Comput Geosci (2017) 21:267–287
DOI 10.1007/s10596-016-9611-2

ORIGINAL PAPER

Gaussian Processes for history-matching: application to an unconventional gas reservoir

Hamidreza Hamdi[1] · Ivo Couckuyt[2] · Mario Costa Sousa[1] · Tom Dhaene[2]

Received: 20 November 2015 / Accepted: 22 December 2016 / Published online: 21 January 2017

Abstract The process of reservoir history-matching is a costly task. Many available history-matching algorithms either fail to perform such a task or they require a large number of simulation runs. To overcome such struggles, we apply the Gaussian Process (GP) modeling technique to approximate the costly objective functions and to expedite finding the global optima. A GP model is a proxy, which is employed to model the input-output relationships by assuming a multi-Gaussian distribution on the output values. An infill criterion is used in conjunction with a GP model to help sequentially add the samples with potentially lower outputs. The IC fault model is used to compare the efficiency of GP-based optimization method with other typical optimization methods for minimizing the objective function. In this paper, we present the applicability of using a GP modeling approach for reservoir history-matching problems, which is exemplified by numerical analysis of production data from a horizontal multi-stage fractured tight gas condensate well. The results for the case that is studied here show a quick convergence to the lowest objective values in less than 100 simulations for this 20-dimensional problem. This amounts to an almost 10 times faster performance compared to the Differential Evolution (DE) algorithm that is also known to be a powerful optimization technique. The sensitivities are conducted to explain the performance of the GP-based optimization technique with various correlation functions.

Keywords Gaussian process · Expected improvement · Tight formation · History-matching

✉ Hamidreza Hamdi
hhamdi@ucalgary.ca

[1] Department of Computer Science, University of Calgary, Calgary, Alberta, Canada

[2] Department of Information Technology (INTEC), Ghent University-iMinds, Ghent, Belgium

Nomenclature

1D	One-dimensional
A	Non-dominated region
Bar sign "‾"	Average value
CCE	Constant Composition Experiment
c(**x**,**x'**)	Kernel or covariance function between two location **x** and **x'**
CGR	Condensate Gas Gatio
d	The dimension of problem
det[**C**]	Determinant of covariance matrix $\boldsymbol{C}$
$\mathbf{D}_n$	Training data set with n samples
$\mathbf{D}_{n+1}$	Augmented data set with $n+1$ samples
DE	Differential Evolution
DFIT	Diagnostic Fracture Injection Test
EI	Expected Improvement
EnKF	Ensemble Kalman Filter
ES	Ensemble Smoother
ES-MDA	Ensemble Smoother for Multiple Data Assimilation
°F	Degree Fahrenheit
f	The output of the truth function
f	A vector containing the output of the truth function in several locations
F	The output of the truth function as a random variable
F	The output of the truth function in several location as a random vector
GP	Gaussian Process

h	Reservoir thickness, ft
$k \times w$	Fracture conductivity, md ft
k_f	Current fracture permeability, md
k_i	Original (initial) fracture permeability, md
KB	Kelly bushing
l	Lateral length, ft
Ln(L)	Negative concentrated log-likelihood
M	Misfit function
n	The number of available samples (simulations)
p	Current pressure, psi
p_i	Initial pressure, psi
pr	Probability distribution
PR-EOS	Peng-Robinson Equation of State
q	Production flow rate, bbl/day (liquid) or MMscf (gas)
$Q_{\text{w\_r}}$	Remaining water in the reservoir after injection, ft^3
$\mathbf{r}$	The correlation vector between sample the $\mathbf{x}^*$ and the data $\mathbf{D}$
$\mathbf{R}$	Covariance matrix $= \sigma^2\mathbf{C}$
$S^2(\mathbf{x}^*)$	The variance of predicted valuey^* corresponding to sample $\mathbf{x}^*$ by GP
SRV	Stimulated Reservoir Volume
$S_{\text{w\_init}}$	Initial water saturation in the model
$S_{\text{w\_SRV}}$	Initial water saturation in the SRV
SRV	Stimulated Reservoir Volume
TVD	True Vertical Depth
$\mathbf{x}_i$	A sample i
x_f	Fracture half length, ft
Y	The posterior distribution of the modeled objective
$\hat{y}$	The predictive mean of the predictive GP
w	Fracture width, ft
W_{SRV}	The width of a 1D SRV,

Subscripts

b	The current best member
$n+1$	The augmented training data set by adding a new sample
o,g,w	oil, gas, water
obs	Observed data
sim	Simulation data

