ENOS
Enabling Onshore CO_2 Storage

A CO_2GeoNet Initiative

ENOS D2.4 | March 2020

Smart characterization of CO_2 storage site using a Bayesian approach

Date	27.03.2020
Author(s)	Alexander Tarakanov, Ahmed H. Elsheikh, Heriot-Watt University, UK

Number of pages	38
Number of appendices	0
Customer	EC
Project name	ENOS
Project number	GA No 653718

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 653718.

Table of Contents

1 Executive Summary

Global climate change has been indicated as one of the most significant challenges to be addressed by humanity. Given the variety of depleted oil and gas fields and other underground reservoirs, CO_2 sequestration is considered as one of the most perspective techniques for reducing of greenhouse gases concentration in the atmosphere. However, the nature of subsurface storage and the high cost of measurements makes it difficult to provide a reliable capacity estimate of potential storage site and further quantify the amount of CO_2 that could be safely stored. Therefore, accurate capacity estimation of a potential CO_2 sequestration site is one of the most significant tasks that needs to be addressed for a practical large-scale implementation of CO_2 underground storage.

The accuracy of CO_2 storage capacity estimation and reliable quantification of risk is highly dependent on amount of available data and thus additional data collection might be needed for most sites which is an expensive procedure for subsurface systems. Therefore, optimization procedure could be utilized to balance the cost of measurements and additional insights provided by additional collected data. In the present work, the issue concerned is addressed via Bayesian experimental design technique. These approaches provides solid mathematical formulation of the optimal data acquisition problem.

In the context of the ENOS project funded by the H2020 European program and dedicated to the enabling of On Shore CO_2 storage, the present work is proposed to developed a smart characterization method of data acquisition and used during the elaboration of a CO_2 injection storage pilot or demonstration site. This report corresponds to the deliverable D2.4 for the task 2.2 of the Work Package 2 of the ENOS project.

In a Bayesian experimental design framework, the amount of information provided by a given measurement is quantified by comparing the probability distribution of model parameters that provide reasonable explanation of the observed data with the prior distribution of model parameters. However, the results of measurements cannot be known before those measurements are collected. Therefore, averaging over the space of possible measurements is performed in order to compute the average or expected amount of information gained from the new observations. This procedure seems to be complicated at the first sight. However, there exist a variety of numerical methods for calculation of the quantities concerned, for instance Markov Chain Monte Carlo methods (MCMC). Therefore, statistical theory provides the tool for quantification of the data insights together with algorithms for numerical evaluation.

The major challenge in the practical implementation of Bayesian experimental design is the associated numerical costs of averaging (i.e. integration) over the possible space of measurements. The latter is critical for subsurface problems when flow simulations are required for single function evaluation. Here we propose to tackle this problem by replacing the expensive numerical flow simulations by a surrogate model. In the present work, this approach concerned is adopted and Polynomial Chaos Expansion (PCE) is utilized for development of the response surface.

Utilization of response surfaces provide a significant speed-up of single expected information gain calculation. However, in the majority of practical problems, the ultimate goal is optimizing the utility function rather than performing a single value evaluation. Therefore, the acceleration provided by replacing the flow simulations with a cheap to evaluate surrogate model is not enough to solve the optimal data gathering problem in a reasonable time. We propose a novel approach to address this issue, where we construct a PCE response surface for the expected information gain without directly evaluating utility function itself.

This novel approach to calculating the expected information gain makes it possible to construct a response surface for the utility function at a minimal cost of (<100 function evaluations). The response surface developed can be utilized to determine the parameters of the optimal experimental design.

The obvious advantage of the proposed methodology is low computational cost of the solution for overall optimization problem if compared with direct approach. However, certain constraints that limit the applicability of proposed method should be satisfied. First of all, it is supposed that measurement errors are independent and normally distributed. These constraints impose limitations on the applicability of the method. However, the assumptions made are not too restrictive and are satisfied in many practical problems. Therefore, the proposed approach to Bayesian experimental design provides a robust technique for optimization of data acquisition processes. The methodological guidelines for implementing the proposed technique for optimizing the data acquisition process is provided in the numerical examples section of this report.

2 Introduction

2.1 Background

In recent decades the rapid growth of the temperature that does not fit in with regular climate behaviour has been observed [IPCC, 2007]. The reason for the effect concerned is dramatic growth of greenhouse gases concentration in the Earth atmosphere [IPCC, 2007]. The scientific community is confident that the concentration growth of observed CO_2 is caused by human activities, and CO_2 emissions make a significant contribution to greenhouse gases concentration [Bryant, 1997].

The surface temperature growth could have far reaching consequences to all humankind. The latter includes the increase of sea-level followed by flooding of coastal regions [IPCC, 2007], forcing significant fraction of Earth population to migrate towards interior parts of the continents. In addition to that, temperature growth or global warming could have significant impact on the distribution of precipitations [IPCC, 2007]. According to the best knowledge of climate mechanisms, such redistribution could make vast regions of the planet too arid or too wet and inhabitant. Those consequences of global warming could lead to tremendous social tension and conflicts and even question the future of the mankind [IPCC, 2007]. Given the scale of consequences of climate change, it is reasonable to take certain measures in order to prevent or reduce the impact of global warming. Moreover, it has been indicated that the sooner those measures are taken, the less significant would be the impact of the climate change [AGU].

As an immediate response to global temperature growth, subsurface CO_2 injection and long-term storage is considered as one of the potential solutions to global warming. The central idea of the approach is to inject CO_2 from atmosphere or any other sources into subsurface reservoirs that can securely store a given amount of CO_2 at the timescale of hundreds or thousands of years.

The constraint on the timescale of secure CO_2 storage limits the set of possible candidates. However, plenty of subsurface reservoir that have a potential to store injected CO_2 for thousands of years are available all over the world. First of all, there exist a variety of natural reservoirs that have been storing CO_2 for geological times (millions of years). Secondly, there is a huge number of depleted oil and gas reservoirs that have trapped hydrocarbons for thousands of years. Therefore, such reservoirs could trap CO_2 and prevent it from leaking due to the properties of the sealing cap rock. Finally, saline water aquifers can satisfy the timescale requirements in the presence of appropriate trapping mechanism. Given the availability of well-studied subsurface oil gas and water reservoir, it is possible to successfully implement CO_2 sequestration technology.

There exist a variety of injection and storage strategies for CO_2 sequestration. The most common one is to inject CO_2 as a supercritical fluid. That allows one to store large amounts of CO_2 due to high density at that case. At the same time, CO_2 occupies all available pore volume under super critical conditions like a normal gas. Therefore, the combination of such properties at supercritical state is beneficial for Carbon Capture and Storage (CCS) operations.

Since CO_2 has been injected into a reservoir various physical and chemical processes start to affect the distribution of the injected CO_2 inside the reservoir rock. Those processes determine trapping mechanisms that prevent CO_2 from escaping from the reservoir.

At the early stages of the injection, the flow of CO_2 is primarily driven by pressure gradients and gravity. As a result, CO_2 migrates towards the cap rock, where it should be sealed by impermeable reservoir rocks. This mechanism is essentially the same as one that keep hydrocarbons confined within an oil or gas reservoirs and is referred to as structural or stratigraphical trapping. At the very beginning of CO_2 injection, this trapping mechanism determines the CO_2 storage capacity.

During the injection, CO_2 displaces water inside the reservoir pore space. After stopping the injection, water starts to occupy the left pore space in the imbibition-like process. In such case, a certain fraction of CO_2 is trapped by capillary forces. This mechanism is referred to as residual trapping. The efficiency of this process is largely determined by the chemical composition of subsurface water and reservoir rocks. It is clear that the mechanism concerned has larger time scale in comparison to structural and stratigraphical trapping and that it starts to affect the distribution of CO_2 only after the injection has stopped.

Structural and residual trapping are known as physical trapping mechanisms, because their efficiency is determined by physical effects like buoyancy and capillary forces. Various chemical processes contribute to CO_2 storage capacity as well. For instance, dissolution of CO_2 in brine or absorption by reservoir rocks effectively capture and immobilize injected CO_2. This mechanism is referred to as solubility trapping. In addition to that, injected CO_2 could react chemically with minerals that form solid matrix of the reservoir. As a result of the reactions, CO_2 precipitates as solid carbonate mineral. This mechanism provide the most secure form of the CO2 storage.

The operational time scale varies significantly over the trapping mechanisms. For example, during the injection of CO_2, structural and stratigraphical trapping is the primary mechanism that determines the capacity of the storage site. The time scale of this process ranges from years to decades, which is a typical lifetime of an active injection phase. After the termination of the injection, residual trapping mechanism affects the CO_2 distribution at the timescale of dozens of years. At the time scale of centuries, significant amount of CO_2 is could be dissolved in subsurface fluids and brine. Finally, at the time scale of thousands of years, chemical reaction with reservoir rocks immobilize significant amount of injected CO_2. The contribution of each of these mechanisms to overall storage capacity at different moments of time is demonstrated on Figure 1a [Bachu, 2008].

It has been mentioned that CO_2 trapping mechanisms vary significantly in terms of timescale. In addition to that, maximal amount of CO_2 that can be stored varies significantly among the trapping mechanism. Typically, slower processes have higher potential capacity as it can be seen on Figure 1b [Bachu, 2008]. However, the exact magnitude of the CO_2 amount that can be potentially stored is highly sensitive to the geometry of the reservoir, chemical composition of reservoir brine and reservoir rocks and other factors [Bachu, 2008]. It is important to note that the overall capacity is determined mainly by the structural or stratigraphical trapping mechanism [Bachu, 2008]. due to high contrast in speed of physical

and chemical processes that determine CO_2 distribution in the reservoir. However, the interplay between different mechanisms reduces the possible risks of CO_2 leakage over time due to dissolution of mobile CO_2 in subsurface fluid or due to chemical reactions with minerals of surrounding rocks.

The variety of candidate sites for CO_2 sequestration, availability of technology for injection of CO_2 make CO_2 sequestration one of the most powerful tools for mitigating climate change effects. However, there exist certain risks for human population and environment associated with subsurface CO_2 storage.

The major risk related to CO_2 sequestration projects is associated with leakage of CO_2 back

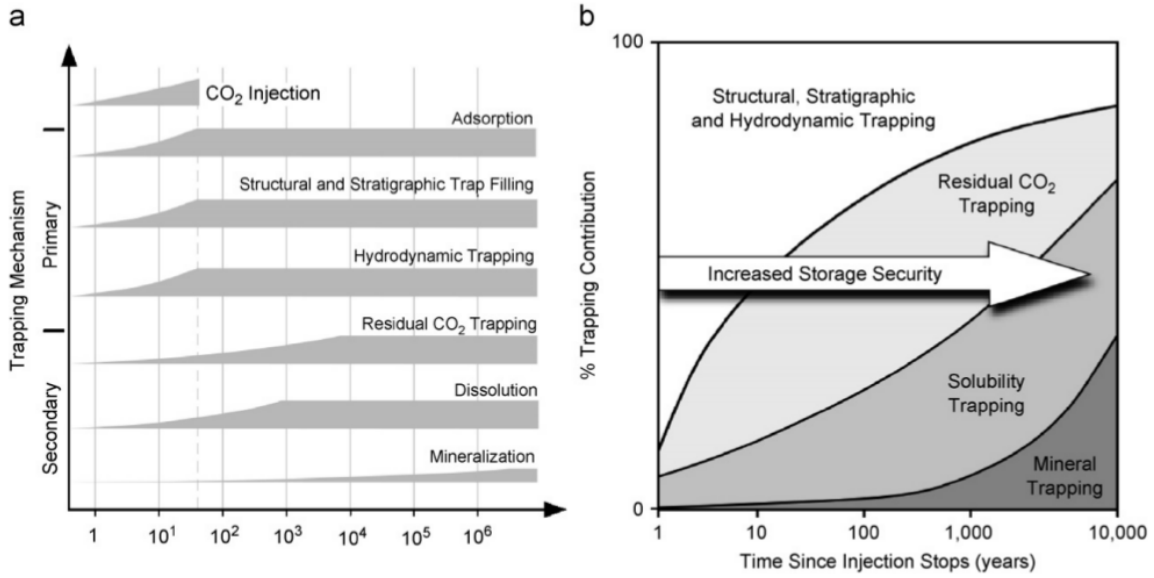

Figure 1: Time scale for different trapping mechanisms (a) and fractions of trapping mechanisms in the overall storage capacity (b) [Bachu, 2008].

to the atmosphere. First of all, the leakage of CO_2 can devalue the effort put in the injection of the CO_2 and make the whole project meaningless. Secondly, leakage from CO_2 storage could have far reaching consequences for people and ecology. For instance, massive CO_2 leakage could result in significant concentrations of CO_2 on the surface and lead even to death of small animals and people due to the asphyxia [Bachu, 2008]. Minor but continuous leakage of CO_2 can significantly change the pH of the underground water and cause an increase of concentration of different heavy metals in the water, making it impossible to utilize that source of water for human needs. Therefore, mitigating potential leakages from CO_2 storage and careful observation of the storage state after the injection of CO_2 are of primary importance to CCS projects.

One of the most common leakage mechanisms is CO_2 migration through fractures in the cap rock. Those fractures can occur because of high pressure of the injected CO_2 or due to chemical reactions that lead to either increase of the pressure of the injected fluid or to weakening of the sealing cap rock [Miocic et al., 2013]. The alternative source of leakage is represented by abandoned wells when storing CO_2 is in depleted hydrocarbon reservoirs. Such leakage could be caused by fractures in the cap rock that develop during the drilling or through the fractures in the cemented region around the well. Those risks could be accessed

through careful analysis of the available data prior to CO_2 injection and by regular monitoring of the reservoir state during the lifetime of the CCS project [Furre et al, 2017].

