as well as the model selection techniques required for selecting the hyper-parameters that describe the covariance function.

4.1 Using Gaussian Processes to Represent Prior Knowledge

A Gaussian process (GP) is a generalization of multivariate normal distribution (MVN) to infinite dimensions. For completeness, let us state here the probability density of a MVN in *D* dimensions:

$$N(\mathbf{x} \mid \boldsymbol{\mu}, \sum) = \frac{1}{(2\pi)^{D/2} \left|\sum\right|^{1/2}} \exp[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \sum^{-1}(\mathbf{x}-\boldsymbol{\mu})], \tag{4.1}$$

where $\boldsymbol{\mu}$ is the mean vector, and $\sum$ is the covariance matrix. Notice that the dimension of $\boldsymbol{\mu}$ is $D\times 1$ while the covariance matrix is a $D\times D$ symmetric positive-definite matrix. A GP is an extension of this MVN to the space of function. In other words, a GP defines a probability measure on a function space. We interpret this probability measure as our prior knowledge about the function before we see any data. In analogy to a MVN, a GP is defined via its mean and covariance, albeit those are now functions. We write:

$$f(\cdot) \mid m(\cdot), k(\cdot,\cdot) \sim \mathrm{GP}(f(\cdot) \mid m(\cdot), k(\cdot,\cdot)), \tag{4.2}$$

where $m(\cdot): \mathbb{R}^D \to \mathbb{R}$ is the *mean function* and $k(\cdot,\cdot): \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is the *covariance function*. Therefore, compared to a multivariate normal we have:

- A random function $f(\cdot)$ instead of a random vector $\mathbf{x}$.
- A mean function $m(\cdot)$ instead of a mean vector $\boldsymbol{\mu}$.
- A covariance function $k(\cdot,\cdot)$ instead of a covariance matrix $\sum$.

Note that the mean function is arbitrary while the covariance function has to be a positive definite function.

In our core example, $\mathbf{x}$ is four dimensional (D=4) and corresponds to the location of the injection and the production wells, and $f(\cdot)$ represents the (expected NPV).

The meaning of Equation (4.2) comes through the MVN. In particular, consider an arbitrary set of N input points represented as an $N \quad D$ matrix:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{pmatrix}, \tag{4.3}$$

as well as the set of function responses on these points represented as an N dimensional vector:

$$\mathbf{f} = \begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}. \tag{4.4}$$

Then, Equation (4.2) *means* that the vector of responses **f** conditional on the inputs **X** and the mean and covariance functions follows a MVN distribution. In particular:

$$\mathbf{f} \mid \mathbf{X}, \mathrm{m}(\cdot), \mathrm{k}(\cdot,\cdot) \sim \mathcal{N}(\mathbf{f} \mid \mathbf{m}, \mathbf{K}), \tag{4.5}$$

with mean vector:

$$\mathbf{m} = \begin{pmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_N) \end{pmatrix}, \tag{4.6}$$

and covariance matrix:

$$\mathbf{K} = \begin{pmatrix} k(\mathbf{x}_1,\mathbf{x}_1) & \dots & k(\mathbf{x}_1,\mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N,\mathbf{x}_1) & \dots & k(\mathbf{x}_N,\mathbf{x}_N) \end{pmatrix}.$$

From another perspective, the mean function gives the expectation of $f(\mathbf{x})$ over our prior uncertainty:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]. \tag{4.7}$$

In order to understand the meaning of the covariance function more deeply, notice that for any $\mathbf{x} \in \mathrm{n}^D$, the variance of $f(\mathbf{x})$ is:

$$\mathbb{V}[f(\mathbf{x})] = \mathbb{E}\left[\left(f(\mathbf{x}) - m(\mathbf{x})\right)^2\right][(f(\mathrm{x}) - m(\mathrm{x}))^2], \tag{4.8}$$

and that, for any $\mathbf{x}' \in \mathrm{n}^D$, the covariance between random variable $f(\mathrm{x})$ and $f(\mathrm{x}')$:

$$\mathbb{C}\left[f(\mathbf{x}), f(\mathbf{x}')\right] = \mathbb{E}\left[\left(f(\mathbf{x}) - m(\mathbf{x})\right)\left(f(\mathbf{x}') - m(\mathbf{x}')\right)\right] = \mathrm{k}(\mathbf{x}, \mathbf{x}'). \tag{4.9}$$

4.2 Covariance Function

In supervised learning, the concept of the covariance function models similarity between observations. The underlying expectation is that when two observations have similar inputs, then they are likely to have similar outputs. In this way, observations that are close to a query point can provide useful information for prediction. Thus the covariance function is an indispensable part of GPR. It encodes the assumptions about the function we wish to learn.

Another name for a covariance function $k(\mathbf{x},\mathbf{x}')$ is *kernel*. In general, the choice of the covariance function depends on the prior assumptions about the regularity of the function space. However, it is usually true that kernels are non-negative $k(\mathbf{x},\mathbf{x}) \geq 0$ and that as the distance between $\mathbf{x}$ and $\mathbf{x}'$ increases $k(\mathbf{x},\mathbf{x}')$ becomes smaller. Some commonly used covariance functions are the squared exponential (SE), the exponential covariance, the linear covariance, the polynomial covariance, the rational quadratic covariance, and the periodic covariance. All of these covariances have free parameters θ, which are known as *hyperparameters*.

In this study, we assume that our prior belief of objective function $f(\mathbf{x})$ conforms with a GP zero prior mean and a squared exponential (SE) covariance:

$$k(\mathbf{x},\mathbf{x}') = s^2 \exp\left\{ -\frac{1}{2}\sum_{i=1}^{D} \frac{(x_i - x_i')^2}{l_i^2} \right\} + \sigma^2 \delta(\mathbf{x} - \mathbf{x}'), \tag{4.10}$$

where the hyperparameters must satisfy $s > 0$ and $l_i > 0$. Here l_i is the characteristic length-scale of the i-th input. Intuitively, it measures how far we need to move in order to make function value of two input points uncorrelated. The SE covariance function

implements automatic relevance determination (ARD). [33] That is, if l_i has a large value then covariance will becomes independent of input i. The parameter s is interpreted as the signal strength. In words, the bigger it is, the more do sample function $f(\cdot)$ from the corresponding GP vary about the mean. The last part of the right hand side models observation noise. The delta function we use resembles Kronecker's delta. It is assumed to be zero unless the two inputs correspond to the same measurement. When we know that our observations are noiseless, then we just set the noise level equal to zero. We only use it when we want to represent the uncertainty in the NVP.

To conclude this section, let us mention once more that the GP constructs a prior probability measure on the space of meta-models which quantifies our state of knowledge about the function of interest. To emphasize this, we will also denote Equation. (4.5) by:

$$f(\cdot) \sim p(f(\cdot)). \tag{4.11}$$

In the following section, we show how observations $\mathcal{D}_N$ can be combined with this prior measure using Bayes theorem, to get the *posterior* probability measure:

$$f(\cdot)\,|\,\mathcal{D}_N \sim p(f(\cdot)\,|\,\mathcal{D}_N). \tag{4.12}$$

The latter is a representation of our state of knowledge about the objective function that is compatible with both the prior knowledge as well as the observations.

4.3 Conditioning a Gaussian Process on Observations

Assume that we have observed a set of N inputs and outputs $\mathcal{D}_N = \{\mathbf{X}_N, \mathbf{f}_N\}$ as defined earlier. Conditioning the prior on this data using the Bayes rule, we get the *posterior* GP measure:

$$f(\cdot)\,|\,\mathcal{D}_N \sim p(f(\cdot)\,|\,\mathcal{D}_N) = \mathrm{GP}(f(\cdot)\,|\,m_N(\cdot), k_N(\cdot,\cdot)), \tag{4.13}$$

where $m_n(\mathbf{x})$ as the *posterior mean function*:

$$m_N(\mathbf{x}) = m(\mathbf{x}) + \mathbf{k}_N(\mathbf{x})\mathbf{K}_N^{1}(\mathbf{m}_N \quad \mathbf{f}_N), \tag{4.14}$$

and $k_n(\mathbf{x},\mathbf{x}')$ is the *posterior covariance function*:

$$k_N(\mathbf{x},\mathbf{x}') = k(\mathbf{x},\mathbf{x}') \quad k_N(\mathbf{x})^T \mathbf{K}_N^{1} k_N(\mathbf{x}'). \tag{4.15}$$

Even though Equation (4.13) contains all posterior information, it is usually more convenient to work with the *point predictive distribution*:

$$y\,|\,\mathbf{x},\mathcal{D}_N \sim p(y\,|\,\mathbf{x},\mathcal{D}_N) = \mathcal{N}\left(y\,|\,m_N(\mathbf{x}), \sigma_N^2(\mathbf{x})\right), \tag{4.16}$$

where the *posterior predictive variance* is simply given by:

$$^{2}_{N}(\mathbf{x}) = k_N(\mathbf{x},\mathbf{x}). \tag{4.17}$$

4.4 Hyper-parameter Selection

In GPR we estimate the hyper-parameters of the covariance function using Bayesian model selection tools. In this study, we use the SE covariance function as default and model selection will be used to find the optimal values for each parameters: the signal strength, the length scale of each dimension, and (if required) the noise level.

The fully Bayesian approach for model selection would be to assign a prior on the hyper-parameters and then characterize their posterior using sampling techniques. However, if the observations are not enough and prior information about the hyper-parameters is vague (e.g., if we assume uniform priors), then one usually obtains good results simply by maximizing the likelihood of the data $p(\mathrm{y}\,|\,\mathrm{X},\theta)$. Since likelihood function is positive, we choose to work with its logarithm. The log likelihood of GPR can be written as:

$$\log p(\mathbf{f}_N \mid \mathbf{X}_N, \ ) = \ \frac{1}{2}\mathbf{f}_N{}^T \mathbf{K}_N{}^1 \mathbf{f}_N \ \ \frac{1}{2}\log\left|\mathbf{K}_N\right| \ \ \frac{N}{2}\log 2 \ , \tag{4.18}$$

where $\mathbf{K}_N$ is the covariance matrix and $= \{s, \ell_1, \ \ , \ell_D, \ ^2\}$ represents all the hyper-parameters of the model. It can be shown that the gradient of the log-likelihood is:

$$\begin{aligned} \frac{}{j}\log p(\mathbf{f}_N \mid \mathbf{X}_N, \ ) &= \ \frac{1}{2}\mathbf{f}_N{}^T \mathbf{K}_N{}^1 \frac{\mathbf{K}_N}{j} \mathbf{K}_N{}^1 \mathbf{f}_N \ \ \frac{1}{2}\mathrm{tr}(\mathbf{K}_N{}^1 \frac{\mathbf{K}_N}{j}) \\ &= \frac{1}{2}\mathrm{tr}((\mathbf{a}_N \mathbf{a}_N{}^T \ \ \mathbf{K}_N{}^1) \frac{\mathbf{K}_N}{j}), \end{aligned} \tag{4.19}$$

where $\mathbf{a}_N = \mathbf{K}_N{}^1 \mathbf{f}_N$.

We solve the log-likelihood maximization problem using the BFGS algorithm[14]. To deal with the positive constraints on the hyper-parameters we work with their logarithms.

CHAPTER 5. BAYESIAN GLOBAL OPTIMIZATION

In this section we address the problem of solving the following design optimization problem:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} f(\mathbf{x}), \tag{5.1}$$

under a limited evaluation budget for the objective function. To start with, assume that we have observed some, initially randomly picked, evaluations of the objective function and that we have built, using the methodology of Chapter 4, a GP representing our posterior state of knowledge. We may use this state of knowledge about the objective to quantify our state of knowledge about the solution to the design optimization problem.

A naïve, but very common approach to characterize our state of knowledge about the optimal solution, would be to replace the objective function with the posterior mean function of the GP characterizing our posterior state of knowledge. Since the posterior mean function is very cheap to evaluate, we may then proceed to solve the design optimization problem. Even though this approach is appealing it hides a lot of problems the most important of which is that the accuracy of the solution we find depends on the accuracy of the surrogate surface.