Greek letters

γ	Fracture reduction factor
δ	Molar composition of components in oil or gas
θ	GP hyperparameters (length scales) of dimension d
λ	An anisotropic distance measure
$\hat{\mu}$	The estimated mean of the GP model knowing data
μ	The prior mean of the GP model
ν	A constant used in defining *Matérn* correlation function
$\hat{\sigma}^2$	The estimated variance of the GP model knowing the data
φ	Porosity
ψ	The normal cumulative distribution function
ψ_{s}	The standard normal cumulative distribution function
ϕ	The normal probability density function
ϕ_s	The standard normal probability density function

1 Introduction

History-matching is a workflow necessary to alter the uncertain model parameters and to ensure that the final model(s) can satisfactorily mimic the dynamic response of a real reservoir. This process is usually performed in order to calibrate a reservoir model for a better production forecast. The subject reservoir models can be either average dynamic models such as well-test analytical models or more expensive numerical simulations with a large number of cells.

Traditional history-matching iterations are usually performed manually with the help of relying on an expert with a great knowledge of understanding the problem. However, with large model sizes and greater level of interactions between unknown parameters, such tasks become more cumbersome. Therefore, an automated or assisted history-matching technique is usually performed to alleviate the need for manual manipulation of the models in such iterative and costly processes. The approaches for assisted history-matching can be divided into two main categories consisting of data assimilation methods and optimization algorithms (e.g., gradient-based and derivative-free algorithms).

Ensemble Kalman Filter (EnKF) [1] is one of the most popular methods for data assimilation that runs in real time to recursively update the model every time the data is available. It is a sequential Monte Carlo algorithm that approximates the posterior distribution. Alternatively, in contrast to the iterative process of the EnKF for the sequential updating of the realizations, Ensemble Smoother (ES) [2, 3] is able to simultaneously assimilate the data in one go within a spatiotemporal setting. Different variants of ES such as iterative ES [4], and Ensemble Smoother with Multiple Data Assimilation (ES-MDA) [5] have also been introduced to enhance the performance of data assimilation method in field problems. The data assimilation methods use a Bayesian framework to combine the priors and observations, as well as the information about their uncertainty to provide the posterior distributions. Although sometimes

data assimilation methods can be defined in a minimization context [6], they are generally classified as uncertainty quantification techniques [7, 8]. In this paper, we look at history-matching as an optimization problem; hence, data assimilation methods are beyond the scope of this paper.

Gradient-based optimization techniques such as Gauss-Newton [9, 10] may not be very useful for many nonlinear complex problems. For those cases where the analytical expressions of the gradients do not exist, the numerical methods for finite differencing can be inaccurate particularly when the objective function is not smooth [11, 12]. Choosing an appropriate step size for differentiation is also difficult to determine because we cannot afford running costly simulations for several trials using various steps [13, 14]. Such algorithms are generally classified as local optimizers which cannot guarantee to find the global minima [15, 16]. On the other hand, the derivative-free methods such as Differential Evolution (DE) [17] frequently require a large number of simulations, which may not be feasible for larger-scale simulation models.

Meta-modeling, which is also known as proxy or surrogate modeling, is a promising approach towards finding the global extrema for low to medium dimensional inverse problems where only limited model evaluations can be performed [18, 19]. For this work, we use the Gaussian Process (GP) technique [20–24], also known as Kriging [25], to model the differences between the outputs of flow simulation models and the measured data (i.e., misfits). Methods such as Gaussian Process work particularly well with a few data points even in higher dimensions. Moreover, they can provide a prediction of the variance (model uncertainty), which has proven to be very informative in many optimization schemes [23, 26]. The purpose of building a proxy model (in this case a GP model) is to avoid unnecessary evaluation of the physical model, which can be very expensive. Instead, we try to build a cheaper model that conveys the typical behavior of the true physical model. At each iteration, the proxy model is used to carefully select new points. This method can considerably reduce the computation cost of the optimization since fewer expensive function evaluations are required for finding the global extrema.