In summary, CO_2 sequestration can be considered as one of the most powerful tools to address the climate change issues in a timely manner. Numerous depleted oil field and saline aquifers all over the world could be utilized for that purposes. Therefore, significant amount of atmospheric CO_2 can be stored in subsurface reservoirs. However, there exist certain risks associated with CO_2 sequestration that should be carefully accessed for success of CO_2 sequestration projects.

2.2 Smart site characterization

Given the high overall cost of the industrial scale CO_2 sequestration and potential hazards of CO_2 leakage, accurate estimate for storage capacity and risk assessment is required. Different studies indicate that this issue can be addressed mainly via refining the knowledge about the reservoir by collecting data from various sources [CO2CRC, 2008; DNV, 2009; NETL, 2010; SiteChar, 2013; Bachu, 2015; Heidug, 2013]. For instance, in the case of depleted oil and gas fields well-log seismic and production data are available along with the reservoir model. However, possible chemical reactions of CO_2 with reservoir fluids and rocks should be analyzed along with the state of the abandon wells and condition of the cap rock [Bachu 2008]. Moreover, additional measurements of fluid and rock properties could be required in order to provide accurate estimate for CO_2 storage capacity. The latter can be illustrated by the Sleipner project, where the CO_2 distribution caused by residual trapping mechanism was not fully understood for several decades [Anne-Kari Furre et al 2017]. In the case of the CO_2 sequestration in the saline aquifers significant amount of measurements should be conducted in order to build accurate reservoir model [Bachu 2008]. The latter requires drilling of exploratory wells and seismic measurements. Therefore, data acquisition is an essential step in the site characterization for CO_2 sequestration.

Variety of methods are available for gathering information about subsurface reservoir. The latter include, drilling of new wells, well test measurements from existing wells, seismic surveys with different resolutions among other sources of data. All of those techniques vary significantly in cost, duration of the measurements and amount of the information provided. Therefore, in order to maximize the information that can be obtained using that various of tools under a given set of financial constraints, the utilization of an optimization-based approach is required. In other words, the problem of data acquisition and site characterization can be naturally transferred to an optimization problem of the experimental design.

In the present work Bayesian optimal experimental design is utilized to optimize data acquisition plans. Bayesian experimental design is a probabilistic approach for optimization of the objective function related to the information about any given physical system. For example, Bayesian experimental design has been successfully applied to minimization of variance of model parameters or predicted values and to the maximization of the information gain itself. The main advantage of the approach is that it provides a streamlined workflow for combining various sources of information within a single optimization routine. In other words,

the method is generic and flexible enough to be easily adopted to almost any data acquisition framework.

2.3 The scope of the project

The main challenge in utilization of Bayesian experimental design is associated with high computational cost. Even a single calculation of the objective function could be quite expensive due to the high dimensional integrals that are the essence of the method concerned. Therefore, the main focus of the present work is to develop efficient numerical schemes with low run-time to optimize data acquisition plans. This issue is addressed through the utilization of surrogate modelling techniques.

First of all, Polynomial Chaos Expansion (PCE) is utilized in order to replace the flow simulators used to estimate the gained knowledge given an observations. That measure allows one to significantly reduce the time needed for objective function evaluation, which is in the same line with the latest research papers on that subject. The construction of PCE surrogate is explained in the Section 4.

In addition to the PCE surrogate for flow simulations, PCE based response surface has been developed to fit the objective function directly. This response surface is constructed with a novel technique that derives the response surface for the objective function from the values of the related quantities. In other words, the response surface for the objective function is constructed without even a single computation of the objective function. The overall computational cost of response surface construction is comparable with dozens of single value calculations. The latter allows one to solve the optimize the data acquisition process at the cost of few function evaluations which is orders of magnitude faster that any direct approach. The latter is the principal result of the project. The details of the novel approach are explained in the Section 5.

Finally, numerical examples that demonstrate the overall workflow of data acquisition optimization for subsurface flow problems is presented in the Section 6.

3 Bayesian experimental design

3.1 Overview

Inverse problems have been playing significant role in engineering and research in the last 50 years. The ultimate goal in such type of problems is to get certain insights about any given system based on the observed data or measurements. Commonly, it is supposed that the given system is approximated by a mathematical model with finite number of parameters. In addition to that, a functional relation between model parameters θ and results of measurements $\boldsymbol{m}$ is formulated as:

$$\mathrm{m} = f(\theta, \mathrm{d}) + \eta \tag{1}$$

Where $\boldsymbol{d}$ is the set of parameters that determines the experimental setup for data collection and η represents the measurements noise. Equation (1) implies that modeled system can be approximated with finite number of parameters. Such constraint does not limit applicability of the approach in the case of subsurface flow problems, where the spatial distribution of porosity and permeability are unknown. For most of practical problems, the distribution of porosity and permeability admits reasonable finite-dimensional approximation via Karhunen-Loeve (KL) expansion. Therefore, all the machinery developed for the systems described by Eq. (1) can be fully applied to subsurface flow problems and to optimization of data acquisition for CO_2 sequestration.

Methods for solution of inverse problem as formulated by Eq. (1) have a long history of development. The solution of the inverse problem heavily relies on the adopted approaches for data gathering. A variety of techniques for experimental design have emerged to address different challenges related to data collection. Historically, the first issue that have been addressed is the appropriate coverage of the design parameter space in order to guarantee that models with different parameters can be distinguished by collected data up to a specified tolerance. The most common approach that addresses this challenge is Full Factorial Design [Montgomery et al., 2012]. That method allows one to infer model parameters with reasonable accuracy. However, the cost of the method grows relatively fast with the problem dimension. In order to tackle the dimensionality issues, methods like Fractional Factorial Design [Montgomery et al., 2012], Latin Hypercube [McKay et al., 1979], Box-Brenken [Box et al., 1963] and Plackett-Burman [Plackett et al., 1946] designs have been proposed. Those methods significantly reduce the overall costs of model parameters inference without essential loss in accuracy when compared to Full Factorial Design. However, these methods of experimental design do not provide optimal scheme of measurements. Therefore, the accuracy of the inferred model parameters could be further improved by optimizing the experimental design $\boldsymbol{d}$.

One of the natural techniques towards optimization of experimental design is to minimize the variance of inferred model parameters or model predictions at any given set of points [Vanlier et al., 2012]. These methods can be considered as particular case of Bayesian experimental design when the model parameters or model predictions are assumed to be normally distributed. The latter typically holds for accurate measurements that allow to infer model parameters with high precision [Miler et al., 1995]. Unfortunately, high-precision

measurements are not always available especially in the case of subsurface flow problems where both the cost of data gathering and dimension of model parameters space is high. Therefore, more generic and robust Bayesian approach is utilized as more appropriate tool for reaching the goals of the ENOS project.

There exist a variety of Bayesian experimental design schemes depending on the objective function of concern [Mosbach et al., 2012; Roshan, 2012]. In the present report, D-optimality criterion is considered. In that approach, the amount of information regarding the model parameters is maximized by tuning the design parameters $\boldsymbol{d}$. More precisely, in any single experiment the information from all the measurements is quantified through the difference between the prior distribution of model parameters and the posterior distribution of the model parameters conditioned to the given observations $\boldsymbol{m}$ and design $\boldsymbol{d}$. However, the actual value of observed data m is not known before the experiment is conducted. Therefore, different experimental schemes are ranked by the objective function in D-optimal experimental design, which is computed as an average value of information gain over the space of possible experimental outcomes. The mathematical details of this idea are summarized below.

In a Bayesian setting, the model parameters are distributed in agreement with a prior distribution:

$$p(\theta) = p(\theta|\mathbf{d}) \tag{2}$$

Where $p(\theta|\boldsymbol{d})$ and $p(\theta)$ are probability density distributions with respect to the model parameters θ. In the present work, it is supposed that prior distribution does not depend on parameters of experimental design, which is reflected in the Eq. (2). Typically, one of the tabular distributions (aka. uniform or normal) is utilized as a prior distribution. In the present work, the reparametrization of the model parameters is utilized in order to work with variables with a well-known probability distribution. Given the prior distribution of model parameters and observation $\boldsymbol{m}$, Bayesian theorem for conditional probabilities can be utilized in order to derive the expression for the posterior distribution of the model parameters:

$$p(\mathbf{m}|\theta, \mathbf{d})p(\theta|\mathbf{d}) = p(\theta|\mathbf{m}, \mathbf{d})p(\mathbf{m}|\mathbf{d}) \tag{3}$$

Where $p(\theta|\boldsymbol{m}, \boldsymbol{d})$ is the density of the posterior distribution, $p(\boldsymbol{m}|\boldsymbol{d})$ is the density distribution of observations $\boldsymbol{m}$ for a given experimental design $\boldsymbol{d}$ and $p(\boldsymbol{m}|\theta, \boldsymbol{d})$ is the density of probability distribution of observing m for a given model parameters θ and design $\boldsymbol{d}$ or likelihood. It is natural to assume that the difference between model predictions and observed data follows a normal distribution:

$$p(\mathbf{m}|\theta, \mathbf{d}) = \frac{1}{(2\pi\sigma^2)^{\dim(\mathbf{m})/2}} \exp\left(- \frac{(\mathbf{m} - f(\theta, \mathbf{d}))^2}{2\sigma^2} \right) \tag{4}$$

Where σ is the standard deviation and $\dim(\boldsymbol{m})$ is the dimension of the observed data. Given the form of prior and the likelihood, the equation for $p(\boldsymbol{m}|\boldsymbol{d})$ can be derived as a normalization factor for the integral of prior distribution and likelihood:

$$p(\mathbf{m}|\mathbf{d}) = \int p(\theta|\mathbf{m},\mathbf{d})p(\mathbf{m}|\mathbf{d})d\theta = \int p(\mathbf{m}|\theta,\mathbf{d})p(\theta|\mathbf{d})d\theta \quad (5)$$

Therefore, all the necessary ingredients are available in order to estimate the posterior distribution using:

$$p(\theta|\mathbf{m},\mathbf{d}) = \frac{p(\mathbf{m}|\theta,\mathbf{d})p(\theta|\mathbf{d})}{p(\mathbf{m}|\mathbf{d})} \quad (6)$$

The information inferred about the model parameters is quantified via Kullback-Leibler divergence [Kullback et al., 1951] between prior and posterior distributions of the model parameters [Huan et al., 2013]:

$$D_{KL}(\mathbf{m},\mathbf{d}) = \int p(\theta|\mathbf{m},\mathbf{d}) \log\left(\frac{p(\theta|\mathbf{m},\mathbf{d})}{p(\theta|\mathbf{d})}\right) d\theta \quad (7)$$

Finally, averaging with respect to all possible experimental results is computed:

$$U(\mathbf{d}) = \int D_{KL}(\mathbf{m},\mathbf{d})p(\mathbf{m}|\mathbf{d})d\mathbf{m} \quad (8)$$

Where $U(\boldsymbol{d})$ is the expected information gain or the utility function that is maximized with respect to $\boldsymbol{d}$ in order to determine the parameters of the experiment as following:

$$\mathbf{d}^* = \underset{\mathbf{d}}{\operatorname{argmin}}\big(U(\mathbf{d})\big) \quad (9)$$

3.2 Numerical integration and Bayesian experimental design

In optimal Bayesian experimental design, the expected information gain $U(\boldsymbol{d})$ is maximized in order to determine the optimal experimental setup. According to Eq. (5) and Eq. (8), the calculation of a single value of the utility function requires sequential multiple integrations. First of all, $p(\boldsymbol{m}|\boldsymbol{d})$ requires integration over the space of model parameters θ, for instance, like in the Eq. (5). Secondly, $U(\boldsymbol{d})$ is computed as the integral of KL-divergence over the space of possible experimental outcomes m with respect to probability distribution $p(\boldsymbol{m}|\boldsymbol{d})$. In certain cases, it possible to compute the integrals concerned analytically [Roshan, 2012]. However, for generic cases, the functions involved in the integration could be highly non-linear. Therefore, only numerical integration is feasible.

There exist a variety of techniques for numerical calculation of $p(\boldsymbol{m}|\boldsymbol{d})$ and $U(\boldsymbol{d})$ [Zhang et al., 2016]. Most common are approximations of the probability distribution of interest by normal distribution [Mosbach et al., 2012]. However, such methods have limited applicability because of strong underlying assumptions about the probability distribution [Miler., 1995]. In the case of subsurface systems, it is quite challenging to justify that the probability distribution of quantities of interest are close to the normal distribution. Therefore, universal Markov Chain Monte Carlo (MCMC) integration technique is utilized in order to enhance the robustness of the approach and to expand the range of problems that can be solved.

In MCMC methods, the integral of a given function $g(\boldsymbol{x})$ is approximated as the mean of functions values at finite number of points called Markov Chain:

$$\int g(\mathrm{x})p(\mathrm{x})d\mathrm{x} = \sum_i \frac{1}{N} g(\mathrm{x}_i) \tag{10}$$

Where $p(\boldsymbol{x})$ is a certain probability distribution and $g(\mathrm{x})$ is integrable function, $\boldsymbol{x}_i$ are points in $\mathbb{R}^n$and i is the index that ranges from 1 to the total number of sampled points N. The essence of MCMC techniques is sampling of points x_i in order to provide accurate estimate of the integral in Eq. (10).

There exist a variety of techniques for generating the set of $\boldsymbol{x}_i$, like Metropolis-Hastings algorithm [Hastings, 1970], Gibbs sampling technique [Gelfand, 1990] and Hamiltonian MCMC [Neal, 2011]. In the present work, Metropolis-Hastings approach is utilized as it provide a good balance between robustness and complexity of implementation. In Metropolis-Hastings algorithm, the set of $\boldsymbol{x}_i$ is generated sequentially via a Markov process [Gelman et al., 2014]. The latter is a two-step procedure of generating a proposal for x_{i+1} followed by an acceptance or rejection step of the proposed point. This procedure is entirely local and is determined by the current state $\boldsymbol{x}_i$. At the first step, a value for x_{i+1} is proposed via sampling of the displacement $\delta\boldsymbol{x}$:

$$\mathrm{x}' = \mathrm{x}_i + \delta\mathrm{x} \tag{11}$$

Typically, that shift $\delta\boldsymbol{x}$ of $\boldsymbol{x}_i$ is sampled from a tabular distribution. In the present work, a normal distribution $\mathcal{N}(0, \sigma_{\mathrm{MH}})$ with zero mean and standard deviation σ_{MH} is utilized for generation of $\delta\boldsymbol{x}$. The parameter σ_{MH} is tuned in agreement with best practices of MCMC applications [Gelman et al., 2014].