The idea in Bayesian global optimization (BGO) is to interrogate the posterior GP for design points of high expected value and sequentially iterate between making the most valuable observations and conditioning the posterior GP on them until our data acquisition budget is exhausted. Stating the problem in its full mathematical generality results in a very hard dynamic programing problem.[39]

Since the data acquisition problem in its most general form is computationally intractable, we restrict our attention to the more modest goal of finding good sequential one-step-look-ahead data acquisition policies. These data acquisition policies are also called *myopic* data acquisition policies. A good myopic data acquisition policy must include a tradeoff between two competing goals: *exploration* and *exploitation*. Exploration refers to our need to have strategies that broadly explore the design space on which we are largely uncertain about the value of the objective. Exploitation refers to the desire to use our existing knowledge that the objective is high on certain regions of the design space, zoom in those regions and search for an even better solution. As we will see later on, in BGO, exploration is related to the posterior variance and exploitation to the posterior mean of the GP representing our state of knowledge about the objective.

In what follows, we discuss generic myopic data acquisition policies (Sec. 5.1) including special choices suitable for design optimization involving deterministic surrogates (Probability of improvement in Sec. 5.1.1 and expected improvement in Sec. 5.1.2) and provide special examples 5.1.3. In Sec. 5.2, we propose a generalization of the expected improvement data acquisition policy to the case of design optimization under uncertainty and demonstrate the efficacy of our approach using numerical examples.

5.1 Acquisition Functions for Bayesian Optimization

Assume that we have made N observations $\mathcal{D}_N$ and that our posterior state of knowledge can be neatly summarized through the point predictive distribution given in Equation (4.16):

$$y \mid \mathbf{x}, \mathcal{D}_N \sim p(y \mid \mathbf{x}, \mathcal{D}_N) = \mathcal{N}\left(y \mid m_N(\mathbf{x}), \sigma_N^2(\mathbf{x})\right).$$

Our problem in this section is to find the most valuable design point $\mathbf{x}_{N+1}$ to observe next. The core idea behind the proposed solution is founded on the concept of the *value of information.*[32] Specifically, assume that we made a hypothetical new observation at point $\mathbf{x}$ and that the objective function value was y. Denote by $v(\mathbf{x}, y; D_n)$ the value of making this observation. Since, we do not actually know what y could be, we have to integrate it out of the picture using the point predictive distribution of the posterior GP as Equation. (4.16). Doing this, we may define the expected value of observing $\mathbf{x}$:

$$v(\mathbf{x}; \mathcal{D}_N) = \mathbb{E}_{y|\mathbf{x}, \mathcal{D}_N}\left[v(\mathbf{x}, y; \mathcal{D}_N)\right] = \int v(\mathbf{x}, y; \mathcal{D}_N) p(y \mid \mathbf{x}, \mathcal{D}_N) dy. \tag{5.2}$$

We are now in a position to define a myopic data acquisition policy associated with value function. It boils down to a sequential iterative solution of optimization problem:

$$\mathbf{x}_{N+1} = \text{argmax}_{\mathbf{x}} v(\mathbf{x}; \mathcal{D}_N). \tag{5.3}$$

That is, the myopic optimization policy sequentially observes the objective function on the design points expected to be most valuable. A natural *stopping criterion* for such a generic policy is to stop when the maximum expected value of the current iteration of the policy is smaller that a threshold. Mathematically, we may stop when:

$$v(\mathbf{x}_{N+1}; \mathcal{D}_N) < \epsilon, \tag{5.4}$$

where $\epsilon > 0$ is a an arbitrary tolerance.

In real applications, one must pay special attention to avoid stopping prematurely due to the S-curve phenomenon of information.[39] The S-curve phenomenon expresses the possibility that in many problems exhibit three distinct faces when it comes to information collection. Initially, when data is limited, a single observation might not be informative enough for the objective a fact that is demonstrated with a small value. However, as more data is accumulated, individual observations start becoming more valuable. This behavior can be attributed to the fact that the GP has learnt exploitable information from the data. Finally, as we keep getting closer to the objective, individual observations start becoming less valuable. This pattern clearly resembles an "S", and hence the name "S-curve" of information. In order to avoid premature stopping, one must avoid the very first concave part of the "S". This can be achieved by making sure that we do not stop before a specific number of observations have been made.

In the rest of this subsection, we discuss two choices for the value function: The probability of improvement (PI) and the expected improvement (EI).

5.1.1 Probability of Improvement

The probability of improvement (PI) policy sequentially picks the design that is most likely to yield an improvement over the current maximum.[31] In particular, let

$$\tilde{y}_N = \max_{1 \le n \le N} y_n, \tag{5.5}$$

be the current observed maximum value. PI picks the design point **x** that has the highest probability (according to posterior GP) a value greater than $\tilde{y}_N$:

$$\mathrm{PI}(\mathbf{x}) = \mathrm{P}[f(x) \ge \tilde{y}_N] := \int_{\tilde{y}_N}^{+\infty} p(y \mid \mathbf{x}, \mathcal{D}_N) = \Phi\left(\frac{\tilde{y}_N - m_N(\mathbf{x})}{\sigma_N(\mathbf{x})}\right), \tag{5.6}$$

where we have used the fact that the point predictive distribution is a Gaussian and exploited the properties of the cumulative distribution function of the standard normal.

According to the work of Donald, the PI is an exploitation-only policy. [30] As a consequence, it is commonly acknowledged that its overall performance is not particularly good.

5.1.2 Expected Improvement

The expected improvement (EI) policy was first proposed in 1978 by Mockus *et al.*[37] As its name suggests, EI involves computing how much improvement we expect to achieve; especially over the domain we are largely uncertain. EI picks the design point $\mathbf{x}$ that has the highest expectation of difference between predicted function value y and a value current maximum $\tilde{y}_N$. If we hypothetically know the data point $(\mathbf{x}, y)$, then the improvement we can get from this data point is:

$$I(\mathbf{x}, y) = \max\{0, y - \tilde{y}_N\}, \tag{5.7}$$

however, we do not know the true value of y, unless we actually evaluate objective function. In order to take away the effect of unknown y, we integral y out and find the expectation of improvement function. As mentioned before, our posterior state of knowledge is given in Equation (4.16):

$$y \mid \mathbf{x}, \mathcal{D}_N \sim p(y \mid \mathbf{x}, \mathcal{D}_N) = \mathcal{N}\left(y \mid m_N(\mathbf{x}), \sigma_N^2(\mathbf{x})\right).$$

The expectation of the improvement is simply the expected value of the improvement found by integrating over this density:

$$\begin{aligned}\mathrm{EI}_N(\mathbf{x}) &= \mathbb{E}\left[I(\mathbf{x},y)\right] = \int_{\tilde{y}_N}^{+\infty} I(\mathbf{x},y)\,p(y\mid\mathbf{x},\mathcal{D}_N)dy \\ &= \int_{\tilde{y}_N}^{+\infty} (y-\tilde{y}_N)\mathcal{N}\left(y\mid m_N(\mathbf{x}),\sigma_N^2(\mathbf{x})\right)dy \\ &= \int_{\tilde{y}_N}^{+\infty} y\cdot\mathcal{N}\left(y\mid m_N(\mathbf{x}),\sigma_N^2(\mathbf{x})\right)dy - \tilde{y}_N\int_{\tilde{y}_N}^{+\infty}\mathcal{N}\left(y\mid m_N(\mathbf{x}),\sigma_N^2(\mathbf{x})\right)dy,\end{aligned} \tag{5.8}$$

denote $z=\dfrac{y-m_N(\mathbf{x})}{\sigma_N(\mathbf{x})}$ then we have $dz=\dfrac{dy}{\sigma_N(\mathbf{x})}$ .Separately calculate the integration of Equation (5.8). The first part is simplified as:

$$\begin{aligned}&\int_{\tilde{y}_N}^{+\infty} y\cdot\mathcal{N}\left(y\mid m_N(\mathbf{x}),\sigma_N^2(\mathbf{x})\right)dy \\ &= \int_{\tilde{y}_N}^{+\infty}\left(\sigma_N(\mathbf{x})z+m_N(\mathbf{x})\right)\frac{1}{\sqrt{2\pi}\sigma_N(\mathbf{x})}\cdot\exp\left(-\frac{1}{2}\mathrm{z}^2\right)\sigma_N(\mathbf{x})dz \\ &= \sigma_N(\mathbf{x})\cdot\int_{\tilde{z}_N}^{+\infty} z\phi(\mathrm{z})\,dz + m_N(\mathbf{x})\cdot\int_{\tilde{z}_N}^{+\infty}\phi(z)\,dz \\ &= \sigma_N(\mathbf{x})\cdot\left[\phi(\tilde{z}_N)+c\right]_{-\infty}^{\tilde{z}_N} + m_N(\mathbf{x})\cdot\left[1-\Phi(\tilde{z}_N)\right] \\ &= \sigma_N(\mathbf{x})\cdot\phi(-\tilde{z}_N)+m_N(\mathbf{x})\cdot\Phi(-\tilde{z}_N),\end{aligned} \tag{5.9}$$

the second part is simplified as follows:

$$\begin{aligned}&-\tilde{y}_N\int_{\tilde{y}_N}^{+\infty}\mathcal{N}\left(y\mid m_N(\mathbf{x}),\sigma_N^2(\mathbf{x})\right)dy \\ &= -\tilde{y}_N\left[1-\int_{-\infty}^{\tilde{y}_N}\mathcal{N}\left(y\mid m_N(\mathbf{x}),\sigma_N^2(\mathbf{x})\right)dy\right] \\ &= -\tilde{y}_N\left[1-\int_{-\infty}^{\tilde{z}_n}\mathcal{N}(\sigma_N^2(\mathbf{x})\cdot z+m_N(\mathbf{x})\mid m_N(\mathbf{x}),\sigma_N^2(\mathbf{x}))\sigma_N(\mathbf{x})dz\right] \\ &= -\tilde{y}_N\left[1-\Phi(\tilde{\mathrm{z}}_n)\right] \\ &= -\tilde{y}_N\Phi(-\tilde{\mathrm{z}}_n),\end{aligned} \tag{5.10}$$

then combine the two parts back together:

$$\begin{aligned}&\sigma_N(\mathbf{x})\phi(-\tilde{\mathrm{z}}_N)+m_N(\mathbf{x})\Phi(-\tilde{\mathrm{z}}_N)-\tilde{y}_N\Phi(-\tilde{\mathrm{z}}_N) \\ &= \sigma_N(\mathbf{x})\cdot\phi\left(\frac{m_N(\mathbf{x})-\tilde{y}_N}{\sigma_N(\mathbf{x})}\right)+\left(m_N(\mathbf{x})-\tilde{y}_N\right)\Phi\left(\frac{m_N(\mathbf{x})-\tilde{y}_N}{\sigma_N(\mathbf{x})}\right),\end{aligned} \tag{5.11}$$

in above equations, $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal density and cumulative distribution function.

Then, the point "best" point to observe is the one that maximizes the expected improvement, i.e.:

$$\mathbf{x}^* = \arg\max \mathrm{EI}_N(\mathbf{x}). \tag{5.12}$$

Summarizing, the expected improvement after observing an arbitrary point is

$$\mathrm{EI}_N(\mathbf{x}) = \sigma_N(\mathbf{x}) \cdot \phi\left(\frac{m_N(\mathbf{x}) - \tilde{y}_N}{\sigma_N(\mathbf{x})} \right) + \left(m_N(\mathbf{x}) - \tilde{y}_N \right) \Phi\left(\frac{m_N(\mathbf{x}) - \tilde{y}_N}{\sigma_N(\mathbf{x})} \right). \tag{5.13}$$

If our objective was to find the global minimum (instead of the global maximum), then we would have:

$$\mathrm{EI}_N(\mathbf{x}) = \sigma_N(\mathbf{x}) \cdot \phi\left(\frac{\tilde{y}_N - m_N(\mathbf{x})}{\sigma_N(\mathbf{x})} \right) + \left(\tilde{y}_N - m_N(\mathbf{x}) \right) \Phi\left(\frac{\tilde{y}_N - m_N(\mathbf{x})}{\sigma_N(\mathbf{x})} \right), \tag{5.14}$$

where $\tilde{y}_N$ now is the current observed minimum (instead of the current observed maximum).