GP-based optimization methods have been successfully applied to many nonlinear problems with low/medium dimensions [27–31]. However, the use of the GP modeling approach for history-matching of the reservoir engineering problems has been limited. Recently, Hamdi et al. [32] presented an example for numerical well testing of a faulted reservoir using a GP-based method. In this paper, we use the IC fault model [33] to highlight the efficiency of the GP-based methods in exploring and exploiting the misfit surface in search of global minima that is compared with the other methods such as quasi-Newton and DE algorithms. We also show the applicability of the GP-based optimization methods in a real fractured tight gas condensate well to estimate the model parameters. We use the example of a tight gas well because typically, a limited set of parameters are used in practice when performing production data analysis of such reservoir problems [34–37]. The aim is to use the compositional simulations of a simplified numerical model for characterization of the *in situ* fluid model and the well/reservoir parameters, which construct a 20-dimensional inverse problem. The compositional simulations are used to simulate the separator's fluid composition, hydrocarbon, and water rates in order to calculate the misfits using the corresponding measured data. At this stage, the objective function becomes a weighted sum of individual misfit values. We compare the performance of the GP-based optimization method with the Differential Evolution [38] algorithm as it was investigated in [39]. Furthermore, the impact of the initial design on the performance of the GP-based method is discussed and the effect of correlation function on the misfit convergence is investigated. The used single-objective optimization algorithm is freely available as part of the Surrogate Modeling (SUMO) toolbox, which can be downloaded from http://sumo.intec.ugent.be/SUMO.

2 Proxy modeling and prediction

In modeling computer experiments, we assume that the outputs of a truth mode (i.e., observations) that are gathered in a vector $\mathbf{f}$ can be regarded as realizations of a stochastic process. In general, there is no restriction on the type of the assumed stochastic process [40]. However, the use of the Gaussian distribution family provides a simple, convenient, and natural way to formulate the proxy modeling framework in an analytically tractable way [41]. More specifically, Mockus [41] showed that the use of Gaussian distribution is well suited for optimization tasks and can provide satisfactory results.

Gaussian Process is a probability distribution over an infinite number of random variables, such that the distribution over any finite number of which follows a multi-variate Gaussian distribution [20, 24]. In our case, the random variables represent the value of a function $f(\mathbf{x})$ at location $\mathbf{x}$. There is no assumption about the type of the objective function. It can be any function such as a misfit function in reservoir history-matching or any benchmark function with various local minima. The GP model defines the distribution over function values $f(\mathbf{x})$ which is then updated in light of training data. The updated GP model is used to predict the unseen test cases.

To represent the simulation misfit values with a GP model, we perform a limited number of simulations (n) by sampling from the d-dimensional reservoir parameter space

(initial design). The sample point is shown by $\mathbf{x}_i$ and the output is represented by $f_i = f_i(\mathbf{x}_i)$ This creates an initial training set $\mathbf{D}_n = \{(\mathbf{x}_i, f_i), i = 1, \ldots n\}$ where $\mathbf{x}_i$s create an $n \times d$ design matrix of the sample parameters and f_is build an $n \times 1$ column vector of the correlated misfit values ($\mathbf{f}$). The GP is used for modeling the unknown simulation misfit function. For this finite set of observations, the GP model is simply a joint Gaussian distribution $\mathbf{F} \sim N(\mu\mathbf{1}, \mathbf{C})$, where the process mean μ is a constant (but unknown) and $\mathbf{1}$ is an $n \times 1$ column vector of ones. For the ease of notational convenience, the uppercase letters are used to represent the random variables. The vector or matrix quantities are shown in bold throughout this manuscript.

A GP model is completely characterized by its mean and its covariance matrix $\mathbf{R} = \sigma^2 \times \mathbf{C}$ (σ^2 is the process variance.) In this paper, a constant-mean GP model is used to represent the truth model. This can be seen as an ordinary Kriging model as described in geostatistics [25] where the mean function is an unknown constant and needs to be estimated. Each entry (i, j) of the kernel (or covariance) matrix is obtained by a covariance function $c(\mathbf{x}_i, \mathbf{x}_j)$. In this work, the stationary Matérn correlation function [46] with $\nu = 3/2$ is adapted to fill the kernel matrix, that is:

$$c\left(\mathbf{x}, \mathbf{x}'\right)_{\upsilon=3/2} = \sigma^2 \left(1+\sqrt{3}\lambda\right) e^{-\sqrt{3}\lambda} \tag{1}$$

$$\lambda = \sqrt{\sum_{m=1}^{d} \theta_m \left(\mathbf{x}_m - \mathbf{x}'_m\right)^2} \tag{2}$$

in which, $\mathbf{x}$ and $\mathbf{x}'$ are any two sample locations (inputs), and λ is the anisotropic distance measure with d hyperparameters θs. For the (second order) stationary random processes, the mean is constant and the covariance is only a function of the distance between the locations [25]. The goal of representing the outputs using a proxy (GP) is to be able to predict the truth model response at any untrained location without the need to run the truth model. Such predictions can be made using ordinary Kriging by employing a frequentist [23, 40] or a Bayesian approach [42, 43]. The frequentist analysis of an (ordinary) Kriging predictor is based on a linear combination of the observed values where the Maximum Likelihood Estimation (MLE) is used to estimate the Kriging model's parameters (e.g., [μ, σ^2, θs]). The Bayesian approach, on the other hand, allows us to specify a prior on the Kriging parameters and derive the full conditional posterior distribution.