Since the proposal has been generated, values of the probability density distributions are computed at the proposed $p(\boldsymbol{x}')$ and the initial point $p(\boldsymbol{x}_i)$. The decision regarding the acceptance or rejection of a new sample is made based on relative value of those values:

$$r = \frac{p(\mathrm{x}')}{p(\mathrm{x_i})} \tag{12}$$

Where r denotes the ratio of the probability density functions. If $r \geq 1$, then the proposed sample is accepted, and generated sample is added to the chain $\boldsymbol{x}_{i+1} = \boldsymbol{x}'$. Otherwise, a random real number ζ is sampled from the uniform distribution on the unit interval $\boldsymbol{\mathcal{U}}[0; 1]$. If $\zeta \leq r$ then the proposed sample is accepted and $\boldsymbol{x}_{i+1} = \boldsymbol{x}'$. Otherwise, the sample is rejected and the current point x_i is added to the chain $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i$. This step can be schematically represented as follows:

$$\mathrm{x}_{i+1} = \begin{cases} \mathrm{x}', & \text{if } r \geq 1 \\ \mathrm{x}', & \text{if } \zeta \leq r < 1 \\ \mathrm{x}_i, & \text{if } r < \zeta \leq 1 \end{cases} \tag{13}$$

It is worth to mention that generating a Markov Chain with the Metropolis-Hastings algorithm only requires the ratio of the probability density functions. Therefore, the algorithm concerned can be naturally utilized for unnormalized distributions.

The initial point $\boldsymbol{x}_1$ can be sampled randomly from any reasonable distribution. In order to exclude dependence on the initial state of the Markov Chain, first several elements of the chain are not considered in the calculation of the integral Eq. (10). This part of the MCMC chain is referred to as burn-in phase [Gelman et. al., 2014]. In the present work, 20% of the chain is utilized as a burn-in and the integral in Eq. (10) is approximated with the remaining part of the chain.

There exist several numerical techniques that utilize MCMC for calculating the Bayesian evidence factor Eq. (5) and the expected information gain Eq. (8). The most common MCMC method for computation of Bayesian evidence factor utilize the following representation of the $p(\boldsymbol{m}|\boldsymbol{d})$:

$$p(\mathbf{m}|\mathbf{d}) = \int p(\mathbf{m}|\theta,\mathbf{d})p(\theta|\mathbf{d})d\theta = \frac{1}{N}\sum_i p(\mathbf{m}|\theta_i,\mathbf{d}) \tag{14}$$

Here, the Markov Chain of the model parameters is sampled from the prior distribution $p(\theta|\boldsymbol{d})$. The advantage of this technique is simplicity. However, there is a possibility of convergence issues if the posterior distribution is very narrow [Schniger, 2014]. The issue with narrow posterior distribution is tackled with an alternative representation of Bayesian evidence factor:

$$\frac{1}{p(\mathbf{m}|\mathbf{d})} = \int \frac{p(\theta|\mathbf{d})}{p(\mathbf{m}|\mathbf{d})}d\theta = \int \frac{p(\theta|\mathbf{m},\mathbf{d})}{p(\mathbf{m}|\theta,\mathbf{d})} = \frac{1}{N}\sum_i \frac{1}{p(\mathbf{m}|\theta_i,\mathbf{d})} \tag{15}$$

Where the samples θ_i are generated from the posterior distribution. The advantage of this approach is the sensitivity to regions with high density of posterior distribution, that make the most significant contribution to the integral in the Eq. (15). However, certain numerical issues arise from samples that are sampled from regions with low density of the likelihood. The latter could cause significant bias in MCMC estimate for Bayesian evidence factor $p(\boldsymbol{m}|\boldsymbol{d})$ [Schniger, 2014].

The approach for calculation of the Bayesian evidence factor as done in Eq. (14) is transferred directly to calculating the objective function $U(\boldsymbol{d})$. In this case, elements of MCMC chain are generated from $p(\boldsymbol{m}|\boldsymbol{d})$, which is computed directly with a MCMC at each acceptance/rejection step. Therefore, $U(\boldsymbol{d})$ is approximated as follows:

$$U(\mathbf{d}) = \frac{1}{N}\sum_i D_{KL}(\mathbf{m}_i,\mathbf{d}) \tag{16}$$

Where N is the length of MCMC chain and it is tuned in order to achieve the required degree of accuracy. It is well-known that MCMC estimate for the integral in the Eq. (10) convergence to the exact value as $1/\sqrt{N}$:

$$\left| \frac{1}{N} \sum_{i} g(\mathbf{x_i}) - \int g(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \right| \tag{17}$$

This convergence rate is not very high. Therefore, sufficiently long Markov Chains should be utilized in order to meet accuracy requirements. Numerical experiments demonstrate that Markov Chains of length from 10,000 to 50,000 provide sufficient accuracy for calculating both the expected information gain $U(\boldsymbol{d})$ and Bayesian experimental design $p(\boldsymbol{m}|\boldsymbol{d})$. Therefore, the calculation of single value of the objective function requires at least 10,000 evaluations of $p(\boldsymbol{m}|\boldsymbol{d})$ and each of those values requires at least 10,000 evaluations of the model functions $f(\theta, \boldsymbol{d})$, which requires a run of a flow simulator. Therefore, the computation of a single value of objective function is a computationally expensive procedure. Moreover, the general optimization of $U(\boldsymbol{d})$ is not feasible. In order to tackle that issue, surrogate models are utilized to replace flow simulator.

In addition to that, we propose to develop a Polynomial Chaos Expansion (PCE) based response surface for the objective function $U(\boldsymbol{d})$ directly. This novel approach admits further reduction of the computational cost by the construction of an accurate surrogate model for $U(\boldsymbol{d})$ at the cost of several dozens of function evaluation. The details of this novel approach are described in Section 5.

4 Polynomial Chaos Expansion

4.1 Introduction

PCE is a popular technique for surrogate modelling and function approximation as a series of orthogonal polynomials (e.g. Hermite, Legendre, Chebyshev and etc.) [Weiner, 1938]. The method has a high convergence rate especially when approximating smooth functions. In addition to that, statistical analysis of the data can be significantly simplified if PCE surrogate model has been developed [Blatman et al., 2011]. For instance, mean, variance and Sobol' sensitivity indices can be computed directly from the PCE coefficients [Kaintura et al., 2018]. Such properties of PCE made it quite popular for sensitivity analysis and uncertainty quantification. In the case of data acquisition optimization, the latter features of PCE provide a compelling reason for the utilization of PCE in this project.

Remarkable properties of PCE for statistical analysis and uncertainty quantification are attributed by the relation between family of orthogonal polynomials and the statistics of the input data. In PCE framework, it is assumed that input parameters x are distributed according to probability distribution with density $\mathcal{K}(\boldsymbol{x})$. Therefore, the inner product of two square-integrable functions $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ can be naturally introduced as:

$$\langle f_1, f_2 \rangle = \int f_1(\mathrm{x}) f_2(\mathrm{x}) \mathcal{K}(\mathrm{x}) d\mathrm{x} \tag{18}$$

Given the density of probability distribution $\mathcal{K}(\boldsymbol{x})$ and associated inner product, the family of orthogonal polynomials can be constructed via Gram-Schmidt orthogonalization process [Abramowitz, 1964] applied to the polynomial function of the following form:

$$p^*_{\alpha_1,\ldots,\alpha_n}(x_1,\ldots,x_n) = x_1^{\alpha_1}\ldots x_n^{\alpha_n} \tag{19}$$

Where $\alpha_1, \ldots, \alpha_n$ are non-negative integer numbers, $x_1, \ldots, x_n$ are components of the n-dimensional vector $\boldsymbol{x}$. Therefore, for a family of orthogonal polynomials can be naturally associated with a probability distribution. Hence, PCE series can be formed for any given square-integrable function $f(\boldsymbol{x})$:

$$f(\mathrm{x}) = \sum_A c_A p_A(\mathrm{x}) \tag{20}$$

where A is an index of PCE or polynomial basis function $p_A(\boldsymbol{x})$ and c_A is the corresponding expansion coefficient. Due to utilization of Gram-Schmidt orthogonalization process, PCE basis functions $p_A(\boldsymbol{x})$ satisfy the following constraint:

$$\langle p_A, p_B \rangle = \int p_A(\mathrm{x}) p_B(\mathrm{x}) \mathcal{K}(\mathrm{x}) d\mathrm{x} = \delta_{AB} \langle p_A, p_A \rangle \tag{21}$$

Where A and B are multi-indices of PCE basis functions and $\delta_{AB}$is a Kronecker symbol. In the present work, the following normalization of PCE basis functions is utilized:

$$\langle p_A, p_B \rangle = \delta_{AB} \tag{22}$$

Gram-Schmidt orthogonalization process allows one to develop PCE formalism for almost any arbitrary probability distribution or kernel function $\mathcal{K}(\boldsymbol{x})$ with finite moments of any order. However, in most applications certain assumptions regarding the statistics of the input data are made. The most common one is that the components $x_1, \dots, x_n$ of the input parameters vector $\boldsymbol{x}$ are mutually independent. This assumption does not limit the applicability of PCE, because different transformation technique could be utilized to map the data to a space with mutually independent parameters [Rosenblatt, 1952]. Moreover, it is possible to adjust the transformation in such a way that independent parameters have a tabular distribution. Therefore, in the present work, only uniform distribution $\mathcal{U}[-1; 1]$ are considered. In that case, the density of probability distribution is expressed as:

$$\mathcal{K}(\mathbf{x}) = \mathcal{K}_1(x_1)...\mathcal{K}_n(x_n) \tag{23}$$

Where $\mathcal{K}_a(x_a)$ is a density function of the probability distribution for x_a (uniform or normal distribution). It can be shown that in such setting, polynomial basis functions are expressed as products of single-variate PCE basis functions constructed for each of the factors in Eq. (18):

$$p_A(\mathbf{x}) = p_{\alpha_1}^{(1)}(x_1)...p_{\alpha_n}^{(n)}(x_n) \tag{24}$$

Here $A = \{\alpha_1, \dots, \alpha_n\}$ is a multi-index of the basis function, $p_{\alpha_k}^{(k)}(x_k)$ is PCE basis function of degree α_k associated with the kernel $\mathcal{K}_k(x_k)$. It is simple to show that $p_{\alpha_k}^{(k)}(x_k)$ are Legendre polynomials for $\mathcal{U}[-1; 1]$ and Hermite probabilistic polynomials for $\mathcal{N}(0; 1)$ [Abramowitz, 1964]. Therefore, in the present work products of Legendre and Hermite polynomials are utilized for surrogate modelling.

4.2 Polynomial chaos expansion and statistical moments

The general idea for calculation of variance and different sensitivity indices with PCE is not sensitive to particular type of orthogonal polynomials utilized for construction of PCE-basis function. Therefore, sensitivity indices can be computed analytically without any Monte Carlo simulations for any kernel $\mathcal{K}(\boldsymbol{x})$ as long as the assumption about mutual independence of parameter vector components Eq. (22) holds. Technique utilized for such kind of calculations plays important role in the developed approach for Bayesian experimental design. Therefore, it is explained in details in the current section.

Regardless of the family of orthogonal polynomials, the routine for sensitivity analysis and uncertainty quantification is essentially the same. For instance, mean value of the function $f(\boldsymbol{x})$ is given by constant term in PCE:

$$\mathbb{E}[f] = \int f(\mathbf{x})\mathcal{K}(\mathbf{x})d\mathbf{x} = \sum_A c_A \langle 1, p_A \rangle = c_A \delta_{A,0} = c_0 \tag{25}$$

Where $\mathbb{E}[f]$ is the mean or expected value of f, c_0 is the notation for PCE coefficient with multi-index $0, \dots, 0$. similar technique can be utilized for calculating the variance of f:

$$\mathrm{Var}(f) = \int \left(f(\mathbf{x}) - \mathbb{E}[f]\right)^2 \mathcal{K}(\mathbf{x}) d\mathbf{x} = \sum_{A_1, A_2 \neq 0} c_{A_1} c_{A_2} \langle p_{A_1}, p_{A_2} \rangle \tag{26}$$

Given the orthogonality of basis functions Eq. (21), the variance of f can be computed as follows:

$$\mathrm{Var}(f) = \sum_{A \neq 0} c_A^2 \tag{27}$$

Calculation of the partial variances and Sobol' indices is slightly more technically involved, but it follows the same lines as the variance calculation in Eq. (25) and Eq. (26). First of all, partial variance of f is defined as the function variance averaged with respect to some of the coordinates. In other words, if $r_1, .., r_n$ is a reordering of $1, \ldots, n$, for any m from 1 to n the following function can be defined:

$$f_{r_1,\ldots,r_m}(\mathbf{x}) = \int f(\mathbf{x}) \mathcal{K}_{r_{m+1}}(x_{r_{m+1}}) \ldots \mathcal{K}_{r_n}(x_{r_n}) dx_{r_{m+1}} \ldots dx_{r_n} \tag{28}$$

Given the PCE expression of f, it is simple to demonstrate with arguments identical to derivation of Eq. (24), that PCE series of $f_{t_1,\ldots,r_m}(\boldsymbol{x})$ have the following form:

$$f_{r_1,\ldots,r_m}(\mathbf{x}) = \sum_{\{A | \alpha_{r_{m+1}} = \ldots \alpha_{r_n} = 0\}} c_A p_A(\mathbf{x}) \tag{29}$$

In other words, the PCE series for $f_{r_1,\ldots r_m}$ are formed by those polynomials of PCE series for $f(\mathbf{x})$ that have degree zero with respect to $x_1, \ldots, x_n$. This set of indices is denoted as Ω in order to simplify the notations. Given the above definition of $f_{r_1,\ldots,r_n}$, the partial variance of f can be computed as the variance of $f_{r_1,\ldots,r_m}$:

$$\mathrm{Var}_{r_1,\ldots,r_m}(f) = \sum_{\{A \in \Omega\}} c_A^2 \tag{30}$$

It is simple to show that regular variance of f is related to partial variances via the following relation:

$$\mathrm{Var}(f) = \sum_{m=1}^{n-1} \sum_{r_1 < \ \ldots \ < r_m} \mathrm{Var}_{r_1,\ldots,r_m}(f) \tag{31}$$

Therefore, partial variances of f are typically normalized and considered as Sobol' sensitivity indices:

$$S_{r_1,\ldots,r_m} = \frac{\mathrm{Var}_{r_1,\ldots,r_m}(f)}{\mathrm{Var}(f)} \tag{32}$$

Here $S_{r_1,\ldots,r_m}$ is Sobol' sensitivity index that measures contribution of the interaction of $x_{r\_1}, \ldots, x_{r_m}$ to the overall variance of the function. Therefore, the strength of interaction

between different model parameters can be quantified through Sobol' indices. It should be noted that those indices are computed analytically if PCE is utilized. In the case when alternative response surface is utilized, Sobol' indices are computed numerically with Monte Carlo simulations. The latter requires large number of function evaluation due to exponential growth of the overall number of sensitivity indices with the dimension of the problem.