5.1.3 Applications

In this section, we explore the performance of the EI policy on some toy objective functions. We start with a simple one dimensional example. Our aim is to find the global maximum of:

$$\begin{gathered} f(x) = 4 \cdot \sin\left(20 \times (x - 0.5)^2 \right) + \cos(20x + 0.1)^2, \\ x \in [0,1]. \end{gathered} \tag{5.15}$$

We start with a randomly selected set of 6 function evaluations. Then we apply the Gaussian process methodology of Chapter 4 in order to construct the posterior measure representing our state of knowledge based on this initial pool of data. The mean of this posterior Gaussian process is the blue line of Figure 5.1(a) while the 95% credible interval is shown by the grey shaded area. This blue line is to be compared to the red line

which depicts the true, albeit unknown, objective function. In the same plot, we show the expected improvement as a function of the design space (green dashed line). The most informative observation (to be made next) is marked with a green disk. Figures 5.1(b-f) show the same quantities as we iteratively add 1,…,5 observations. We can see that as the number of iterations increases, the posterior mean gets closer to the actual objective function, especially, in the area around the actual global maximum. The value of the expected improvement, shown on the right-hand scale, is gradually decreasing. This fact indicates the observations become less and less informative.

The second example we consider is 2D objective function known as the "Sasena function". The Sasena function is defined by:

$$y(\mathbf{x}) = 2 + 0.01\times(x_2 - x_1^2)^2 + (1-x_1)^2 + 2\times(2-x_2)^2 + 7\sin(0.5x_2)\cdot\sin(0.7x_1x_2), \quad 0 \le x_1 \le 5, 0 \le x_2 \le 5. \tag{5.16}$$

In this test, the aim is to find global minimum. The starting observation pool consists of twenty randomly picked design points. We present the results in Figure 5.2 and Figure 5.3. This numerical example demonstrates that the EI policy balances exploration and exploitation. At first two iterations the new points chosen are not close to the currents minimum region. The reason they are chosen is due to their high predictive uncertainty. During this phase, EI actually explores. After this phase, the last six observations are all around the current minimum region. This corresponds to the exploitation phase. Finally, notice the dramatic decrease of the EI as the number of iterations increases. This behavior is identical to the one we observed in the 1D toy example.

We end this section by testing the performance of EI on two objective functions with 3 and 6 design inputs. Those functions are the Harmant3 and Harmant6 and our problem is to minimize them.[16] The Hartman 3 function has 3 inputs ($n_{dv} = 3$) and functional form:

$$f(\mathbf{x}) = \sum_{i=1}^{4} a_i \exp\left(\sum_{j=1}^{n_{dv}} B_{ij}(x_j \quad D_{ij})^2 \right),$$
$$\mathbf{a} = [1.0 \quad 1.2 \quad 3.0 \quad 3.2],$$
$$0 \quad x_j \quad 1, j = 1,2, \square \ , n_{dv}, \tag{5.17}$$

with parameters:

$$B = \begin{bmatrix} 3.0 & 10.0 & 30.0 \\ 0.1 & 10.0 & 35.0 \\ 3.0 & 10.0 & 30.0 \\ 0.1 & 10.0 & 35.0 \end{bmatrix}, D = \begin{bmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.03815 & 0.5743 & 0.8828 \end{bmatrix}, \tag{5.18}$$

The Hartman 6 function has 6 inputs ($n_{dv} = 6$), the same functional form as Hartman3, but with parameters:

$$B = \begin{bmatrix} 10.0 & 3.0 & 17.0 & 3.5 & 1.7 & 8.0 \\ 0.05 & 10.0 & 17.0 & 0.1 & 8.0 & 14.0 \\ 3.0 & 3.5 & 1.7 & 10.0 & 17.0 & 8.0 \\ 17.0 & 8.0 & 0.05 & 10.0 & 0.1 & 14.0 \end{bmatrix},$$
$$D = \begin{bmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{bmatrix}. \tag{5.19}$$

The analytical global minimum for function Harmant3 and Harmant6 are known.[26] They are:

$$f(\mathbf{x}^*) = -3.86278,$$

$$\mathbf{x}^* = (0.114614 \quad 0.555649 \quad 0.852547)$$

and

$$f(\mathbf{x}^*) = -3.32237,$$

$$\mathbf{x}^* = (\ 0.20169 \quad 0.150011 \quad 0.476874 \quad 0.275332 \quad 0.311652 \quad 0.6573\ ),$$

respectively.

In these examples, we investigate the dependence of the performance of the EI on the choice of the initial pool. Specifically, we run our algorithm 50 times, each time starting from a different, randomly picked initial observation pool. Figure 5.4 and Figure 5.5 present the results. The red symbols indicate the evolution of the observed minimum as a function of the number of iterations of BGO averaged over the 50 runs of the algorithm. The shaded grey area corresponds to a 95% prediction interval about the average performance. We see that for Hartma3, the average of the best current minimum does converge to the analytic solution quite fast. On the other hand, Hartmant6, requires more than 30 iterations to reach the analytic solution. According to work done by Donald, with 65 initial function evaluations and iteration 121 times, it can reach analytical minimum with 1% error.[30] Nevertheless, in both cases, we see that the choice of the initial pool fades as the number of iterations increase.

5.2 Expected Improvement for Noisy Objectives

In real world design problems, chances are that the objective function is contaminated noise. That is, if we try to evaluate it twice at the same design point, we will get slightly different answers. Ignoring risk aversion, a common approach in such situations is to attempt to maximize the expectation of the objective function:

$$\mathbf{x}^* = \text{argmax}_{\mathbf{x}} \mathbb{E}_{\xi} f(\mathbf{x}, \xi), \tag{5.20}$$

over all uncertain variables ξ.

Even though GPR can easily cope with noisy observations (assuming that the noise is more or less stationary), the standard expected improvement defined in Equations (5.14) and (5.13) fails. The reason is as follows. If a certain observation contains relatively large noise, it may mislead the EI policy to regard this observation as the current maximum. Basically, the problem arises when the EI attempts to interpret an observations magnitude as true value instead of value plus noise. We deal with this problem, by proposing a simple extension to the EI policy that makes it robust to noisy observations.

Assume that we have observed a set of N inputs and outputs $\mathcal{D}_N = \{\mathbf{X}_N, \mathbf{f}_N\}$ as defined earlier. Our posterior state of knowledge is still illustrated in Equation (4.16):

$$y \mid \mathbf{x}, \mathcal{D}_N \sim p(y \mid \mathbf{x}, \mathcal{D}_N) = \mathcal{N}\left(y \mid m_N(\mathbf{x}), \sigma_N^2(\mathbf{x})\right).$$

The posterior variance $^{2}_{N}(\mathbf{x})$ can be seen as having two parts:

$$^{2}_{N}(\mathbf{x}) = \ ^{2}_{0,N}(\mathbf{x}) + \ ^{2}, \tag{5.21}$$

where $^{2}_{0,N}(\mathbf{x})$ is the part of the posterior GP variance that does not depend on the noise, and ^{2} is the posterior GP variance attributed to noise. This distinction is useful in defining the expected improvement later on.

Similarly to the definition of the current observed maximum $\tilde{y}_N$ we use for the noiseless case, we define the current observed *maximum projection*:

$$\tilde{m}_N = \max_{1 \le n \le N} m_N(\mathbf{x}_n). \tag{5.22}$$

This is just the maximum of the projection of the current observations to the mean of the posterior GP. It is intuitively clear that, in the presence of noise, it is a much better indicator of our information about the maximum of the objective function than the current observed maximum.

Using these definitions, we define the extended EI to be:

$$\mathrm{EI}_N(\mathbf{x}) = \sigma_{0,N}(\mathbf{x}) \cdot \phi\left(\frac{m_N(\mathbf{x}) - \tilde{m}_N}{\sigma_{0,N}(\mathbf{x})}\right) + \left(m_N(\mathbf{x}) - \tilde{m}_N\right)\Phi\left(\frac{m_N(\mathbf{x}) - \tilde{m}_N}{\sigma_{0,N}(\mathbf{x})}\right). \tag{5.23}$$

Even though our definition seems ad hoc, in reality it can be formally derived through arguments similar to the ones we used for the derivation of the standard EI. In particular, it can be derived by taking the expectation of the improvement over the predictive distribution about $f(\mathbf{x})$ instead of y.

5.2.1 Applications

We start with a toy 1D application. The objective function that we want to maximize is:

$$\begin{gathered} f(x) = 4 \times \left(1 - \sin(x + 8e^{x-7})\right) + \sigma \times \mathcal{N}(0,1), \\ x \in [0,6], \end{gathered} \tag{5.24}$$

where noise submits to σ times standard normal distribution $\mathcal{N}(0,1)$ and $\sigma = 1.5$. Note that our scheme does not "see" the objective function. All it sees is samples from Eq.

(5.23) at given design points. Finally, note also that the noise level is also assumed unknown. It is automatically approximated by the GPR.

Figure 5.6 shows picked according to extended EI policy. Here, the red line is the objective function without noise, the blue line is the mean function given by posterior, and the grey areas denote 95% prediction intervals about the mean. Notice that in the presence of noise, the model no longer interpolates the data. The green line is the extended expected improvement. In subfigure (a), we see that with just three observations that model is unable to identify any specific trend. However, after a few more observations are added, the method gradually builds a better and better picture about both the objective function and the measurement noise. We see that, after an initial exploration phase, the policy passes into an exploitation phase during which it first identifies the location of the maximum, and then it tries to estimate its value more accurately by repeatedly sampling in that region. The fact that the initial EI (see subfigure (a)) is very small, then increases, and finally decreases again, is a manifestation of the S-curve phenomenon of information.

As a final test, we also consider the 2D Sasena function with Gaussian noise $\sigma \times \mathcal{N}(0,1)$ where $\sigma = 0.5$. The evolution of the current best projected minimum as a function of the number of observations is shown in Figure 5.8. The spikes observed at the very beginning are cause by the limited data used to build the surrogate in combination with the maximum likelihood technique used to train it. In particular, what happens is that the model becomes overconfident (it essentially overfits the observations). This means that the estimate noise is smaller than the true one, which causes the current

projected minima to be closer to the noise-contaminated current observed minima. The situation is gradually remedied as more observations are made.

Based on all results above, we can conclude that the extended EI policy maintains all the advantage that EI. At the same time, it can successfully approach to global optimum even with noisy objective function observations.