2.1 Frequentist analysis of ordinary Kriging

To predict the function value at an unsampled location $\mathbf{x}^*$, the Kriging predictor $\hat{y}(\mathbf{x}^*)$ is written as a linear combination of the observed data, i.e., $\hat{y}\left(\mathbf{x}^*\right) = \sum_{i=1}^{n} a_i f_i$ where a_is are the weights and f_is are the observed values. In classical frequentist approach, we can treat the predictor and the observed data as random values i.e. $Y\left(\mathbf{x}^*\right) = \sum_{i=1}^{n} a_i F_i$. After using the random counterpart of our data, we make the unbiasedness assumption, that is $E[Y(\mathbf{x}^*) - F(\mathbf{x}^*)] = 0$, in which $F(\mathbf{x}^*)$ is the truth response of the function. The unbiasedness assumption amounts to having a constant mean over the random process (stationary). The optimum weights (a_is) are obtained by minimizing the Mean Square Error (MSE) $E[a_i F_i - F(\mathbf{x}^*)]^2$. Therefore, the Kriging predictor is called Best (because of minimizing the variance) Linear (because of its linear combination form) Unbiased (because of the constant mean assumption) Predictor which is simply called BLUP [25, 40, 44]. The optimal Kriging weights can immediately give (1) the Kriging predictor $\hat{y}(\mathbf{x}^*)$ and (2) the associated (minimized) MSE or Kriging variance denoted by $s^2(\mathbf{x}^*)$ according to these equations:

$$\hat{y}\left(\mathbf{x}^*\right) = \mu + \mathbf{r}^{\mathbf{T}}\mathbf{C}^{-\mathbf{1}}\left(\mathbf{f} - \mathbf{1}\hat{\mu}\right) \tag{3}$$

$$s^2\left(\mathbf{x}^*\right) = \sigma^2\left[1 - \mathbf{r}^{\mathbf{T}}\mathbf{C}^{-\mathbf{1}}\mathbf{r} + \frac{\left(1 - \mathbf{r}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{r}\right)^{\mathbf{2}}}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}}\right] = \sigma^2 \mathrm{V}_\theta \tag{4}$$

where $\mathbf{r}$ is the correlation vector between the predicted point $Y(\mathbf{x}^*)$ and the observation vector $\mathbf{F}$ assuming the same Matérn correlation function with $\nu = 3/2$. The second term in the MSE equation ($\mathbf{r}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{r}$) has a negative sign that indicates a reduction in uncertainty as the Kriging predictor is correlated with the observations, while the last term ($\frac{\left(1-\mathbf{r}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{r}\right)^{\mathbf{2}}}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}}$) has a positive sign which reflects an increased uncertainty in prediction as μ is unknown and needs to be estimated.

In Eqs. 3 and 4, the process mean (μ) and variance (σ^2) are estimated by maximizing the Gaussian likelihood function. The corresponding *estimated values* for mean and variance are shown by $\hat{\mu}$ and $\hat{\sigma}^2$ that can be analytically obtained as follows:

$$\hat{\sigma}^2 = \frac{\left(\mathbf{f} - \mathbf{1}\hat{\mu}\right)^{\mathrm{T}} \mathbf{C}^{-\mathbf{1}} \left(\mathbf{f} - \mathbf{1}\hat{\mu}\right)}{n} \tag{5}$$

$$\hat{\mu} = \frac{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{f}}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}} \tag{6}$$