4.3 Numerical methods for polynomial chaos expansion

In the present work, PCE coefficients are computed numerically, based on the known values y_i of function $f(\boldsymbol{x})$ at finite number of points $\boldsymbol{x}_i$, where $i = 1, \dots, N$. For that purpose, a regression approach has been adopted for truncated PCE series in Eq. (19) so that polynomials of degree less or equal to a given d are considered. The degree of polynomial functions is tuned to provide the highest accuracy without overfitting the data. Given the PCE truncation scheme, the vector of PCE coefficients $\boldsymbol{c}$ can be computed via minimization of the mean-square error functional:

$$\mathcal{F}(\mathbf{c}) = \frac{1}{N}\sum_i \left(y_i - \sum_A c_A p_A(\mathbf{x}_i) \right)^2 \tag{33}$$

It is well-known that direct minimization of the mean-square error could provide unstable solution or a response surface that is not accurate at the points outside the training data. Therefore, Elastic Net [Mol et al., 2009] regularization techniques is utilized. Therefore, PCE coefficients are computed as the solution of the following minimization problem:

$$\mathbf{c}^* = \underset{\mathbf{c}}{\operatorname{argmin}}(\mathcal{F}(\mathbf{c}) + \lambda_1|\mathbf{c}|_1 + \lambda_2|\mathbf{c}|_2^2) \tag{34}$$

Where λ_1 and λ_2 are hyperparameters that are computed via cross-validation and $|\boldsymbol{c}|_1$ and $|c|_2$ are the ℓ_1 and ℓ_2 norms of $\boldsymbol{c}$, respectively.

The minimization problem in Eq. (34) is solved with coordinate-descent algorithm [Hasite et al., 2010] implemented in scikit-learn library [Pedregosa et al., 2011]. The implemented PCE technique has been naturally integrated with the scikit-learn library enabling the use of PCE in any machine-learning pipeline allowing one to combine different regression methods and cross-validation techniques.

5 Polynomial chaos expansion for Bayesian experimental design

In the present work PCE is utilized in Bayesian experimental design in two ways. The first one follows classical techniques for Bayesian experimental design. In this approach, model function $f(\theta, \boldsymbol{d})$ from Eq. (1) is approximated with PCE surrogate model. In the scope of CO_2 sequestration and data gathering optimization such approximation makes it possible to replace a run of the flow simulator by an evaluation of a polynomial function, which is dramatically faster. Therefore, surrogate modeling significantly reduces the computational cost of evaluating $U(\boldsymbol{d})$ and makes it feasible to perform Bayesian experimental design for the highly non-linear systems of subsurface reservoirs. Secondly, PCE serves as a fundamental tool to approximate the utility function $U(\boldsymbol{d})$. This technique allows one to compute a response surface for $U(\boldsymbol{d})$ at the cost of several dozens of function evaluations, which is critical for the optimization routine.

The proposed approach utilizes the reparametrization trick that is essentially the same as one utilized to derive the objective function for variational autoencoders [Welling et al., 2019]. In Eq. (8) the expected information gain is represented as an expected value of the KL-divergence:

$$U(\mathbf{d}) = \int D_{KL}(\mathbf{m}, \mathbf{d}) p(\mathbf{m}|\mathbf{d}) d\mathbf{m} \tag{35}$$

Bayesian evidence factor can be computed as an averaged likelihood over the prior distribution of model parameters Eq. (5). Therefore, the following equation for $U(\boldsymbol{d})$ is valid:

$$U(\mathbf{d}) = \int D_{KL}(\mathbf{m}, \mathbf{d}) p(\mathbf{m}|\theta, \mathbf{d}) p(\theta|\mathbf{d}) d\theta d\mathbf{m} \tag{36}$$

It is useful to rewrite the equation for $U(\boldsymbol{d})$ as repeated integration over $\boldsymbol{m}$ and θ:

$$U(\mathbf{d}) = \int \left(\int D_{KL}(\mathbf{m}, \mathbf{d}) p(\mathbf{m}|\theta, \mathbf{d}) d\mathbf{m} \right) p(\theta|\mathbf{d}) d\theta \tag{37}$$

In the present work, we assume that the vector of measurements is combined with the model predictions and normally distributed error Eq. (4). Therefore, $\boldsymbol{m}$ can be substituted as $f(\theta, \boldsymbol{d}) + \eta$ inside the inner integral in Eq. (36):

$$U(\mathbf{d}) = \int \left(\int D_{KL}(f(\theta, \mathbf{d}) + \eta, \mathbf{d}) p(f(\theta, \mathbf{d}) + \eta|\theta, \mathbf{d}) d(f(\theta, \mathbf{d}) + \eta) \right) p(\theta|\mathbf{d}) d\theta \tag{38}$$

In the inner integration, the vector of model parameters θ is constant. Therefore $U(\boldsymbol{d})$ can be computed as following:

$$U(\mathbf{d}) = \int \left(\int D_{KL}(f(\theta, \mathbf{d}) + \eta, \mathbf{d}) p(f(\theta, \mathbf{d}) + \eta|\theta, \mathbf{d}) d\eta \right) p(\theta|\mathbf{d}) d\theta \tag{39}$$

The error of measurements is normally distributed. Therefore, the likelihood function in the Eq. (38) can be replaced by density of normal distribution:

$$p\big(f(\theta, \mathbf{d}) + \eta|\theta, \mathbf{d}\big) = \mathcal{N}(\eta, 0, \sigma) \tag{40}$$

Here $\mathcal{N}(\eta, 0, \sigma)$ is the density of a normal distribution with zero mean and standard deviation σ at the point η. Therefore, expected information gain can be expressed as an integral over the noise and model parameters:

$$U(\mathbf{d}) = \int D_{KL}\big(f(\theta, \mathbf{d}) + \eta, \mathbf{d}\big) N(\eta, 0, \sigma) p(\theta|\mathbf{d}) d\eta d\theta \tag{41}$$

In the present work, both prior distribution of model parameters and probability distribution of noise are considered to be normal. Uniform distribution of design parameters $\boldsymbol{d}$ is

introduced as well. Therefore, it is natural to utilize PCE to build a surrogate model for the function inside the integral in Eq. (40):

$$D_{KL}\big(f(\theta,\mathbf{d})+\eta,\mathbf{d}\big)=\sum_{A_1,A_2,A_3} c_{A_1,A_2,A_3}p_{A_1}(\mathbf{d})p_{A_2}(\theta)p_{A_3}(\eta) \tag{42}$$

Where A_1, A_2, A_3 are multi-indices and $p_{A_1}, p_{A_2}, p_{A_3}$ are PCE basis functions that depend only on θ and η, respectively. Given the response surface for $D_{\mathrm{KL}}(f(\theta,\boldsymbol{d})+\eta,\boldsymbol{d})$, it is possible to derive a response surface for $U(\boldsymbol{d})$ with a similar method to the one utilized for derivation of the expression of the partial variances and Sobol' indices Eq. (27):

$$U(\mathbf{d})=\int D_{KL}\big(f(\theta,\mathbf{d})+\eta,\mathbf{d}\big)N(\eta,0,\sigma)p(\theta,\mathbf{d})d\eta d\theta=\sum_{A_1} c_{A_1,0,0}p_{A_1}(\mathbf{d}) \tag{43}$$

Therefore, if PCE series for $D_{\mathrm{KL}}(f(\theta,\boldsymbol{d})+\ \eta,\boldsymbol{d})$ is known, then PCE approximation for $U(\boldsymbol{d})$ can be computed by selecting only that terms from Eq. (41) that have degree zero with respect to η and θ . Unfortunately, approximating $D_{\mathrm{KL}}(f(\theta,\boldsymbol{d})+\eta,\boldsymbol{d})$ could be a challenging task due to the curse of dimensionality. Therefore, in the present work it is proposed to generate the training data for $D_{\mathrm{KL}}(f(\theta,\boldsymbol{d})+\ \eta,\boldsymbol{d})$ as a function of $\boldsymbol{d}$, θ and η and approximate it with polynomials that depend on d only. Therefore, the PCE coefficients for $U(\boldsymbol{d})$ are computed as a solution to the following minimization problem:

$$\mathbf{c}^*_{\cdot,0,0}=\underset{\mathbf{c}_{\cdot,0,0}}{\operatorname{argmin}}\left(\frac{1}{N}\sum_i\left(D_{KL}\big(f(\theta_i,\mathbf{d}_i)+\eta_i,\mathbf{d}_i\big)-\sum_A c_{A,0,0}p_A(\mathbf{d}_i)\right)^2+\lambda_1|\mathbf{c}_{\cdot,0,0}|_1+\lambda_2|\mathbf{c}_{\cdot,0,0}|_2\right) \tag{44}$$

Where the hyperparameters λ_1 and λ_2 are tuned to minimize the cross-validation error. It can be shown that if the model parameters θ_i and η_i are sampled from prior distribution and from distribution of measurements errors respectively, then the solution for Eq. (44) approximates the actual PCE coefficients of $U(\boldsymbol{d})$ and coincide with them if size of the training data Ngoes to infinity. Therefore, the proposed technique can be applied for calculation of PCE coefficients for the expected information gain.

The practical difficulty with application of Eq. (44) for calculation of PCE coefficients is the high level of noise because $D_{\mathrm{KL}}(f(\theta_i,\boldsymbol{d_i})+\ \eta_i,\boldsymbol{d_i})$ oscillates near the average value $U(\boldsymbol{d})$. Therefore, a significant number of training points is required to compute $c^*_{\cdot,0,0}$ with reasonable accuracy. Our numerical experiments demonstrate that it requires around 100,000 training points to develop an accurate PCE approximation for the expected information gain. Even with such size of the training data, the computational cost of derivation of PCE surrogate for $U(\boldsymbol{d})$ is comparable with 10 evaluations of $U(\boldsymbol{d})$ at different points. Therefore, the proposed technique has a potential to significantly accelerate calculation of the optimal experimental design parameters.

6 Numerical Examples

In this section, we evaluate the proposed PCE-based approach to solve the optimal Bayesian experimental design on several numerical examples. The proposed technique is first validated against simple analytical models that admits an approximate analytical solution. Further evaluations are presented on subsurface flow problem relevant to the capacity estimation of CO_2 storage sites.

6.1 Test case 1

In the current test case, two models in the form of Eq. (1) are considered:

$$\mathbf{m}_a = f_a(\theta, \mathbf{d}) + \eta_a \tag{45}$$

where $a = 1, 2$ is the model index, m_a are the observable data for model a, θ and d are one-dimensional design parameters, η_a is a normally distributed noise with standard deviation σ. The nonlinear functions $f_a(\theta, \boldsymbol{d})$ are defined by:

$$f_1(\theta, \mathbf{d}) = \theta^3 d^2 + \theta \exp\left(|d - 0.2|\right) \tag{46}$$

$$f_2(\theta, \mathbf{d}) = \theta^3 d^2 + \theta \exp\left(-20(d - 0.2)^2\right) \tag{47}$$

Models from Eq. (46) and Eq. (47) describe similar systems. The principal difference between these two models is in the degree of smoothness. The model defined by Eq. (46) has a discontinuity in the first order derivatives while the model defined by Eq. (47) has continuous derivatives of any order. Due to that, systems described by Eq. (46) and Eq. (47) are referred to as non-smooth and smooth, respectively.

For both models, the parameter θ is a uniformly distributed random variable in $\mathcal{U}[0; 1]$ and $d \in [0; 1]$. For small values of σ the following approximation for the KL-divergence can be derived:

$$D_{KL,a}(f_a(\theta, \mathbf{d}) + \eta, \mathbf{d}) \approx D_{KL,a}(f_a(\theta, \mathbf{d}), \mathbf{d}) = \log\left(\frac{1}{(2\pi)^{1/2}\sigma}\frac{\partial f_a}{\partial \theta}\right) - 1/2 \tag{48}$$

Here, we neglect the value of η. Therefore, Eq. (48) is only valid for small values of σ.