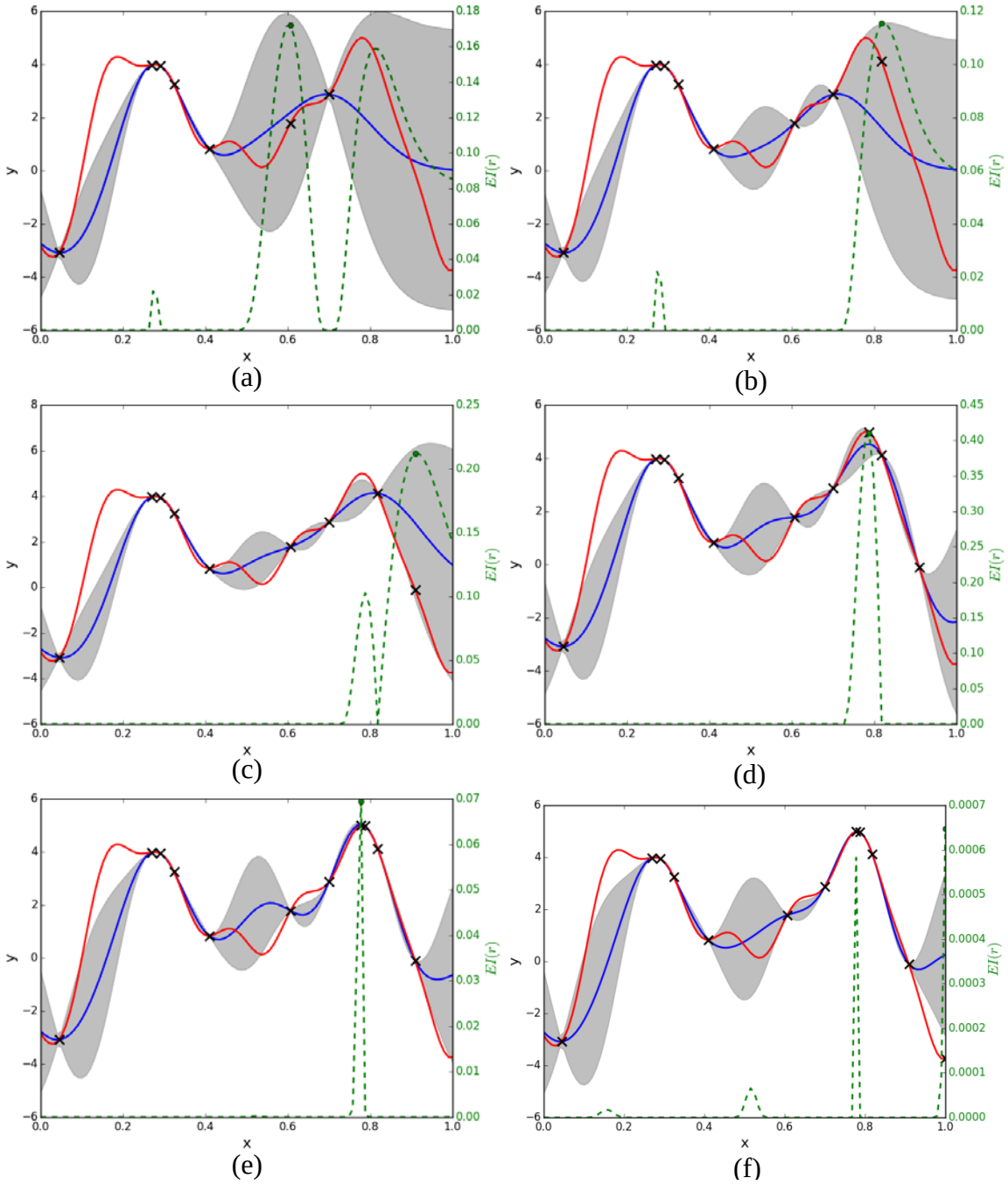

Figure 5.1. Application of EI on a simple one-dimensional test function. Start with 6 initial observations; n stands for number of iteration
(a)n=1;(b)n=2(c)n=3;(d)n=4;(e)n=5;(f)n=6.

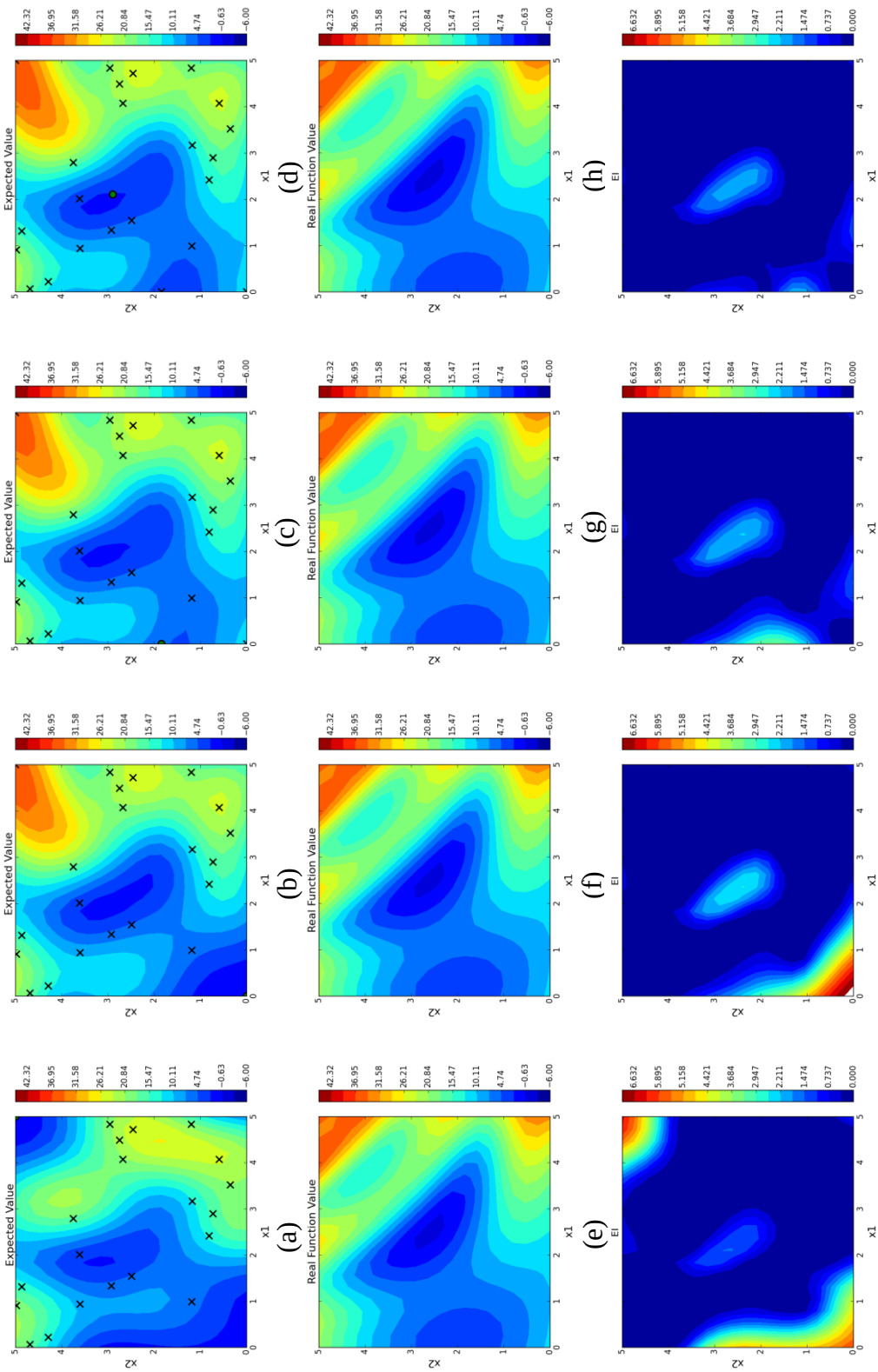

Figure 5.2. Application of EI on a simple two-dimensional Sasena function, n stands for number of iterations; (a)~(d) posterior prediction for n=1,2,3,4; (e)~(h) actual function value; (i)~(l) EI function value for n=1,2,3,4.

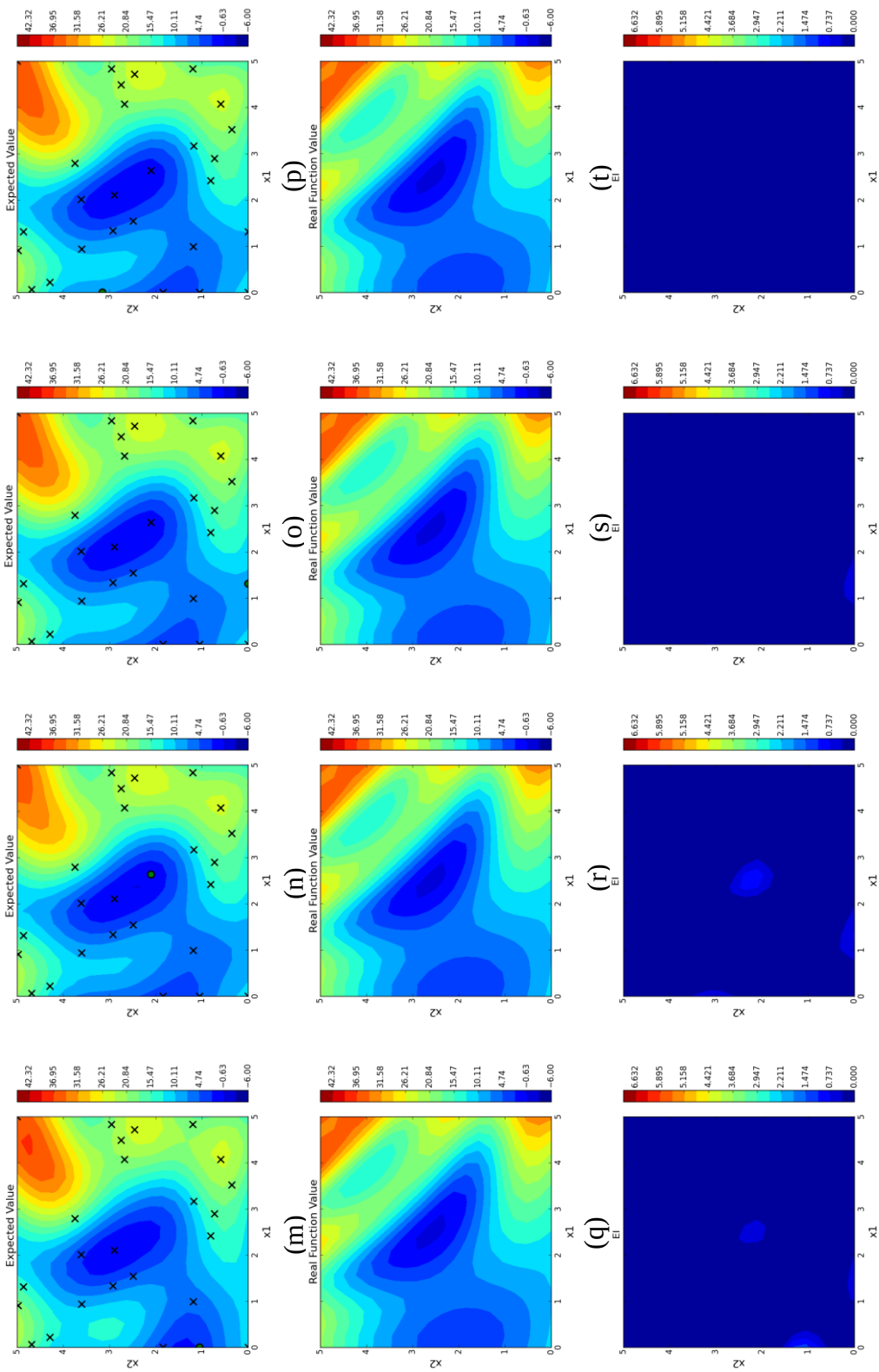

Figure 5.3. Application of EI on a simple two-dimensional Sasena function, n stands for number of iterations; (m)~(n) posterior prediction for n=5,6,7,8; (q)~(t) actual function value; (u)~(x) EI function value for n=5,6,7,8.

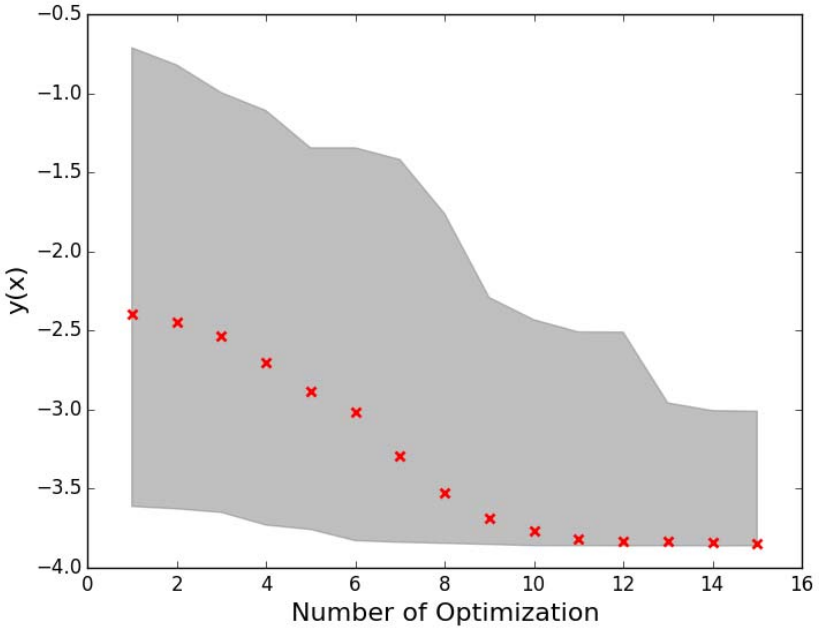

Figure 5.4. Application of EI on Harmant 3 function, sampling 50 times.