In Eq. 5, n is often replaced by $n - 1$ to take into account the loss in the degrees of freedom associated with estimating μ [45]. Without having an estimation of the hyperparameter vector ($\boldsymbol{\theta}$), we are unable to calculate $\mathbf{C}$,$\hat{\mu}$, and $\hat{\sigma}^2$ as they all depend on $\boldsymbol{\theta}$. Therefore, the hyperparameters (θs) also need to be estimated. Plugging the estimated mean ($\hat{\mu}$) and variance ($\hat{\sigma}^2$) into the Gaussian likelihood function, we come up with a new expression which is called *the concentrated*

likelihood function that only depends on the hyperparmeters (θs). In practice, the hyperparameters are estimated by minimizing the negative concentrated log-likelihood [20], which is defined as:

$$\ln(L) = -\frac{n}{2}\ln\left(\hat{\sigma}^2\right) - \frac{1}{2}\ln(\det[\mathbf{C}]) \tag{7}$$

in which det[**C**] is the determinant of matrix **C**. For practical purposes, we use the Matlab™ "fmincon" optimization function [46] to minimize ln(L). Therefore, to use Eqs. 3 and 4, we first need to estimate the hyperparameters by minimizing (7); then, we are able to compute the estimated mean and variance (5) and (6). Having estimated all Kriging parameters, we can now use Eqs. 3 and 4 to define a posterior Gaussian distribution for predicting the misfit value (or the objective function) in any untrained location $\mathbf{x}^*$, i.e., $Y(\mathbf{x}^*) \sim N(\hat{y}(\mathbf{x}^*), s^2(\mathbf{x}^*))$.

2.2 Bayesian interpretation of ordinary Kriging

In Bayesian framework approach, the GP is used as *a prior* to compute a predictive (*posterior*) distribution for making prediction for unseen function values $Y(\mathbf{x}^*)$ in light of training data **F** [20, 26]. The reader should bear in mind that this inference takes place only for the GP model parameters (e.g., [σ^2, μ, $\boldsymbol{\theta}$]), not the parameters of the underlying physical model such as porosity and permeability.

Following the properties of GP, any finite set of the function outputs have a joint Gaussian distribution. Therefore, our prediction $Y(\mathbf{x}^*)$ is jointly distributed with the observations, i.e.,

$$\begin{pmatrix} \mathbf{F} \\ Y(\mathbf{x}^*) \end{pmatrix} \sim N\left(\begin{bmatrix} \mu\mathbf{1} \\ \mu \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{C} & \mathbf{r} \\ \mathbf{r}^{\mathbf{T}} & 1 \end{bmatrix} \right) \tag{8}$$

Using the conditional properties of Gaussian distributions and after factorizing the joint distribution using the Schur complement [20, 47, 48], we can obtain the conditional distribution (or the posterior) in a closed form as [24, 48, 49]:

$$Y^*|\mathbf{F}, \sigma^2, \mu, \boldsymbol{\theta} \sim N\left(\mu + \mathbf{r}^{\mathbf{T}}\mathbf{C}^{-1}(\mathbf{f} - \mu\mathbf{1}), \sigma^2\left[1 - \mathbf{r}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{r}\right]\right) \tag{9}$$

This estimation is similar to the simple Kriging predictor in geostatistics where the GP parameters are known. In order to have similar results as in the frequentist analysis of ordinary Kriging, we have to use Jeffrey's non-informative improper prior distributions [50] for μ and σ^2 that is pr(μ) $\propto 1$ for $-\infty < \mu < \infty$ and pr(log σ^2)$\propto 1$ (or pr(σ^2) $\propto 1/\sigma^2$) for $0 < \sigma^2 < \infty$ (see chapters 1 and 2 in [51]). The improper distributions are used to reflect the fact that we have little knowledge on those parameters. They are called improper as the integral of their density distributions are infinite. Therefore, the joint distribution pr(σ^2, μ, $\boldsymbol{\theta}$)$\propto$pr($\boldsymbol{\theta}$)/σ^2 serves as *a prior* distribution for the inference.

For ordinary Kriging with an unknown mean, we are interested to compute the marginal posterior of Eq. 9 over μ, i.e., $Y^*|\mathbf{F}, \sigma^2, \boldsymbol{\theta}$ as:

$$\text{pr}(y^*|\mathbf{f}, \sigma^2, \boldsymbol{\theta}) = \int_{-\infty}^{\infty} \text{pr}(y^*|\mathbf{f}, \sigma^2, \mu, \boldsymbol{\theta}) \times \text{pr}(\mu|\mathbf{f}, \sigma^2, \boldsymbol{\theta}) d\mu \tag{10}$$

Following the Zellner's approach [43, 51] the posterior probability of the mean is a normal distribution, i.e., pr($\mu|\mathbf{f}$, σ^2, $\boldsymbol{\theta}$)$\sim N(\hat{\mu}, \sigma^2(\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1})^{-1})$. Knowing $Y^*|\mathbf{F}$, μ, σ^2, $\boldsymbol{\theta}$ from Eq. 9 (i.e. simple Kriging), it can be shown that the marginal posterior distribution (10) is also a Gaussian distribution, i.e., $Y^*|\mathbf{F}, \sigma^2, \boldsymbol{\theta} \sim N(\hat{y}(\mathbf{x}^*), \sigma^2 V_\theta)$. This is similar to the result obtained for the ordinary Kriging using the frequentist approach (3) and (4).