In the present test case, $200{,}000$ values ofθ and η are sampled from prior distribution $\mathcal{U}[0; 1]$ and from normal distribution with variance σ, respectively. The following values of σ are considered: 3.0×10^{-3}, 1.0×10^{-3}, 3.0×10^{-4} and 1×10^{-4}. The range of parameter σ is selected in such a way that assumption about small magnitude of σ is valid for the smallest σ considered and is violated for the highest one. In such setting, $200{,}000$ samples for $\boldsymbol{d}$ are generated for uniform distribution $\mathcal{U}[0; 1]$. For each generated sample the KL-divergence is computed numerically using Eq. (41). The first term in Eq. (41) is computed with $50{,}000$ MCMC-samples generated from the posterior distribution and the second term in Eq. (41) is computed with $200{,}000$ MCMC-samples generated from the prior distribution Eq. (2). Since the KL-divergence values have already been generated for all samples, $\boldsymbol{d}$ is rescaled to $\mathcal{U}[-1; 1]$ in order to allow for using Legendre polynomials as a basis function in the PCE

expansion. Polynomials up to degree eight were utilized in order to produce a response surface for $U(\boldsymbol{d})$ in accordance with Eq. (44) For the purposes of validation, the utility function values at any given $\boldsymbol{d}$ is computed via Eq. (41) and Eq. (48), where the integration is replaced by averaging over $200{,}000$ samples generated in agreement with the prior distribution $\mathcal{U}[0;1]$ and the likelihood defined in Eq. (4). Results of comparison of two techniques for $U(\boldsymbol{d})$ calculation are shown in Figure 2 and Figure 3.

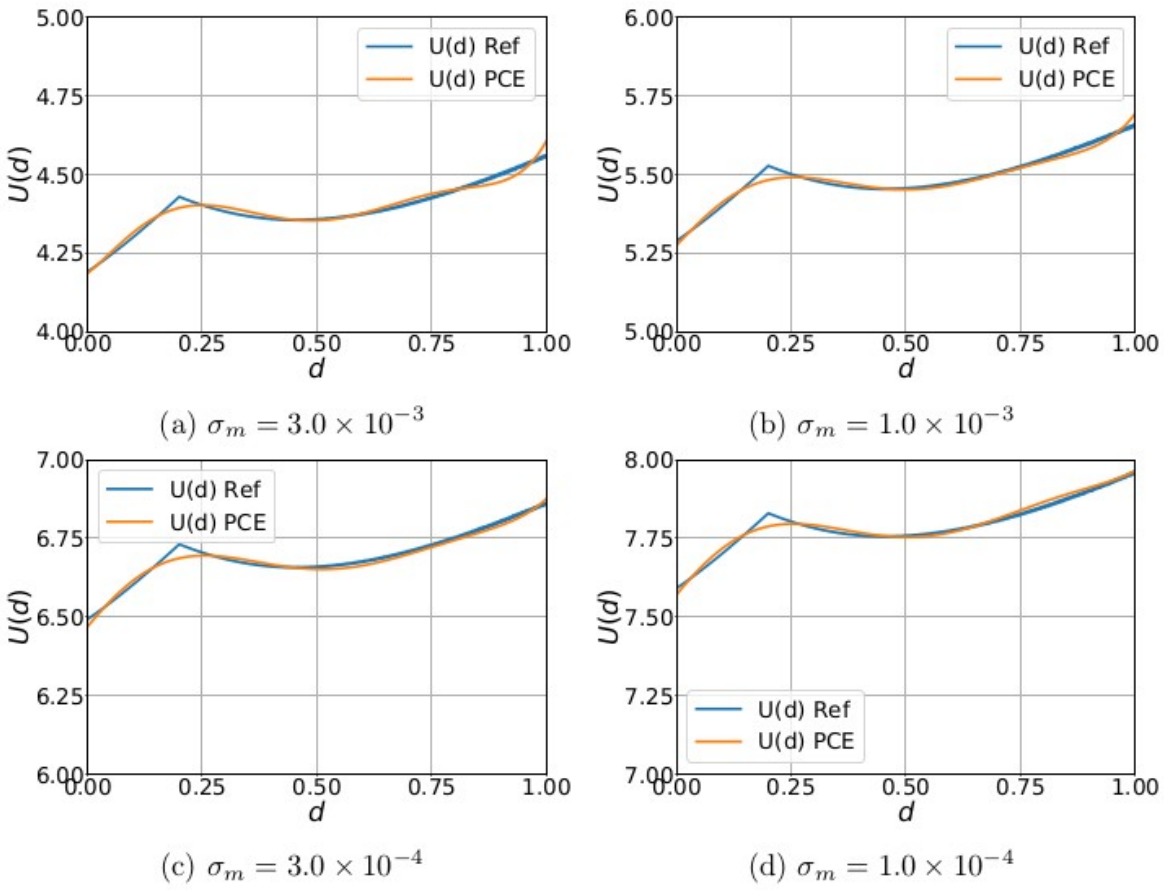

Figure 2: Plots of the expected information gain $U(\boldsymbol{d})$ versus the design parameter $\boldsymbol{d}$ for non-smooth objective functions defined in Eq. (46) for different values of σ .

It can be observed that for both of the test cases the proposed PCE approach provides a relatively accurate approximation of the utility function $U(\boldsymbol{d})$. Almost exact match can be observed for the cases of small σ where the KL-divergence approximate defined in Eq. (48) is supposed to be valid. For small values of σ, both methods that correspond to Eq. (41) and Eq. (48) provides similar estimates for $U(\boldsymbol{d})$. However, for high σ values some divergence between those techniques is observed. This is expected because Eq. (48) is not supposed to work in those cases. In addition to that, the PCE response surface for $U(\boldsymbol{d})$ failed to reproduce the discontinuity in the derivative of $U(\boldsymbol{d})$ as it can be observed in Figure 2. This is an expected behaviour of PCE response surface because of the smooth basis functions. What is more important, the design value $\boldsymbol{d}$ corresponding to the local optimum is accurately reproduced. Therefore, the present test case demonstrates that the introduced PCE based technique for estimating the utility $U(\boldsymbol{d})$ function is more accurate for smooth problems. However, the

proposed PCE based approach could still be used for both smooth and non-smooth cases to estimate the optimal Bayesian experimental design, because the design values $U(\boldsymbol{d})$ maximizing the utility $U(\boldsymbol{d})$ are accurately approximated.

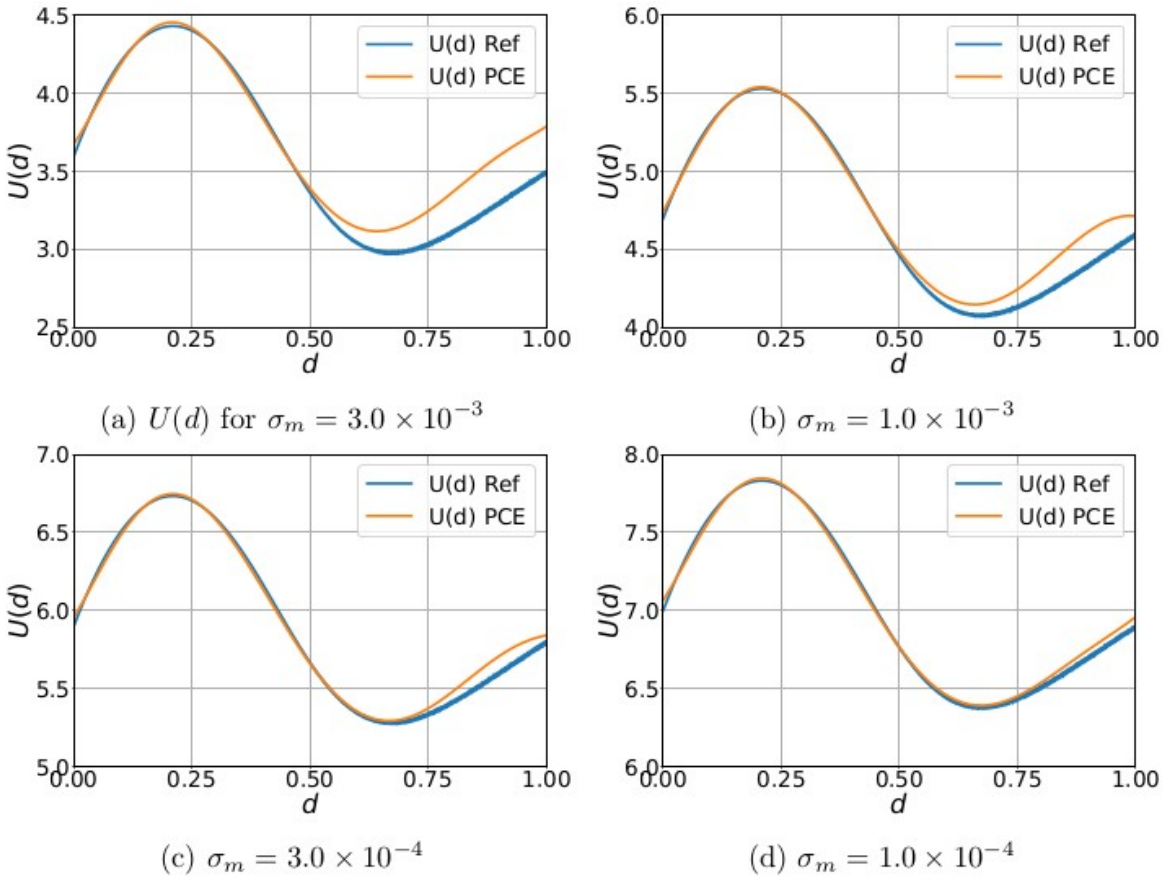

Figure 3: Plots of the expected information gain $U(\boldsymbol{d})$ versus the design parameter $\boldsymbol{d}$ for smooth objective functions defined in Eq. (47) for different values of σ.

6.2 Test case 2

In this test case, we consider a two-phase subsurface flow related to the forecast of hydrocarbon oil production. On one hand, the accuracy of the forecast is directly related to the quantity and quality of available data used to estimate the subsurface rock properties. On the other hand, direct measurements of those properties is an expensive process. Therefore, utilizing optimal experimental design techniques to decide which data to be collected in order to produce accurate predictions of hydrocarbons production is of great practical importance. In the present numerical example we optimize the design of experiment in order to maximize the accuracy of subsurface parameters measurements, which in turn reduce the uncertainty in the oil production forecast.

In the present test case, numerical simulations of oil production enhanced by well-known water-flooding technique are studied. During this process water is injected into the reservoir via a group of wells called injection wells (aka. injectors) and displaces oil that saturates the pores of the reservoir rocks. Hydrocarbons, in turn, are produced via another group of wells

called production wells (aka. producers). The fluid flow in the reservoir together with the oil production rates is controlled by the spatial distribution of reservoir properties namely, the porosity field $\phi(\boldsymbol{r})$ and the permeability field $k(\boldsymbol{r})$. Typically, the porosity and permeability are known at several locations in the reservoir where rock samples have been extracted during drilling. These point values are then used within stochastic interpolation frameworks (aka. Geo-statistics [Richard Webster et al., 2015]) to populate the model parameters over the entire domain of interest.

In the present test case, we solve for optimal design of experiment based on Bayesian framework. We consider a five-spot injection pattern where an injection well is located at the centre of a square surrounded by four production wells. Given the symmetry of this pattern, only one quarter of the domain is modelled with one producer and one injector located at the opposite corners of a square domain. The length of the edge of that square is set to $L = 500\mathrm{m}$. The thickness of the reservoir is$h = 10\mathrm{m}$. We do not consider discretization along the vertical direction and we only consider a two-dimensional flow problem. Further, the porosity is considered to be constant value,$\phi(\boldsymbol{r}) = 0.2$. Also, we assume that data is collected by drilling additional wells and results in a measurement of the permeability value at the location and a measurement of the pressure value at specified moments of time. Alternative sources of data, like seismic measurement are not considered in the present example.

In this setting, the vector of design parameters $\boldsymbol{d}$ is formed by the 2D coordinates additional wells and the dimension of the design parameter space can be computed as following:

$$\dim(\mathbf{d}) = 2n_s \tag{49}$$

Where n_s is total number of new wells. In the present example only two cases are considered: $n_s = 1, 2$.

The vector of model parameters θ is introduced via Karhunen-Loeve (KL) expansion for the spatially discretized log-permeability field $\log(k(\boldsymbol{r}))$. The log-permeability distribution is assumed to be a linear combination of the reference permeability field representing the general trend of the field and random perturbation that is defined stochastically:

$$\log(k(\mathbf{r})) = \log(k_{\mathrm{ref}}(\mathbf{r})) + \zeta(\mathbf{r}) \tag{50}$$

where $k_{ref}(\boldsymbol{r})$ is the reference permeability field at the point $\boldsymbol{r}$, $k(\boldsymbol{r})$ is a value of permeability field at $\boldsymbol{r}$ and $\zeta(\boldsymbol{r})$ represents the perturbations to the logarithm of the reference permeability. The perturbation $\zeta(\mathbf{r})$ is set as zero at the locations of injector and producer wells as the permeability is known at those grid blocks. For generating multiple realizations of the $\log\big(k_{ref}(\boldsymbol{r})\big)$, KL expansion is applied to spatially discretized permeability field. In more details, it is assumed that values of $\log\big(k_{ref}(\boldsymbol{r})\big)$ at grid-blocks are exponentially correlated:

$$\begin{aligned}\langle \log(k_{\mathrm{ref}}(\mathbf{r}_1)), \log(k_{\mathrm{ref}}(\mathbf{r}_2))\rangle &= \exp\left(-\frac{|\mathbf{r}_1 - \mathbf{r}_2|}{L_{\mathrm{ref}}}\right)\\ \langle \zeta(\mathbf{r}_1), \zeta(\mathbf{r}_2)\rangle &= \exp\left(-\frac{|\mathbf{r}_1 - \mathbf{r}_2|}{L_{\mathrm{p}}}\right)\end{aligned} \tag{51}$$

where $L_{ref} = 0.3L$ and $L_p = 0.1L$ is the correlation length for reference permeability and perturbation and $L = 500\mathrm{m}$ is the side of the square reservoir. For both $\log(k(\boldsymbol{r}))$ and $\zeta(\boldsymbol{r})$ KL expansion is performed:

$$\log(k_{\mathrm{ref}}(\mathbf{r})) = \sum_j \lambda_{\mathrm{ref},j}\chi_{\mathrm{ref},j}\xi_{\mathrm{ref},j}(\mathbf{r})$$
$$\zeta(\mathbf{r}) = \sum_j \lambda_{\mathrm{p},j}\chi_{\mathrm{p},j}\xi_{\mathrm{p},j}(\mathbf{r}) \quad (52)$$

where $\lambda_{ref,j}$ and $\xi_{ref,j}(\boldsymbol{r})$ are eigenvalues and eigenfunctions respectively for KL expansion for the random field with correlation length L_{ref}, $\lambda_{p,j}$ and $\xi_{p,j}(\boldsymbol{r})$ are eigenvalues and eigenfunctions respectively for KL expansion for the random field with correlation length L_p respectively, $\chi_{ref,j}$ and $\chi_{p,j}$ are uncorrelated normally distributed random variables with zero means and standard deviations $\sigma_{ref} = 2.0$ and $\sigma_p = 0.5$. In the present work permeability field is normalized in such a way that zero values of $\chi_{ref,j}$ correspond to permeability of $1\mathrm{mD}$ ($1\mathrm{mD}$=$9.869233\times10^{-16}\mathrm{m}^2$). Both reference permeability distribution and perturbation are generated stochastically by sampling of random values from appropriate normal distribution for $\chi_{ref,j}$ and $\chi_{p,j}$.