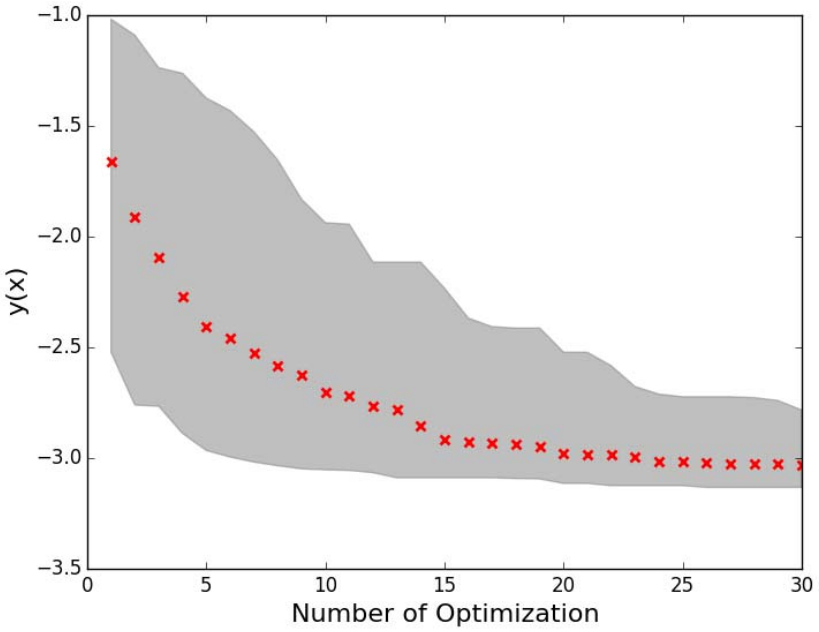

Figure 5.5. Application of EI on Harmant 6 function, sampling 50 times.

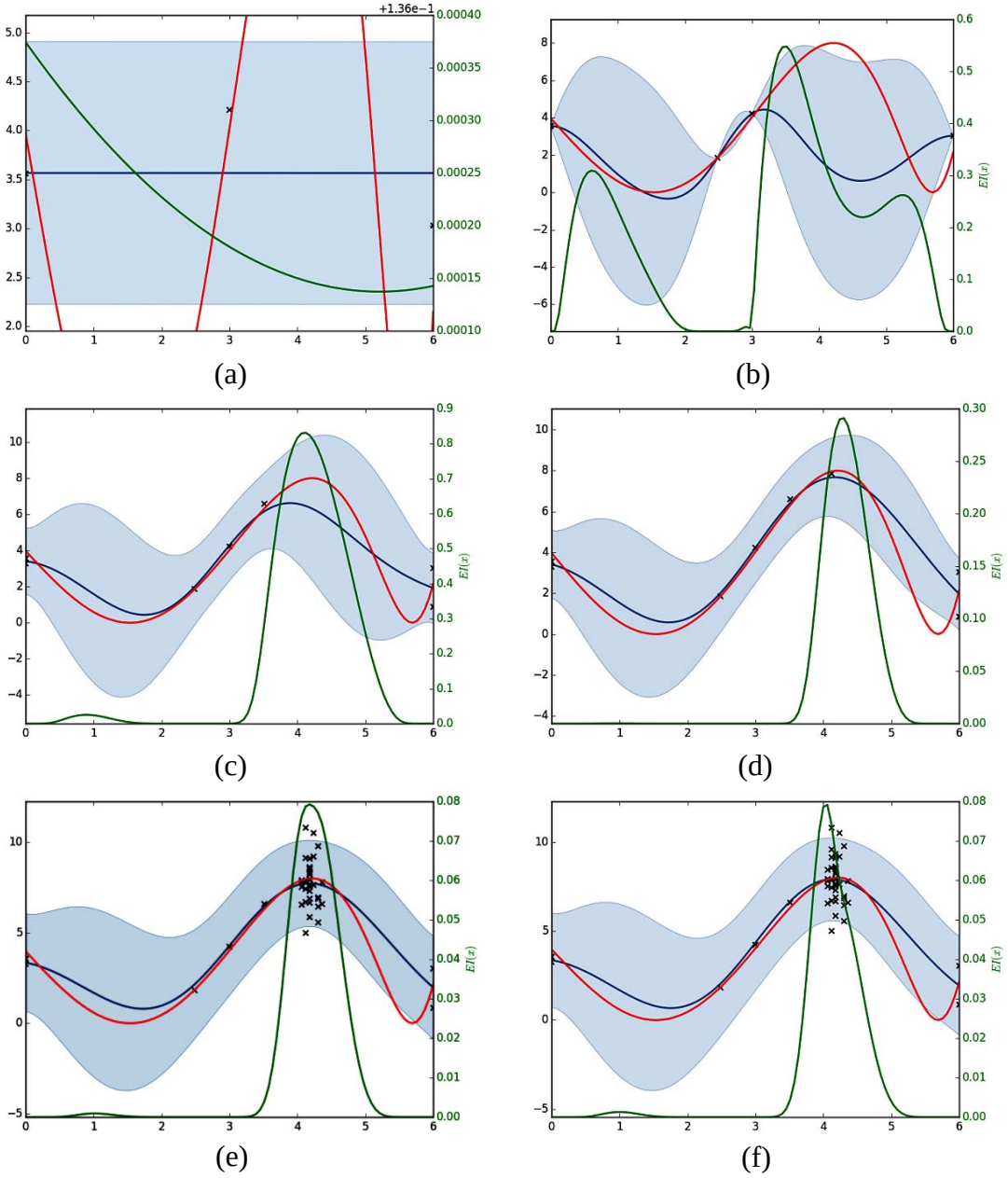

Figure 5.6. Application of extended EI on noisy 1D function, n stands for number of iterations; (a)n=0;(b)n=2;(c)n=4;(d)n=5;(e)n=6;(f)n=41.

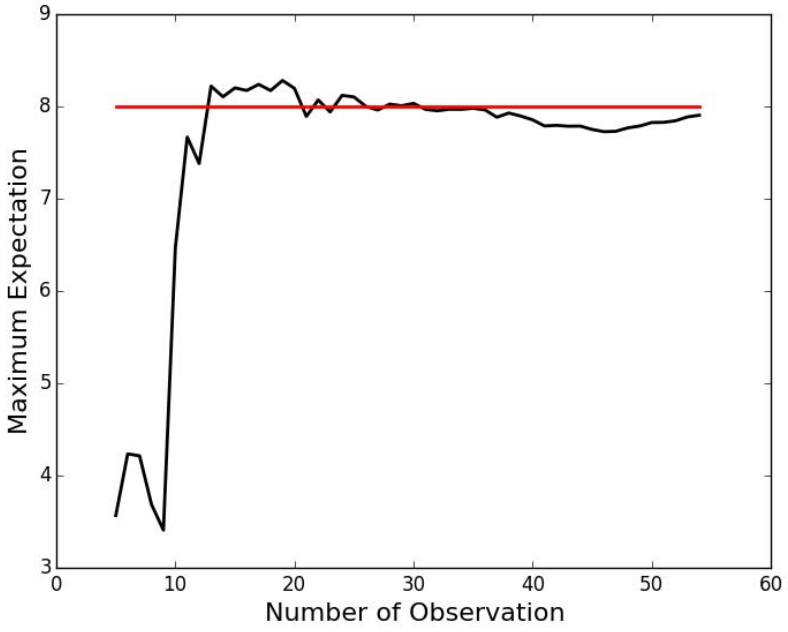

Figure 5.7. Application of extended EI on noisy 1D function, with 3 initial observations.

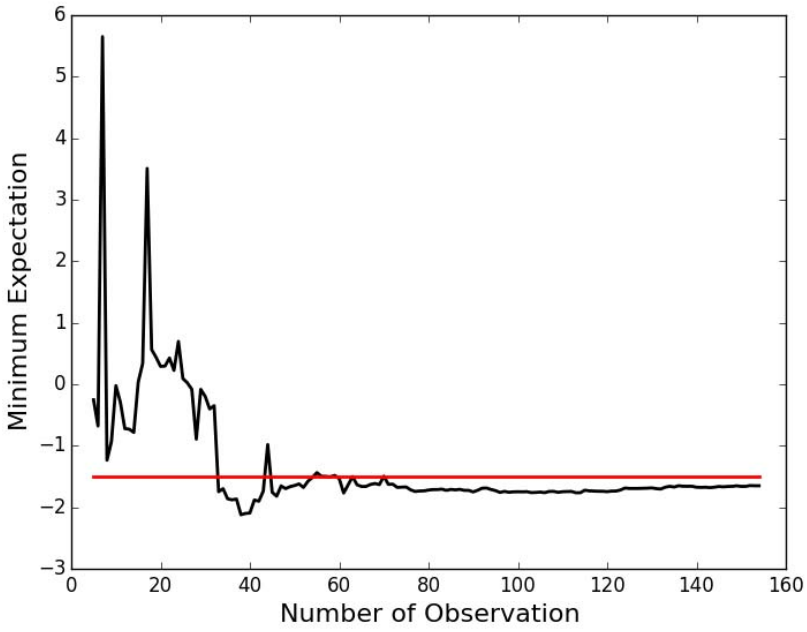

Figure 5.8. Application of extended EI on noisy 2D Sasena function, with 20 initial observations.

CHAPTER 6. AN APPLICATION: WELL PLACEMENT PROBLEM

In the oil well placement problem, our aim is to identify the best well location that can maximize NPV function. Due to limited budget, computational difficulties, and the uncertainty in the specification of the oil price as well as the permeability, this problem is a global optimization problem with a noisy NPV. In this section, we address this problem by considering four scenaria of increasing difficulty: (1) Oil price and permeability are known (noise-free case); (2) Oil price is modeled as log-normal random walk in Equation (3.3) and the permeability is known; (3) Both oil price and permeability are uncertain. All the geophysical parameters, such as permeability tensor, the porosity, the viscosities of the various phases, are taken from the SPE Comparative Solution Project.[21] For comparison purposes, we also approximate the solution to the first two scenaria using a random search optimization approach.[43] Specifically, we evaluate the NPV for this scenario at 16384 randomly selected well locations and pick the one that has the maximum value. We used a latin hyper-cube random design [29] using the tools implemented in the Python package py-design.[11] Note that for the second scenario, the solution using a random search is possible because we can actually evaluate the expectation of the NPV with respect to the oil price by employing sample averages that do not depend on the solutions of the PDE. This is not possible for the third scenario.

Based on limited observations, we construct a surrogate using GPR to quantify our state of knowledge about NPV and then use the (extended) EI policy to actively select the most valuable design points.

6.1 Well Location Result from Random Search

Figure 6.1 is the permeability field showing optimal well locations marked as a cross as identified by the random search optimization. Figure 6.1(a) is given with the assumption that oil price is a constant number during the whole process of production while the oil price in Figure 6.1(b) is given by log-random walk model. The best well locations are selected via a random search that uses 16384 points in each case. For the case of random oil price, the expected NPV for each random well location is evaluated by a sample average using 10,000 samples. From the results, we can see that the best well locations are identical. This is not a coincidence. It is due to the fact that the expected NPV is a linear function of the oil price. However, in the maximum value of NPV function is different.

6.2 No Uncertainty

With the assumption that oil price is constant and the permeability exactly known, we address the following optimization problem:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} f_T(\mathbf{x}), \tag{6.1}$$

where $f_T(\mathbf{x})$ is the NPV function given in Equation (3.1) with modification $c_o(t) = c_o$. Figure 6.2 shows the evolution of the current best observed NPV as a function of the number of PDE evaluations. The red line is the maximum NPV found by the random

search. The maximum NPV value increases in steps, and finally surpasses the best value found by the random search. By only using 5 initial observations and optimize 100 times, BGO method already gives considerably better results than a plain vanilla random search. Figure 6.3 depicts the current best well location. Note that best well locations found by BGO are close to the random search results, but not identical. It is evident, however, that BGO finds a better solution at a fraction of the cost.

6.3 Aleatoric Uncertainty Existing in Oil Price

We now consider the case in which the NPV depends on an uncertain oil price. Specifically, our optimization problem becomes:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \mathbb{E}_{\{c_o(t)\}_{t=1}^T}\left[f_T(\mathbf{x})\right], \tag{6.2}$$

where $f(\mathbf{x})$ is the noisy form of Equation (3.1) based on assumption $c_o(t) = c_o(t,\xi)$.

In practice, when making the investment decision, we should not only aim to maximize the expectation but also minimize the risk associated with NPV. Mathematically, we would like to maximize the expectation of NPV while minimizing its variance. Intuitively, more risk always leads to larger expected reward, i.e., these two objectives are negative correlated. This multiple objective optimization can be addressed by employing the Pareto front concept. In order to quantitatively find the trade-off between expectation and uncertainty, we create the Figure 6.4 to visually check the result by random search. In Figure 6.4, if we draw a line that can cover all the blue dots in the figure, then this line is called Pareto Front line. Pareto front line represents the best achievable trade-off between expectation and variance.

Note that we computed the variance of the NPV using the following formula:

$$
\begin{aligned}
Var\left(\sum_{i=1}^{N} y_i(x)P_i\right) &= \sum_{i,j=1}^{N} y_i(x)y_j(x)Cov(P_i,P_j) \\
&= \sum_{i=1}^{N} y_i^2(x)Var(P_i) + \sum_{i\neq j} y_i(x)y_j(x)Cov(P_i,P_j) \\
&= \sum_{i=1}^{N} y_i^2(x)Var(P_i) + 2\sum_{1\le i\le j\le N} y_i(x)y_j(x)Cov(P_i,P_j) \\
&= \sum_{i=1}^{N} y_i^2(x)Var(P_i) + 2\sum_{1\le i\le j\le N} y_i(x)y_j(x)Cov(P_i,P_j),
\end{aligned}
\tag{6.3}
$$

where $y_i(x)$ stands for the oil production rate for day i.

To solve the noisy optimization problem, we employ the extended EI data acquisition policy. In Figure 6.5 we plot the evolution of the current maximum observed projected value as a function of the number of observations. Starting with just 5 initial random observations, we reach a solution as good as the random search after 30 iterations. After 200 iterations of iterations, we find a solution with value about 1% larger than the best random search value. In Figure 6.6 we show the best well locations found in this case. They are also near the random search result, albeit not identical.

6.4 Aleatoric Uncertainty in Oil Price and Epistemic Uncertainty in Permeability

In this final example, we consider oil well placement problem with quantified uncertainties for both the oil price and the permeability. The design optimization problem we have to solve is: $\mathbf{x}^*$:

$$
\mathbf{x}^* = \arg\max_{\mathbf{x}} \mathbb{E}_{\{c_o(t)\}_{t=1}^{T},K}\left[f_T(\mathbf{x})\right], \tag{6.4}
$$

where the oil price is as before and K represents uncertain permeability. We apply the BGO approach using the extended version of the EI to this problem.

Figure 6.7 depicts the evolution of the current observed projected maximum as a function of the number of observations. Convergence is slower than before, albeit steady. The spikes are due to the discounts on the Bayesian formalism induced by the fact that the underlying GP is trained via maximum likelihood. Figure 6.8 shows the best well location that the algorithm discovers. Comparing the results in Figure 6.8 and Figure 6.6, we see that epistemic uncertainty in permeability brings makes BGO to require more iterations in order to converge.

We end this section by studying the uncertainty of the NPV for a fixed well location via Monte Carlo. Specifically, we take 1,000 joint samples of all uncertain quantities, and compute the NPV for each sample. Then, we construct the histogram of the net present value. Figure 6.9 depicts the result we obtain for two scenaria: 1) uncertain oil price; 2) uncertain oil price and permeability. We can see the variance for latter case is much larger than that in former. This is expected, for more uncertainty should bring more variance. Notice that the mean values of two test cases are different, with the mean of the second scenario being significantly smaller than the mean of the first scenario. A more detailed uncertainty propagation technique would require state-of-the-art methodologies. We refer the reader to the extensive literature.[7-9, 17]

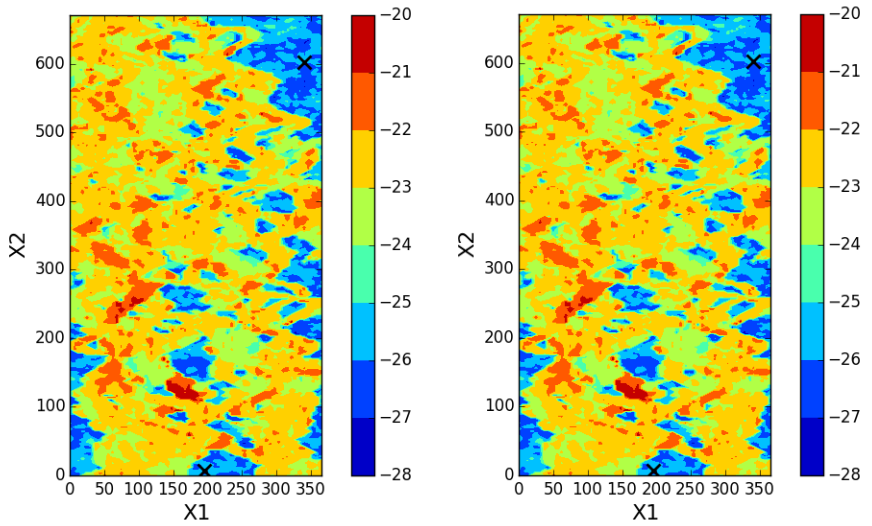

Figure 6.1. Permeability field showing best well location by random search.

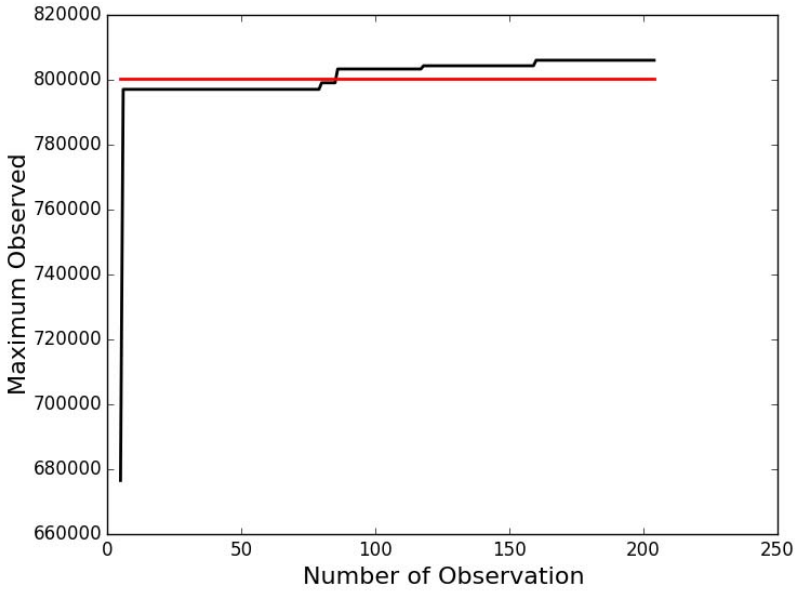

Figure 6.2. BGO with EI criterion for oil well placement prediction when oil price is constant, starting with 5 observations and optimize 200 times.

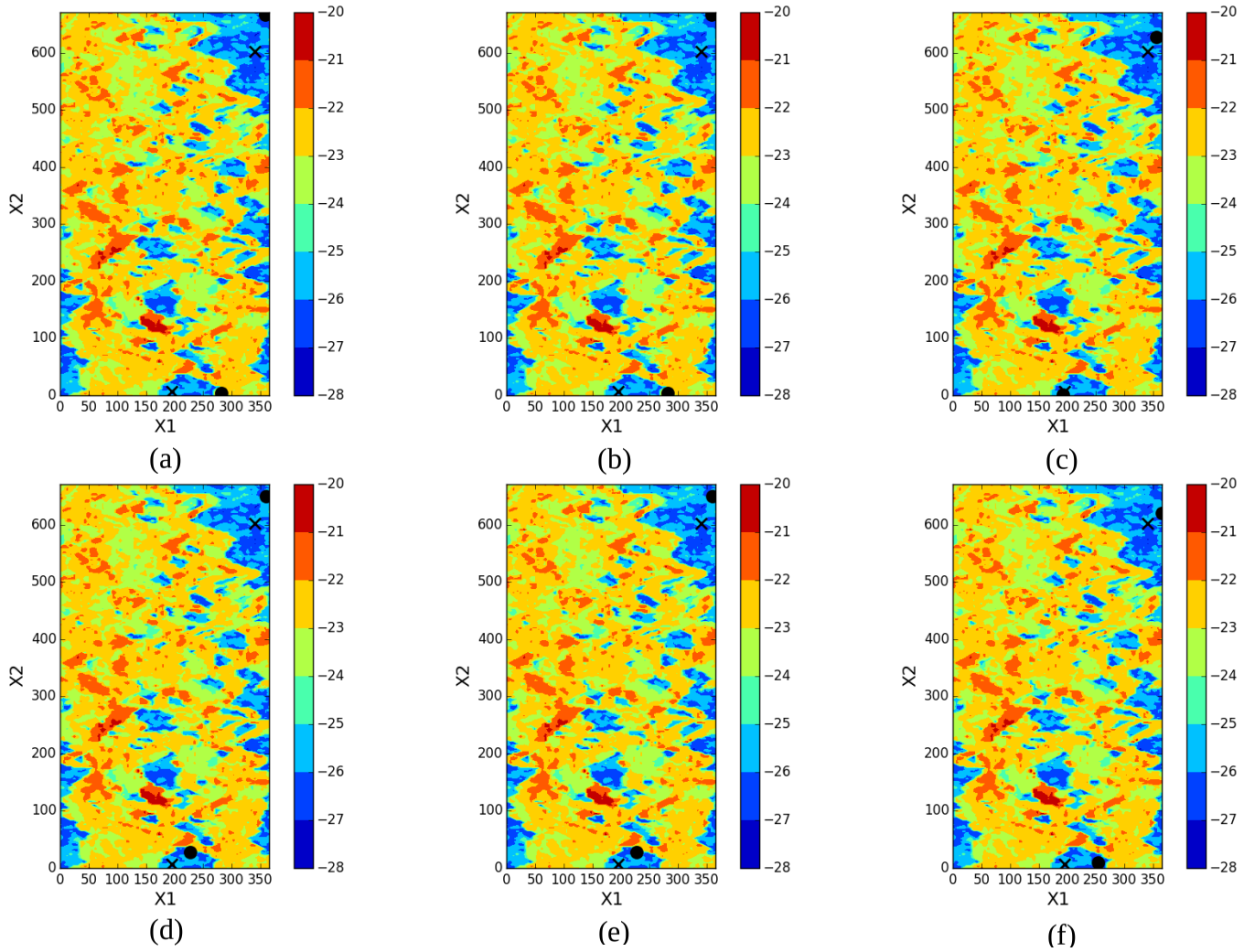

Figure 6.3. Permeability field showing the positions of current wells. The symbols cross indicates frequentist well location and dot indicates predicted well location, respectively; (a) n=1;(b)n=50;(c)n=100;(d)n=125;(e)n=150;(f)n=195.