In the present example, the KL expansion for the reference permeability distribution is truncated and only first 50 eigenfunctions are considered. Then a single realization of reference permeability distribution is selected and utilized in all calculations in the present test case. The perturbation $\zeta(\boldsymbol{r})$ to the $\log\big(k_{ref}(\boldsymbol{r})\big)$ is constructed in almost the same fashion. However, different values of correlation length and variance are utilized to perform the KL expansion. Moreover, only first five terms of the KL expansion are considered. In addition to that, two linear constraints on $\zeta(\boldsymbol{r})$ are utilized, because the perturbation vanishes at the injector and production wells. Therefore, $\zeta(\boldsymbol{r})$ is effectively parametrized with five coefficients of the KL expansion: $\chi_{p,1}, \dots, \chi_{p,5}$. It is clear that for a fixed $k_{ref}(\boldsymbol{r})$, the permeability distribution $k(\boldsymbol{r})$ is fully parametrized by the same parameters due to Eq. (52). Therefore, the dimension of the model parameter space $\dim(\theta)$ is simply 5. Figure 4 shows examples of the permeability field generated with the described approach above.

The observations vector m is formed by values of the permeability and pressure at selected locations. The pressure is measured at early stage of water-flooding at four different moments of time that correspond to four different values of total injected volume of water measured as a fraction of total reservoir pore volume or PVI. Two scenarios of pressure measurements are considered. In the first scenario only value of pressure for $\mathrm{PVI} = 1\%$ is measured. In the second scenario, the pressure is measured for $\mathrm{PVI} = 1\%, 2\%, 3\%, 4\%$.

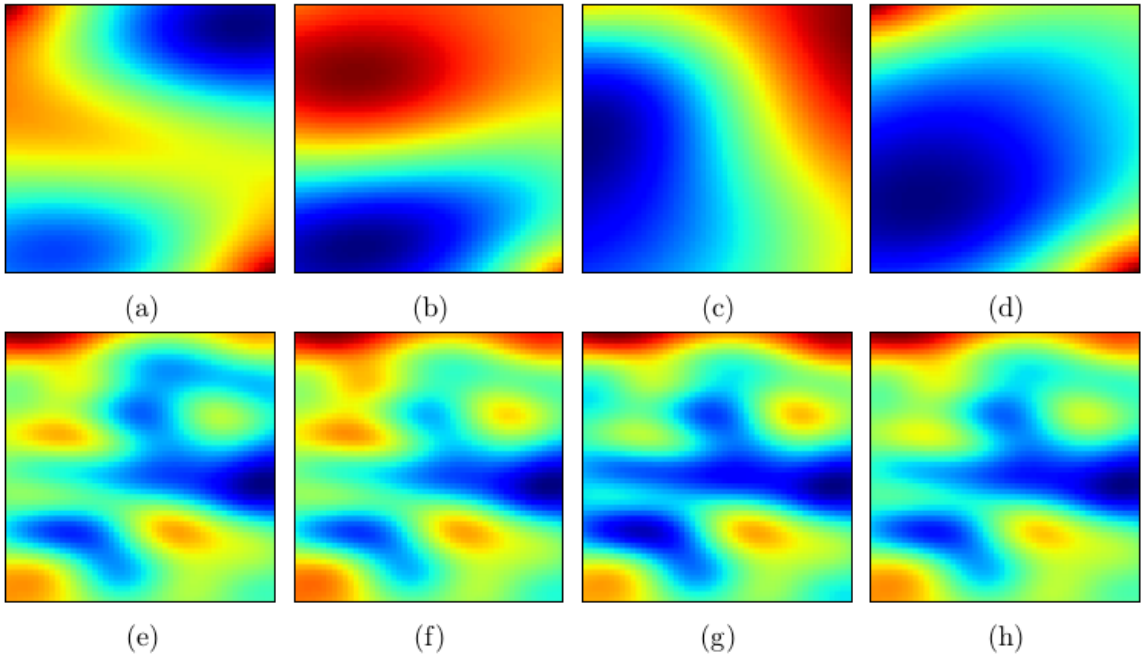

Figure 4: Realizations of the perturbations $\zeta(\boldsymbol{r})$ shown in panels (a, b, c, d) and the corresponding permeability field $log(k(\boldsymbol{r}))$ shown in the panels (e, f, g, h), respectively

In the current numerical example, the observations vector is computed numerically by solving a coupled system of PDEs namely the conservation of mass and conservation of momentum coupled via Darcy law:

$$\frac{\partial \phi s_a \rho_a}{\partial t} - \sum_{\gamma=1}^{3} \frac{\partial}{\partial r^\gamma}\left(\frac{\rho_a k k_a}{\mu_a}\frac{\partial P}{\partial r^\gamma}\right) = Q_a \tag{53}$$

Where $s_a = s_a(\boldsymbol{r})$ is a volumetric fraction or saturation of fluid with index a (water or oil) at the point $\boldsymbol{r}$, $\phi = \phi(\boldsymbol{r})$ is rock porosity at point $\boldsymbol{r}$ and $k = k(\boldsymbol{r})$ is permeability at the point $\boldsymbol{r}$, $k_a = k_a(s)$ is the relative phase permeability that depends only on the fluid saturations s at the point $\boldsymbol{r}$, $P = P(\boldsymbol{r})$ is the pressure at point $\boldsymbol{r}$, ρ_a is the density of fluid a, μ_a is the viscosity of fluid a, $Q_a = Q_a(\boldsymbol{r})$ is source term at the point $\boldsymbol{r}$. In the case of incompressible flow, Eq. (53) admits the following simplification:

$$\phi\frac{\partial s_a}{\partial t} - \sum_{\gamma=1}^{3} \frac{\partial}{\partial r^\gamma}\left(\frac{k k_a}{\mu_a}\frac{\partial P}{\partial r^\gamma}\right) = q_a \tag{54}$$

Where $q_a = Q_a/\rho_a$ is the source term for fluid a normalized to the density of corresponding fluid. For calculating the relative phase permeabilities, Brooks-Corey model [Brooks et al., 1964] is utilized:

$$\begin{aligned} k_w(S_{\mathrm{wn}}) &= k_w^{(0)} S_{\mathrm{wn}}^{p_w} \\ k_w(S_{\mathrm{wn}}) &= k_o^{(0)}(1 - S_{\mathrm{wn}})^{p_o} \end{aligned} \tag{55}$$

where k_w and k_0 are the values of the relative phase permeability for water and oil, respectively and $k_w^{(0)}$ and $k_o^{(0)}$ are the maximum values of the relative phase permeability for water and oil, respectively. The values p_w and p_o are dimensionless parameters of the model and S_{wn} is the normalized water saturation defined as:

$$S_{\mathrm{wn}} = \frac{S - S_{\mathrm{wir}}}{1 - S_{\mathrm{wir}} - S_{\mathrm{owr}}} \tag{56}$$

where S_{wir} and S_{owr} are the irreducible water and oil saturations, respectively. For the purposes of simplicity, incompressible immiscible fluids is considered while neglecting gravity effects.

A uniform square grid is used for simulations and the dimensions of each grid-block is $10\mathrm{m}$ by $10\mathrm{m}$ by $10\mathrm{m}$. In other words, a 50 by 50 by 1 mesh is used for discretization. Pressures at injection and production wells are considered to be constant and equal to $200\ \mathrm{Bar}$ and $100\ \mathrm{Bar}$ respectively. The fluid properties and parameters of Corey model are essentially the same as in [Tarakanov et. al., 2019] and are summarized in the Table 1.

μ_o, cP	μ_w, cP	p_o	p_w	$k_o^{(0)}$	$k_w^{(0)}$
10.0	1.0	2.0	2.0	1.0	1.0

Table 1. Fluid properties and parameters of the model for relative-phase permeability.

The evolution of incompressible flow is fully determined by the pressure differences between the injection well and the production wells and does not depend on the absolute values of those pressures. Therefore, the pressure distribution is rescaled in the following way:

$$P_*(t, \mathbf{r}) = \frac{P(t, \mathbf{r}) - P_0}{P_1 - P_0} \tag{57}$$

where P_0 and $P_1$are the pressures at the injection well and production wells and $P_*(t, \boldsymbol{r})$ is a normalized pressure. In the present test case, normalized pressure $P_*(t, \boldsymbol{r})$ is utilized for construction of observations vector Eq. (1).

Figure 5 shows the pressure and saturation distributions for the reference permeability field $k_{ref}(\boldsymbol{r})$ at different PVI values. The plots demonstrate that reference permeability field is highly heterogeneous, leading to a highly heterogeneous distribution of the saturation field.

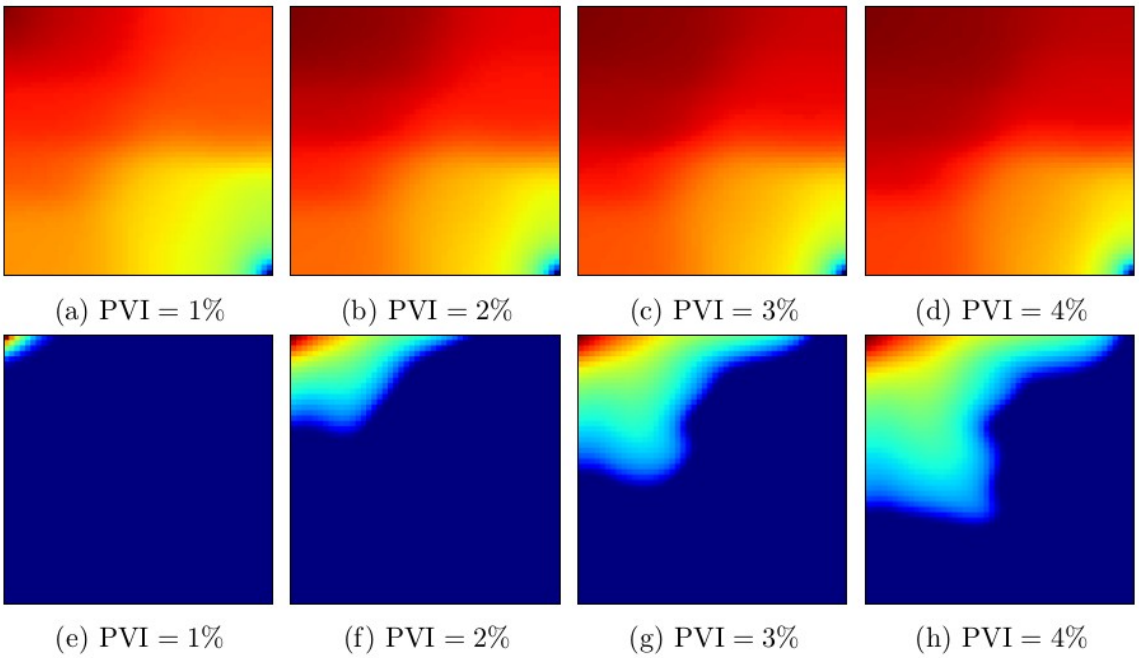

Figure 5: Snapshots of pressure (a, b, c, d) and saturation (e, f, g, h) distributions for and respectively computed for the reference permeability field PVI = 1%, 2%, 3% and 4%.

Direct utilization of flow simulations in Bayesian experimental design is not feasible due to the high computational cost of estimating the utility function $U(\boldsymbol{d})$. Therefore, PCE-based response surface for both $\zeta(\boldsymbol{r})$ and $P_*(t, \boldsymbol{r})$ has been developed. For that purpose, numerical simulations on $5{,}000$ different realizations of perturbation to the reference permeability field have been performed. A total of $4{,}000$ of those simulations are utilized for building (aka. training) the PCE-based response surface and the remaining $1{,}000$ model runs are used for validation and hyper-parameters optimization. For each simulation from the training set, a 20 grid blocks are randomly sampled and the values of $\zeta(\boldsymbol{r})$ and $P_*(t, \boldsymbol{r})$ are added to the training dataset. Finally, the PCE surrogate for permeability perturbation and pressure has been developed as a function of two spatial coordinates and five coefficients of KL expansion. Both spatial coordinates and parameters of perturbation are rescaled in such a way that classical families of orthogonal polynomials can be utilized. Namely, Legendre and Hermite probabilistic polynomials are utilized for the spatial variables and parameters of KL expansion, respectively. Basis polynomials of degree up to eight with respect to all variables are considered in PCE. Additional constraint is imposed on the Hermite polynomials, where only basis functions of degree up to four are utilized. The PCE coefficients are computed via minimization of mean-square error functional with Elastic-Net regularization terms [Hastie et. al., 2010] The accuracy of the response surface on the validation data is around 3%. The cross-plots shown in Figure 6 demonstrate quality of the response surface on both the training and validation dataset.