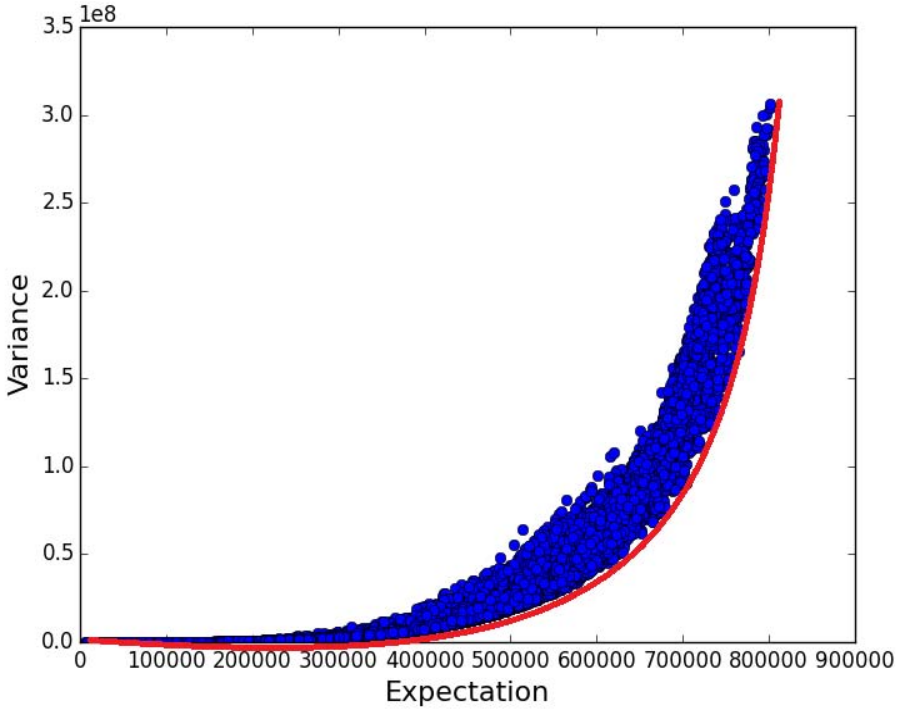

Figure 6.4. Pareto Font line for NPV when oil price has uncertainty.

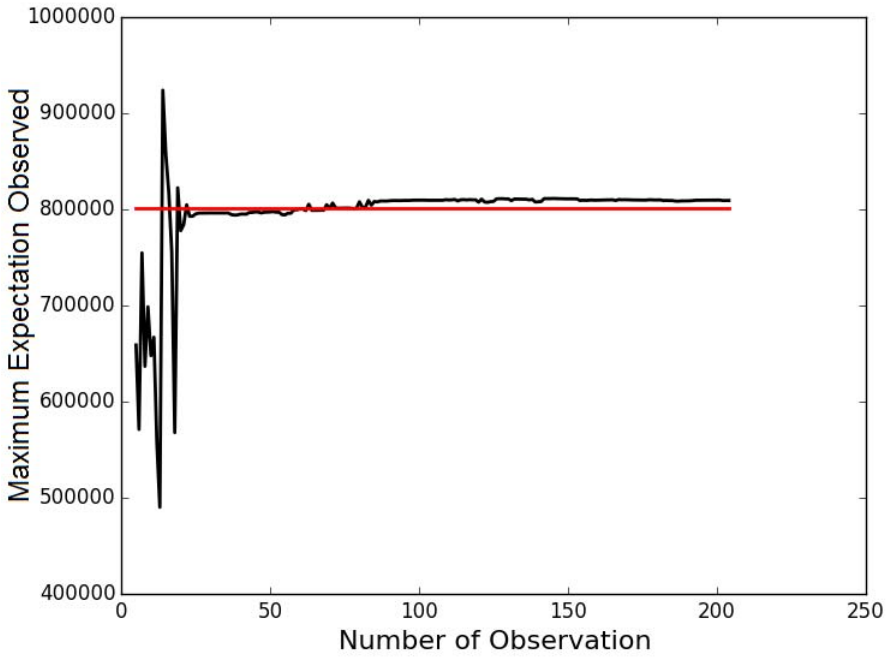

Figure 6.5. BGO with extended EI criterion for oil well placement prediction when oil price has uncertainty.

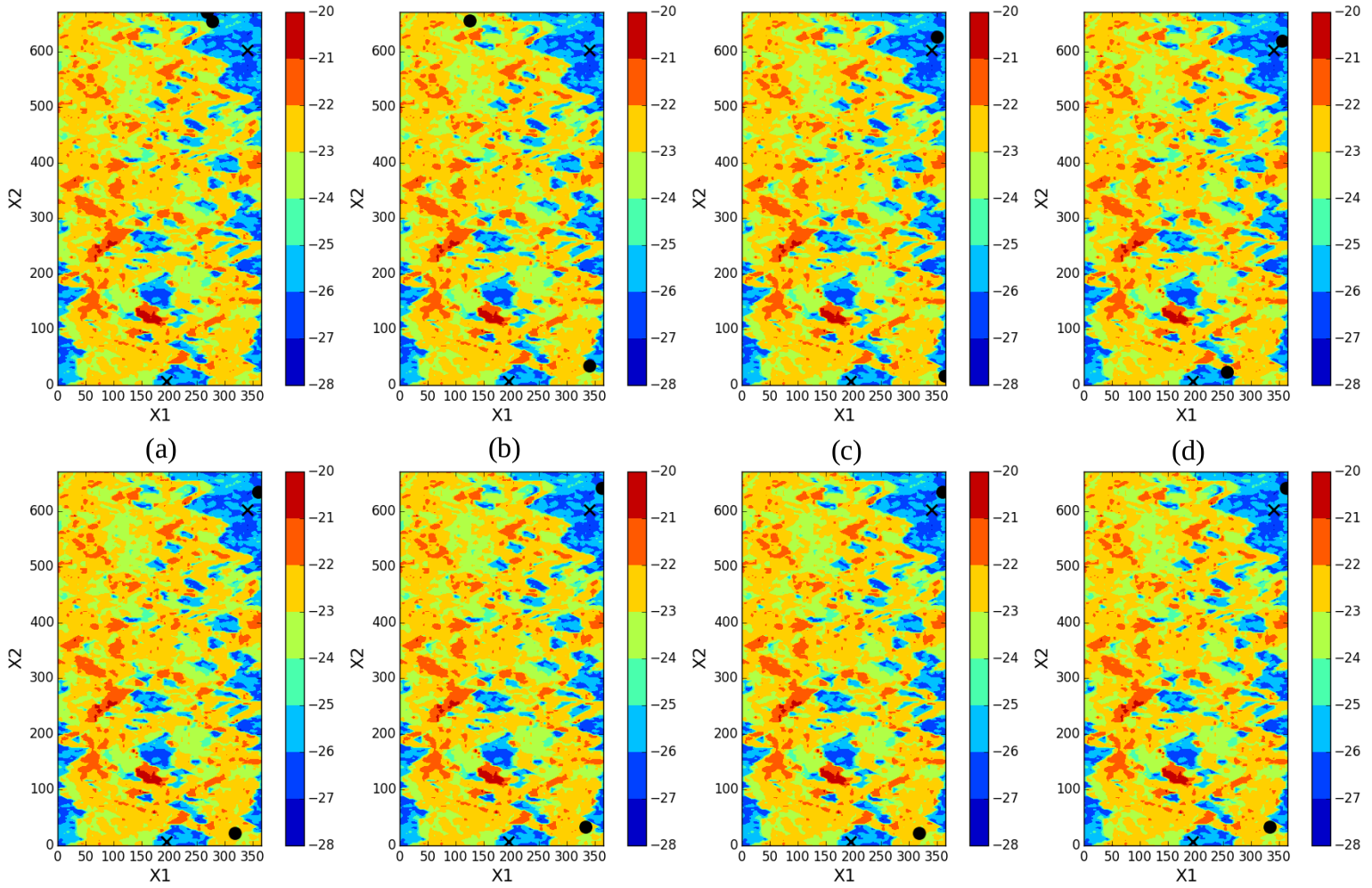

Figure 6.6. Permeability field showing the positions of current wells. The symbols cross indicates well location given by sampling and dot indicates predicted well location respectively; (a)n=1;(b)n=10;(c)n=30;(d)n=50;(e)n=100;(f)n=150;(g)n=175;(h)n=194.

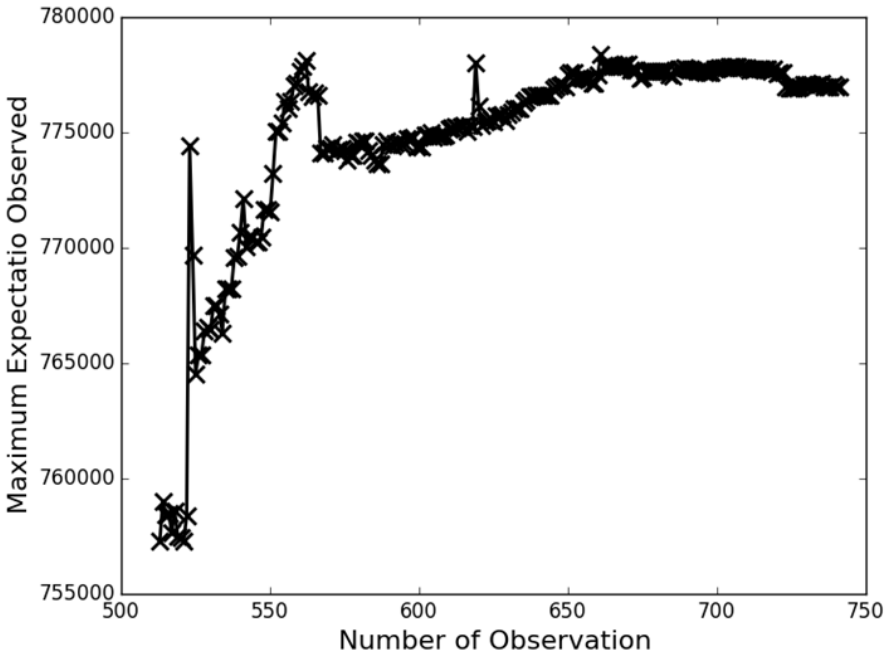

Figure 6.7. BGO with extended EI criterion for oil well placement prediction when oil price and permeability has uncertainty.

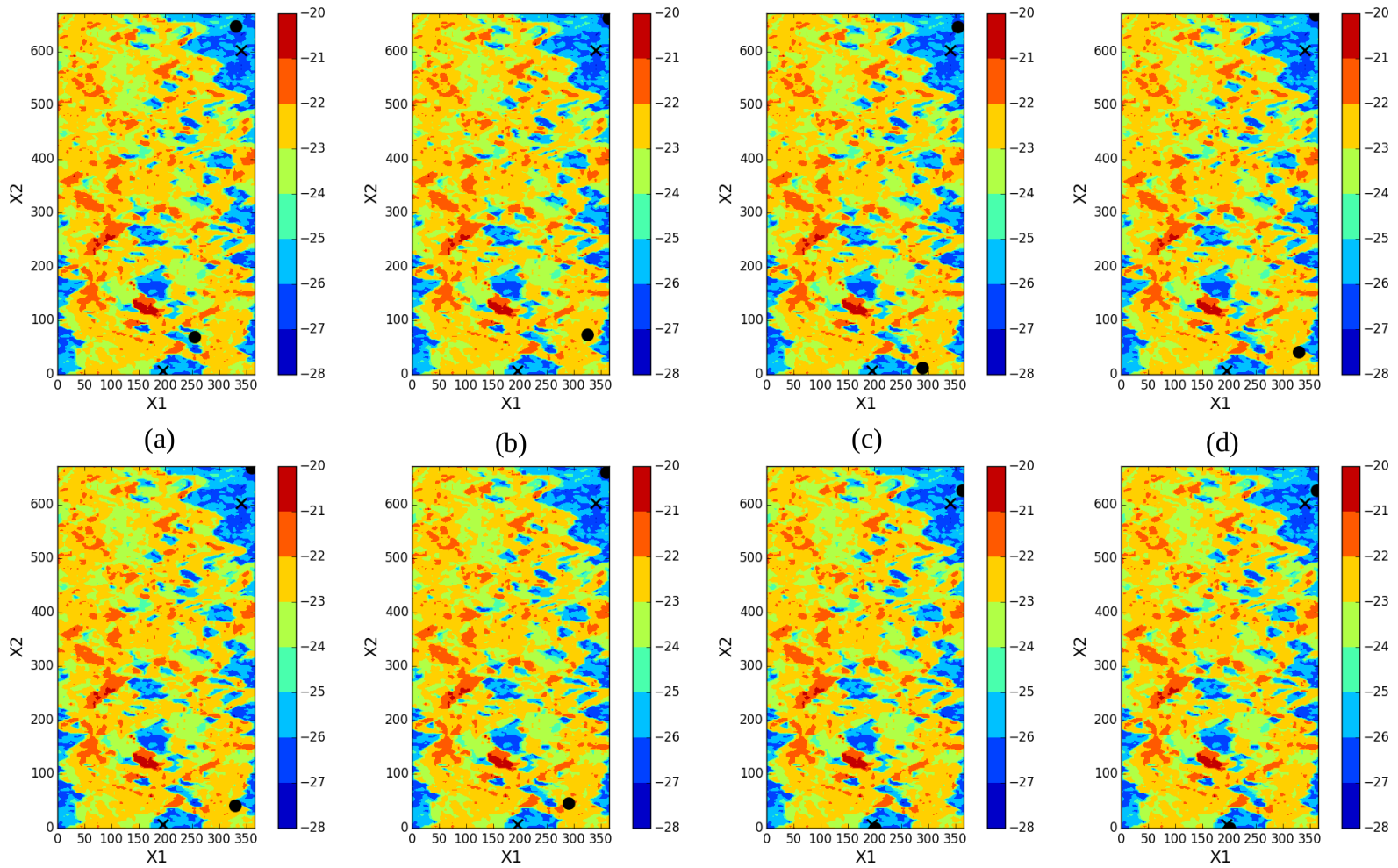

Figure 6.8. Permeability field showing the positions of current wells. The symbols cross indicates well location given by sampling and dot indicates predicted well location; (a)n=1;(b)n=50;(c)n=200;(d)n=300;(e)n=400;(f)n=450;(g)n=475;(h)n=500.

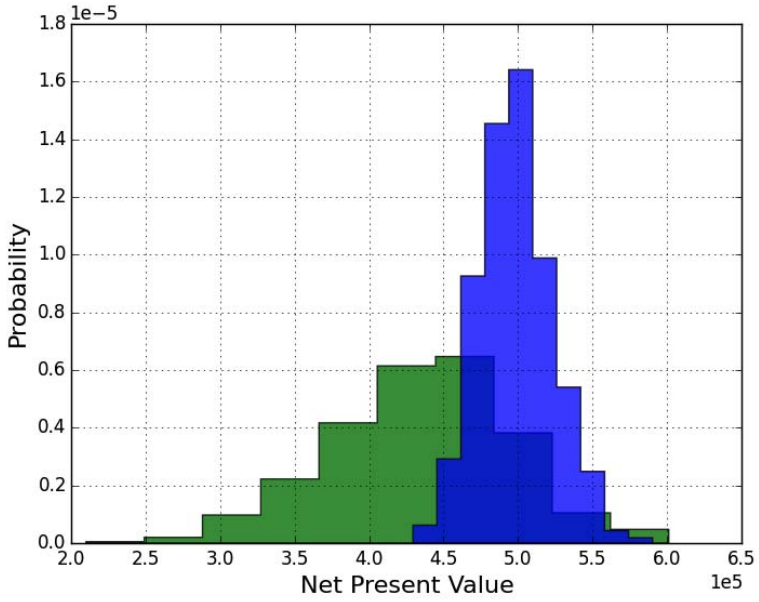

Figure 6.9. Histogram of net present value with 1000 samples. Green histogram line: oil price and permeability are both uncertain; Blue histogram: only oil price is uncertain.

CHAPTER 7. SUMMARY

The purpose of this study is to employ the machinery of BGO for design optimization under uncertainty with limited data-budget in applications of well-placement problem in oil-reservoir modeling. BGO uses GPR to represent our state of knowledge and adaptively select simulation according to extended EI policy.

The mathematic background of GPR and derivation of formula for EI data-acquisition criterion is given in previous chapter. By testing standard functions such has Sasena, Harmant 3 and Harmant 6, proposed extended EI data acquisition is proved can successfully approach to global maximum and minimum when there is noise in observations, which we refer as uncertainty in this study.

Then, exact same method is used to find the best well location under three different circumstances. No uncertainty, uncertainty in oil price and uncertainty in both oil price and permeability. According to results shown in Chapter 6, it is proved that BGO with extended EI policy is successfully applied in design optimization under uncertainty with limited data-budget with applications to the well-placement problem in oil-reservoir modeling by saving cost and improving the computing efficiency.

A big issue neglected in this study is the formulation of stopping criterion, so for future work formulation of stopping criteria for data-acquisition process is worth exploration. With a proper stopping standard, more budgets can be saved and

performance of BGO will be guaranteed. In Sec.6.3, the best well location we finally picked out is the top one in pareto front. The risk for this point is highest, in reality, we may seldom choose. So multiple conflicting optimization problem is also worth exploration.

LIST OF REFERENCES

LIST OF REFERENCES

[1] Aarnes, J. E., Gimse, T., & Lie, K. A. (2007). An introduction to the numerics of flow in porous media using Matlab. In *Geometric modelling, numerical simulation, and optimization* (pp. 265-306). Springer Berlin Heidelberg.

[2] Arbogast, T. (2003). An overview of subgrid upscaling for elliptic problems in mixed form. Contemporary Mathematics, 329, 21-32.

[3] Arbogast, T. (2004). Analysis of a two-scale, locally conservative subgrid upscaling for elliptic problems. SIAM Journal on Numerical Analysis, 42(2), 576-598.

[4] Bangerth, W., Klie, H., Wheeler, M. F., Stoffa, P. L., & Sen, M. K. (2006). On optimization algorithms for the reservoir oil well placement problem.*Computational Geosciences*, *10*(3), 303-319.

[5] Barker, J. W., & Thibeau, S. (1997). A critical review of the use of pseudorelative permeabilities for upscaling. *SPE Reservoir Engineering*, *12*(2), 138-143.

[6] Beckner, B. L., & Xong, X. (1995). *Field development planning using simulated annealing-optimal economic well scheduling and placement* (No. CONF-951002). Society of Petroleum Engineers (SPE), Inc., Dallas, TX (United States).

[7] Bilionis, I., & Zabaras, N. (2012). Multidimensional adaptive relevance vector machines for uncertainty quantification. *SIAM Journal on Scientific Computing*,*34*(6), B881-B908.

[8] Bilionis, I., & Zabaras, N. (2012). Multi-output local Gaussian process regression: Applications to uncertainty quantification. *Journal of Computational Physics*, *231*(17), 5718-5746.

[9] Bilionis, I., Zabaras, N., Konomi, B. A., & Lin, G. (2013). Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification. *Journal of Computational Physics*, *241*, 212-239.

[10] Bilionis, I., & Zabaras, N. (2014). Solution of inverse problems with limited forward solver evaluations: a Bayesian perspective. *Inverse Problems*, *30*(1), 015004.

[11] Bilionis, I. *py-design: Design of Experiments in Python*. 2014; Available from: https://github.com/ebilionis/py-design.

[12] Bilionis, I., Drewniak, B. A., & Constantinescu, E. M. (2015). Crop physiology calibration in the CLM. *Geoscientific Model Development*, *8*(4), 1071-1083.

[13] Bittencourt, A. C., & Horne, R. N. (1997). Reservoir development and design optimization. *paper SPE*, *38895*, 5-8.

[14] Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*,*16*(5), 1190-1208.

[15] Centilmen, A., Ertekin, T., & Grader, A. S. (1999). *Applications of neural-networks in multi-well field development* (Doctoral dissertation, Pennsylvania State University).

[16] Chaudhuri, A., Haftka, R., & Viana, F. (2012, April). Efficient Global Optimization with Adaptive Target for Probability of Targeted Improvement. In*8th AIAA Multidisciplinary Design Optimization Specialist Conference, American Institute of Aeronautics and Astronautics, Honolulu, HI* (pp. 1-13).

[17] Chen, P., Zabaras, N., & Bilionis, I. (2015). Uncertainty propagation using infinite mixture of Gaussian processes and variational Bayesian inference.*Journal of Computational Physics*, *284*, 291-333.

[18] Chen, Z., & Hou, T. (2003). A mixed multiscale finite element method for elliptic problems with oscillating coefficients. *Mathematics of Computation*,*72*(242), 541-576.

[19] Chen, Z., Huan, G., & Ma, Y. (2006). *Computational methods for multiphase flows in porous medi* (Vol. 2). Siam.

[20] Christie, M. A. (1996). Upscaling for reservoir simulation. *Journal of Petroleum Technology*, *48*(11), 1-004.

[21] Christie, M. A., & Blunt, M. J. (2001). Tenth SPE comparative solution project: A comparison of upscaling techniques. *SPE Reservoir Evaluation & Engineering*, *4*(04), 308-317.

[22] Demond, A. H., & Roberts, P. V. (1987). An examination of relative permeability relations for two-phase flow in porous media. *JAWRA Journal of the American Water Resources Association*,*23*(4), 617-628.

[23] Elamvazuthi, I., Vasant, P., & Ganesan, T. (2013). Hybrid optimization techniques for optimization in a fuzzy environment. In *Handbook of Optimization* (pp. 1025-1046). Springer Berlin Heidelberg.

[24] Ghanem, R. G., & Spanos, P. D. (2003). *Stochastic finite elements: a spectral approach*. Courier Corporation.

[25] Güyagüler, B., & Horne, R. N. (2004). Uncertainty assessment of well-placement optimization. *SPE Reservoir Evaluation & Engineering*, *7*(01), 24-32.

[26] Hedar, A. R. (2013). Global optimization test problems. *URL http://www-optima. amp. i. kyoto-u. ac. jp/member/student/hedar/Hedar_files/TestGO. htm*.

[27] Honarpour, M., & Mahmood, S. M. (1988). Relative-permeability measurements: An overview. *Journal of petroleum technology*, *40*(08), 963-966.

[28] Hou, T. Y., & Wu, X. H. (1997). A multiscale finite element method for elliptic problems in composite materials and porous media. *Journal of computational physics*, *134*(1), 169-189

[29] Inman, R. L., Helson, J. C., & Campbell, J. E. (1981). An approach to sensitivity analysis of computer models: Part II-ranking of input variables, response surface validation, distribution effect and technique synopsis. *Journal of Quality Technology*, *13*(4).

[30] Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization,13*(4), 455-492

[31] Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, *21*(4), 345-383.

[32] Lawrence, D. B. (1999). *The economic value of information*. Springer.

[33] Lázaro-Gredilla, M., Quiñonero-Candela, J., Rasmussen, C. E., & Figueiras-Vidal, A. R. (2010). Sparse spectrum Gaussian process regression. *The Journal of Machine Learning Research*, *11*, 1865-1881.

[34] Maini, B. B., & Okazawa, T. (1987). Effects of temperature on heavy oil-water relative permeability of sand. *J. Cdn. Pet. Tech.(May-June 1987)*, 33-41.

[35] Marek, C., & Tomasz, Z. (2011). Mathematics for Finance-An Introduction to Financial Engineering.

[36] Mattax, C. C., & Dalton, R. L. (1990). Reservoir simulation.

[37] Mockus, J., Tiesis, V., & Zilinskas, A. (1978). The application of Bayesian methods for seeking the extremum. *Towards Global Optimization*, *2*(117-129), 2.

[38] Pardalos, P. M., & Resende, M. G. (Eds.). (2001). *Handbook of applied optimization*. Oxford university press.

[39] Powell, W. B., & Ryzhov, I. O. (2012). *Optimal learning* (Vol. 841). John Wiley & Sons.

[40] Pride, S. R., Harris, J. M., Johnson, D. L., Mateeva, A., Nihel, K. T., Nowack, R. L., ... & Fehler, M. (2003). Permeability dependence of seismic amplitudes.*The Leading Edge*, *22*(6), 518-525.

[41] Renard, P., & De Marsily, G. (1997). Calculating equivalent permeability: a review. *Advances in Water Resources*, *20*(5), 253-278.

[42] Smith, T. E., & Dearmon, J. (2014). Gaussian Process Regression and Bayesian Model Averaging: An alternative approach to modeling spatial phenomena.

[43] Spall, J. C. (2005). *Introduction to stochastic search and optimization: estimation, simulation, and control* (Vol. 65). John Wiley & Sons.

[44] Zheng, S. Y., Corbett, P. W., Ryseth, A., & Stewart, G. (2000). Uncertainty in well test and core permeability analysis: a case study in fluvial channel reservoirs, northern North Sea, Norway. *AAPG bulletin*, *84*(12), 1929-1954.