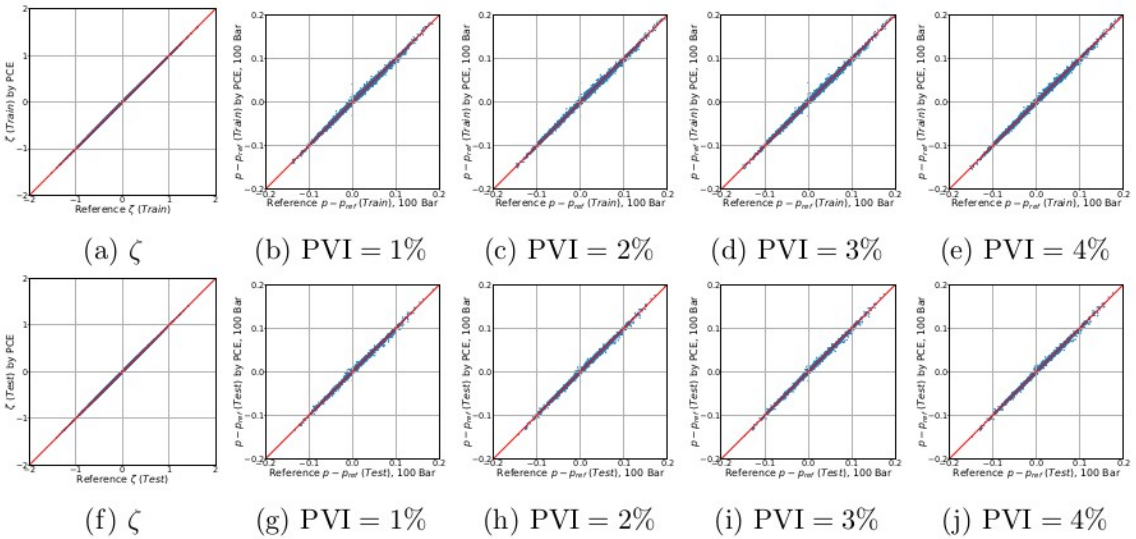

Figure 6: Cross-plots of reference values and predictions of PCE surrogate model for training (a) - (e) and test (f) - (j) data respectively. Figures (a) and (f) correspond to permeability perturbation and remaining figures correspond to deviation from the reference pressure for different values of PVI.

In the proposed method, the model function $f(\theta, \boldsymbol{d})$ is calculated via PCE-based response surface and the differences between the model predictions and observations are assumed to follow a normal distribution and the standard deviation for that difference σ is assumed to be the same for all the components of vector of observables. This is generally true, as long as $\zeta(\boldsymbol{r})$ and $P_*(t, \boldsymbol{r})$ are dimensionless quantities. In this test case, the standard deviation is set to be $\sigma = 1.0 \times 10^{-3}$. A total of $200{,}000$ realizations of design parameters d are sampled in both cases with $n_s = 1$ and $n_s = 2$. All model parameters are rescaled linearly in order to be uniformly distributed in the interval $[-1; 1]$. For each of the samples KL-divergence is computed with MCMC chain of length $50{,}000$. The computed data is then fitted with Legendre polynomials on a rescaled design parameters only. PCE is truncated by the total polynomial degree, which is set to 5. According to Eq. (44) the surrogate model developed represents $U(\boldsymbol{d})$ directly. The response surfaces for $U(\boldsymbol{d})$ are visualized for the case of a single new well, for the two different number of pressure measurements is shown in Figure 7.

According to the colour maps of expected information gain for two scenarios of pressure measurements and for single additional well (shown in Figure 7), there are two peaks of the utility function located at the corners of the model domain opposite to the injection and production wells as demonstrated in Figure 7a. In the second scenario when additional measurements of pressure are added, only one maximum is observed (at the lower left corner of the domain) as shown in Figure 7b. This observation is in agreement with the variance of perturbation of permeability and pressure as shown in Figure 8a and Figure 8b, respectively. In the case of a single measurement of pressure, the variance of permeability perturbation determines the shape of the utility function. In the second scenario, when extra pressure measurements are added the contribution of pressure variance becomes more significant. Therefore, the maximum of the utility function is shifted towards the maximum of the pressure variance. In other words, optimal parameters of experiment according to the Bayesian technique are in the proximity to the point where the sensitivity of the model

predictions to model parameters is the highest in terms of standard deviations. The latter observation is in agreement with common sense of experimental design.

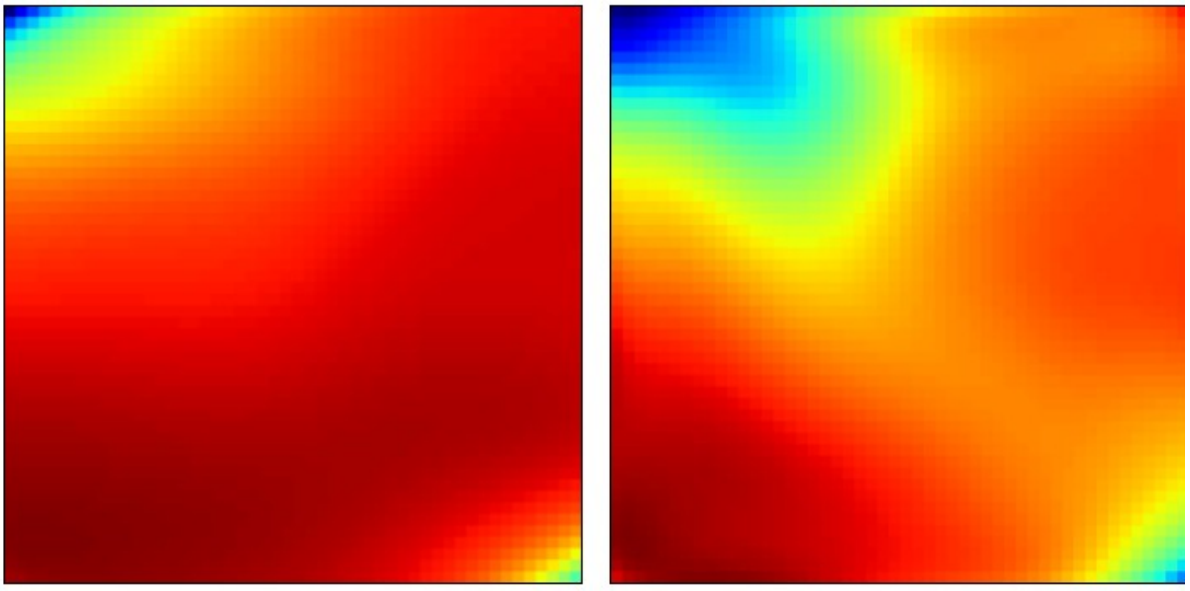

Figure 7: PCE response surface for expected information gain for experiments with one (a) and four (b) pressure measurements.

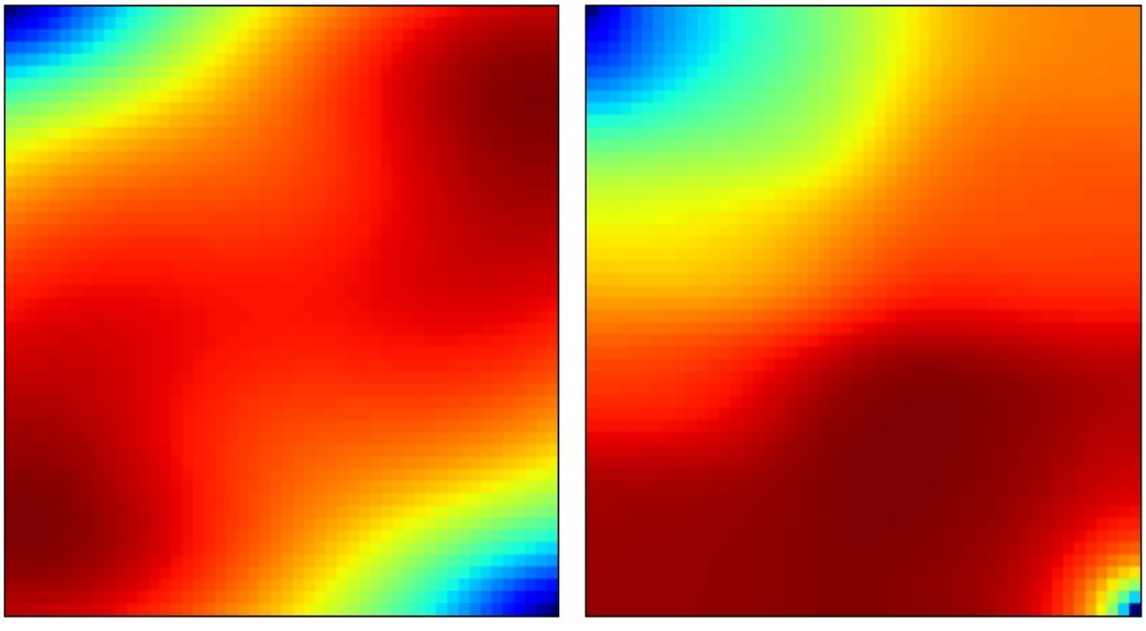

Figure 8: Variance of $\log k$ at a given grid-block (a) and variance of normalized pressure at a given grid-block (b) computed from training data.

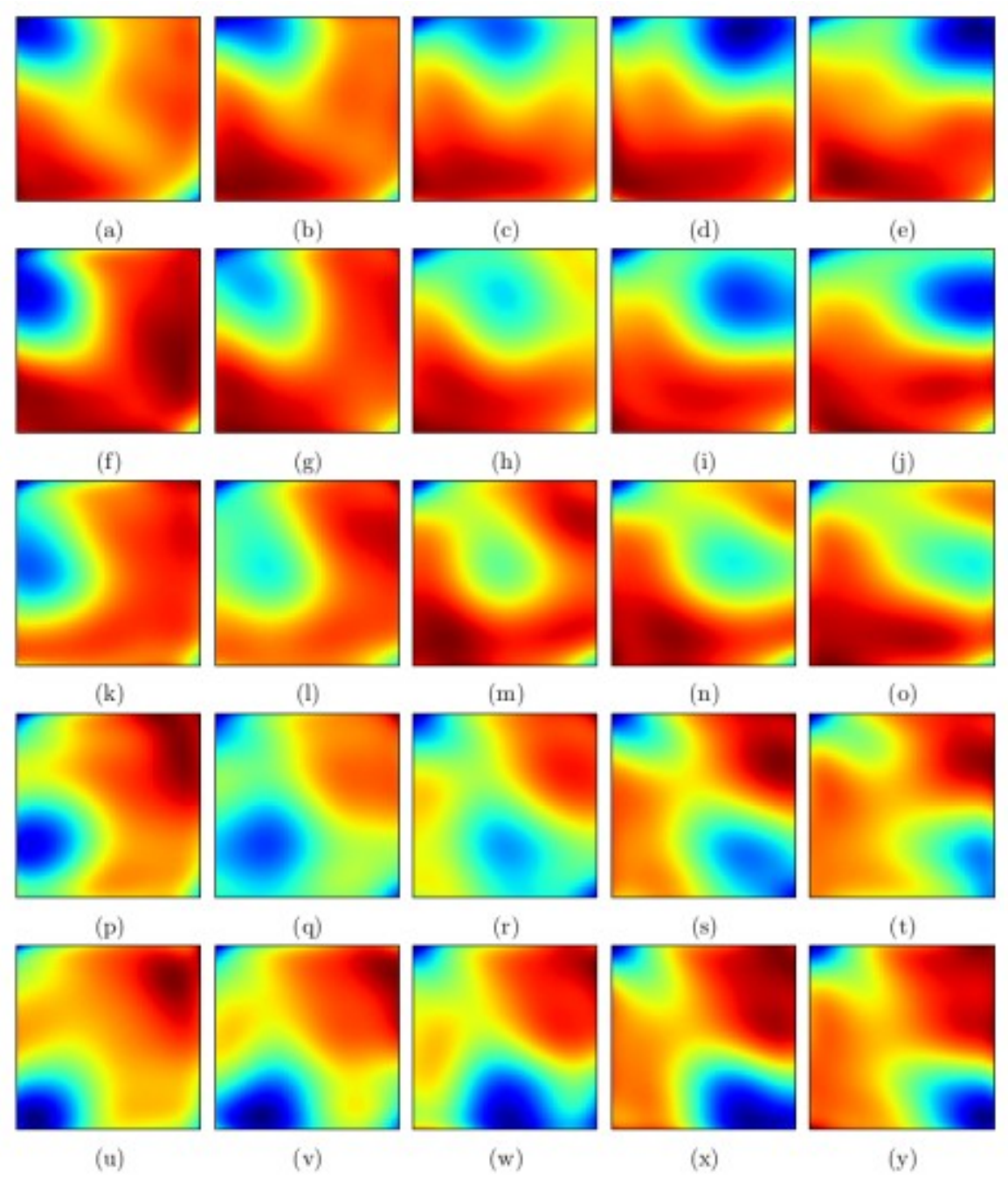

Figure 9: Values of the expected information gain as a function the first well position if the location of the second well is fixed. Each of the figures (a) - (y) corresponds to different coordinates of the second well that corresponds to the minimum of the utility function (dark blue). Case of single pressure and permeability measurement is presented.

The calculation of $U(\boldsymbol{d})$ for the case of two new measurement wells is performed in a similar fashion. In the present scenario of measurements, we focus on the examination of PCE-based response surface for $U(\boldsymbol{d})$ rather than on optimization of utility function. Therefore, optimal design parameters are not provided for the current test case. Instead, the quality of response surface is assessed visually, given the low dimension of the design parameter space and clear

geometric meaning of those parameters (aka well location). For that purpose, a 5 by 5 uniform lattice of possible locations of the second well has been generated and the expected information gain as a function of the location of the first well is plotted in Figure 9, where a single pressure measurement is performed at each new well. For four pressure measurements, the results are shown in Figure 10.

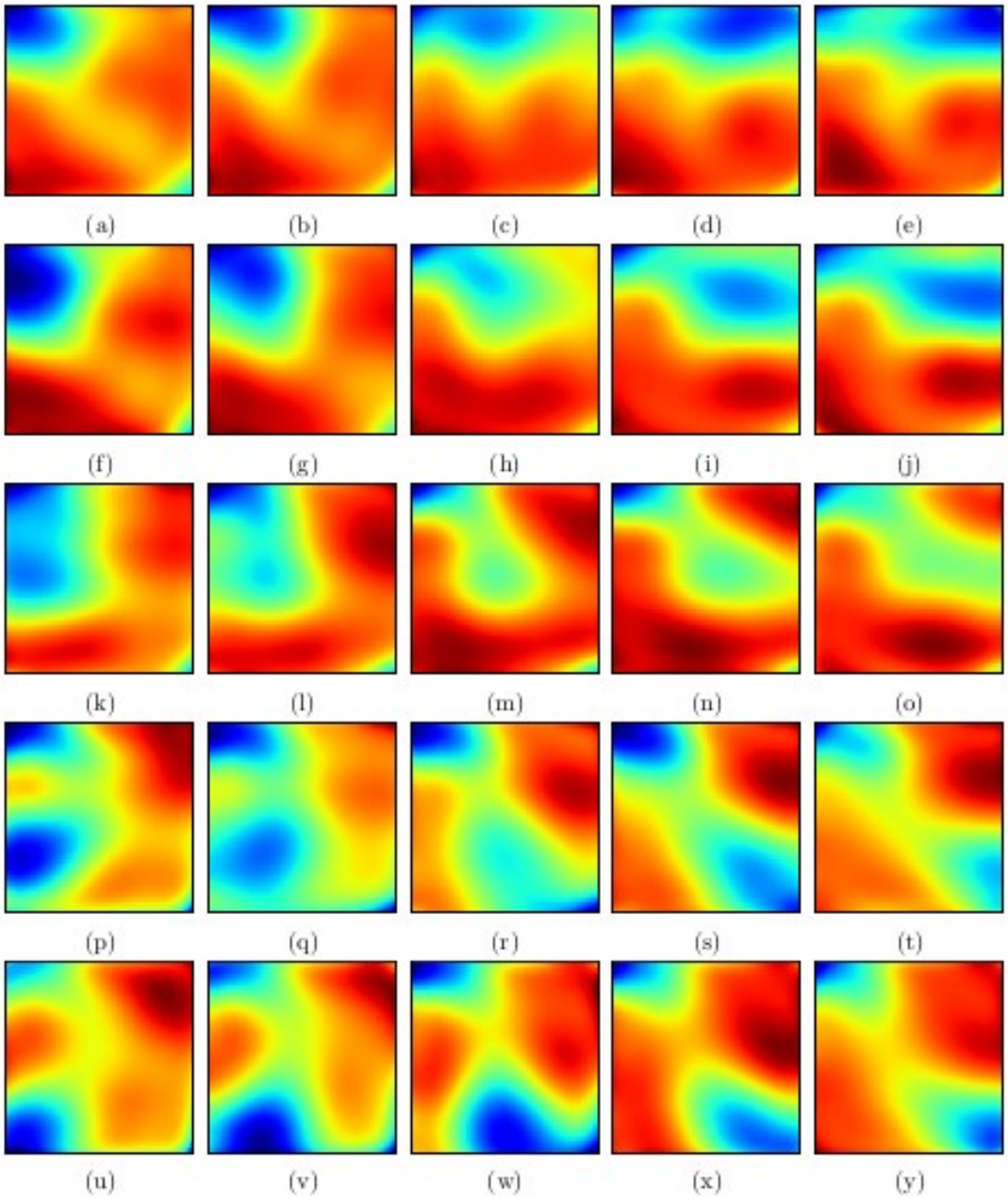

Figure 10: Values of the expected information gain as a function the first well position if the location of the second well is fixed. Each of the figures (a) - (y) corresponds to different coordinates of the second well that corresponds to the minimum of the utility function (dark blue). Case of single permeability and four pressure measurements is presented.

It can be observed from Figure 9 and Figure 10 that the minimum of expected information gain is achieved when the exploration wells are drilled close to each other or close to either the production or injection wells. Moreover, the expected information gain is high when all those wells are far from each other. In addition to that, Figure 9 and Figure 10 demonstrate that optimal experimental design corresponds to the case when measurements are collected at vicinity of domain corners that are far from the location of injection and production wells. The latter is in agreement with the variance distribution of logarithm of permeability and pressure shown in Figure 8a and Figure 8b, respectively. Therefore, the proposed technique provides reasonable estimates for $U(\boldsymbol{d})$ in the scenario concerned.

Finally, the proposed PCE-based expected information gain provides reasonable approximation of the utility function in both cases (single and multiple pressure measurements) and for single and two additional wells utilized for measurements. In all of the cases, the estimates concerned are in agreement with variation of permeability and pressure measurements collected at locations with higher variance of permeability and pressure provide more information about permeability distribution. Additionally, PCE-based expected information gain $U(\boldsymbol{d})$ reflects generic dependencies between location of exploration wells and magnitude of $U(\boldsymbol{d})$. Taking measurements at spatially close points of the reservoir or at the neighbourhood of injection and production wells is definitely not the optimal strategy of experiment. Therefore, PCE-based response surface provides reasonable representation of $U(\boldsymbol{d})$ and allows one to determine parameters of the optimal experimental design.

7 Conclusion

In this report, we introduce a new approach to Bayesian experimental design. The proposed technique utilizes PCE for averaging the KL-divergence with respect to the prior distribution of model parameters and measurement errors. The result of this procedure is a PCE response surface for the expected information gain. The proposed technique provides dramatic acceleration of solution for optimal parameters of data acquisition processes.

The proposed PCE approach has a high degree of flexibility and can be naturally extended to other systems including simulations of CO_2 sequestration process. The only assumptions that have been made include normal prior distribution and normal distribution of measurement errors. Both of these assumptions are natural and valid for a wide variety of practical systems. In addition to that, Rosenblatt transformation [Rosenblatt, 1952] can be applied to build a normal distribution in the parameter space of concern. Therefore, the proposed PCE technique can be applied to vast range of problems including CCS.

References

1. Abramowitz M. and Stegun I. A., 1964. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing.

2. AGU (American Geophysical Union), 1999. Position statement. EOS Trans AGU 1999;80:49

3. IPCC (Intergovernmental panel on climate change), 2007. *Climate change 2007: The physical science basis*. Fourth assessment report, IPCC Secretariat, Geneva, Switzerland, 2007.

4. Bachu S., 2008. *CO2 Storage in Geological Media: Role, Means, Status and Barriers to Deployment.* Progress in Energy and Combustion Science. 34. 254-273. 10.1016/j.pecs.2007.10.001.

5. Bachu S., 2015. *Review of CO2 storage efficiency in deep saline aquifers, International Journal of Greenhouse Gas Control,* Available online 1 February 2015, ISSN 1750-5836

6. Blatman G., Sudret B., 2011. *Adaptive sparse polynomial chaos expansion based on least angle regression.* Journal of Computational Physics, 230(6):2345 – 2367.

7. Box G. E. P., Behnken D. W. 1960. Some new three level designs for the study of quantitative variables. Technometrics, 2(4):455–475.

8. Brooks R.H., Corey A.T., 1964. *Hydraulic properties of porous media.* Hydrology Papers, (3).

9. Bryant E., 1997. *Climate process and change*. Cambridge, UK: Cambridge University Press.

10. CO2CRC, 2008, *Storage capacity estimation, site selection and characterisation for CO2 storage projects*, CO2CRC report RPT08-1001

11. DNV, 2009, *CO2Qualstore - Guideline for selection and qualification of sites and projects for geological storage of CO2,* DNV report 2009-1425.

12. Furre A.-K., Eiken O., Alnes H., Vevatne J. N., Kiær A. F., 2017. *20 Years of Monitoring CO2-injection at Sleipner,* Energy Procedia, Volume 114, Pages 3916-3926, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2017.03.1523.

13. Gelfand A. E., Smith A. F. M., 1990. *Sampling-Based Approaches to Calculating Marginal Densities.* Journal of the American Statistical Association, 85 (410): 398–409.

14. Gelman A., Carlin J. B., Stern H. S., Rubin D. B., 2004, Bayesian Data Analysis, Chapman and Hall/CRC

15. Hastie T., Tibshirani R., 2010. *Regularization paths for generalized linear models via coordinate descent.* Journal of Statistical Software, 33:1 – 22.

16. Hastings, W. K. 1970. *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. Biometrika. 57 (1): 97–109.

17. Heidug W., 2013. IEA Workshop 2013 *Methods to access geologic CO2 storage capacity: status & best practice.*

18. Huan X. and Marzouk Y. M., 2013. *Simulation-based optimal Bayesian experimental design for nonlinear systems.* Journal of Computational Physics, 232(1):288 – 317*. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2012.08.013.*

19. Kaintura A., Dhaene T., Spina D., 2018. *Review of polynomial chaos-based methods for uncertainty quantification in modern integrated circuits*. ELECTRONICS, 7(3):21.

20. Kullback, S., Leibler, R. A., 1951. *On information and sufficiency.* Ann. Math. Statist., 22(1):79–86, 03.

21. McKay M. D., Beckman R. J., and Conover W. J., 1979. *Comparison of three methods for selecting values of input variables in the analysis of output from a computer code*. Technometrics, 21(2):239–245.

22. Miler P., Parmigiani G., 1995. *Optimal design via curve fitting of monte carlo experiments.* Journal of the American Statistical Association, 90(432):1322–1330.

23. Miocic J. M., Gilfillan S. M. V., McDermott C., Haszeldine R. S., 2013. *Mechanisms for CO2 Leakage Prevention – A Global Dataset of Natural Analogues,* Energy Procedia, Volume 40, Pages 320-328, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2013.08.037.

24. Mol C. De, Vito E. De, and Rosasco L., 2009. *Elastic-net regularization in learning theory*. Journal of Complexity, 25(2):201 – 230.

25. Montgomery D.C. 2012. *Design and Analysis of Experiments*, 8th Edition. John Wiley & Sons, Incorporated. ISBN 9781118214718.

26. Mosbach S., Braumann A., Man P. L. W., Kastner C. A., Brownbridge G. P. E., Kraft M., 2012. *Iterative improvement of Bayesian parameter estimates for an engine model by means of experimental design.* Combustion and Flame, 159(3):1303 – 1313.

27. Neal R., 2011. *MCMC Using Hamiltonian Dynamics.* In Handbook of Markov Chain Monte Carlo, edited by Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, 116–62. Chapman; Hall/CRC.

28. NETL, 2010, *Site screening, selection and initial characterization for storage of CO2 in deep geological formations,* DOE report DOE/NETL-401/090808

29. Plackett R. L., Burman J. P., 1946. *The Design of optimum mulifractional experiments*. Biometrika, 33(4):305–325, 06.

30. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M.,. Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., and Duchesnay E., 2011. *Scikit-learn: Machine learning in Python.* Journal of Machine Learning Research, 12:2825–2830.

31. Rosenblatt M., 1952. *Remarks on a multivariate transformation*. Ann. Math. Statist., 23(3):470–472, 09.

32. Roshan J. V., 2012. *Bayesian computation using design of experiments-based interpolation technique.* Technometrics, 54(3):209–225.

33. Ryan E. G., C. Drovandi C. C., Pettitt A. N., 2015. *Simulation-based fully Bayesian experimental design for mixed effects models.* Computational Statistics & Data Analysis, 92:26 – 39. ISSN 0167-9473.

34. Schniger A., Whling T., Samaniego L., Nowak W., 2014. *Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence.* Water Resources Research, 50(12):9484–9513.

35. SiteChar, 2013. *Characterisation of storage sites: Best practices and Guidelines developed from SiteChar project.* SiteChar Report D2.4. 85 pages.

36. Tarakanov A., Elsheikh A. H., 2019. *Regression-based sparse polynomial chaos for uncertainty quantification of subsurface flow models*. Journal of Computational Physics, 399:108909. ISSN0021-9991. doi: https://doi.org/10.1016/j.jcp.2019.108909. URL http://www.sciencedirect.com/science/article/pii/S002199911930614X.

37. Vanlier J., Tiemann C.A., Hilbers P.A.J., Riel N.A.W. van. 2012. *A Bayesian approach to targeted experiment design.* Bioinformatics.

38. Webster R., Oliver M. A. 2015. *Basic Steps in Geostatistics: The Variogram and Kriging*. Springer, Cham. doi: https://doi-org.ezproxy1.hw.ac.uk/10.1007/ 978-3-319-15865-5.

39. Welling M., Kingma D. P., 2019. *An Introduction to Variational Autoencoders*. Foundations and Trends in Machine Learning. 12 (4): 307–392. arXiv:1906.02691

40. Wiener N., 1938. T*he homogeneous chaos.* American Journal of Mathematics, 60(4): 897–936.

41. Zhang J., Li W., Zeng L., Wu L. 2016. *An adaptive Gaussian process-based method for efficient Bayesian experimental design in groundwater contaminant source identification problems.* Water Resources Research, 52(8):5971–5984.

This deliverable is prepared as a part of ENOS project
More information about the project could be found at http://www.enos-project.eu

Be nice to the world!
Please consider to use and distribute this document electronically.

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 653718