By the same token, we can now remove the condition on σ^2 by marginalization of Eq. 10 to find $Y^*|\mathbf{F}$, $\boldsymbol{\theta}$. The conditional posterior distribution of σ^2 is in the form of an inverted-Gamma distribution [51], i.e., pr($\sigma^2|\mathbf{f}$, $\boldsymbol{\theta}$)$\propto \sigma^{-(n+1)}\exp[-(n-1)\hat{\sigma}^2/2\sigma^2]$. In the same manner, the predictive marginal distribution of Y^* given the covariance hyperparameters can be calculated as:

$$\text{pr}(y^*|\mathbf{f}, \boldsymbol{\theta}) = \int_0^{\infty} \text{pr}(y^*|\mathbf{f}, \sigma^2, \boldsymbol{\theta}) \times \text{pr}(\sigma^2|\mathbf{f}, \boldsymbol{\theta}) \text{d}\sigma^2 \tag{11}$$

Handcock and Stein [42] analytically calculated this integral and realized that this conditional distribution is in the form of a shifted Student's t distribution, i.e., $Y^*|\ \mathbf{F}, \boldsymbol{\theta} \sim t_{n-1}(\hat{y}(\mathbf{x}^*), \hat{\sigma}^2 V_\theta)$. Finally, we can try to remove the condition on $\boldsymbol{\theta}$ to find $Y^*|\mathbf{F}$. However, in the latter case, the marginalization on $\boldsymbol{\theta}$ is extremely complicated and requires numerical methods such as Markov chain Monte Carlo (McMC) to sample from the predictive distributions [42, 49].

It should now be clear that making a Gaussian assumption can largely simplify the inference and estimation problem as many formulations can take simplified analytical forms. For example, in the case of non-Gaussian likelihood training the GP can get complicated and might require some transformations [24]. If we do not make the Gaussian assumptions, the predictor $Y(\mathbf{x}^*)$ given the data (i.e., $E[Y(\mathbf{x}^*)|\mathbf{F}]$) may not be linear and therefore the linear frequentist approach may not serve as a good predictor for $Y(\mathbf{x}^*)|\mathbf{F}$ [45]. The frequentist linear predictor is a limit of Bayesian approach when we use a GP as the stochastic model, and also when we use improper uniform distributions as priors for μ and σ^2. Although it is obvious that the Bayesian approach is a thorough method to address the uncertainty, in practice, due to the complexity of the numerical methods involved in the Bayesian approach (i.e., McMC), the frequentist approach is preferred for the modeling tasks [25].

3 Expected improvement

After building a proxy model using the observed values of the objective function (e.g., the misfit values of the reservoir simulation), we need further analysis to be able to use the proxy model for the optimization task. In particular, we need to use the proxy model to realize the good locations that most likely have lower objective functions. If we realize such locations, we can get more samples from these areas and run the actual simulator to see the outcome. This procedure can be repeated many times which can result in more accurate proxy models for the purpose of optimization. However, the good locations for the truth model (global minima) do not necessarily reside in the areas that minimize our proxy model. This is because we have an uncertainty associated with our fitted surface (i.e., Kriging variance). Such a naive minimization idea can easily jeopardize the global minimization trial. For instance, in Fig 1 (left), the initial design has failed to sample the region where the global minimum is located (i.e. $0.6 < x < 0.8$). If we try to minimize the Kriging function itself without any additional criteria we come up with a proposal location at $x = 0.3$ (Fig. 1, left), which only helps exploit the space and get closer to a possible local minimum around $x = 0.2$ (Fig 1, right). Because we have neglected to incorporate the uncertainty of the fitted surface into the optimization procedure, no matter how many times we repeat this algorithm we cannot find the global minimum. Therefore, we have to define some additional sampling criteria to take into account the uncertainty of the fitted surface and make a balance between the exploration and the exploitation of the parameter space in the search for the global minimum.

Expected improvement (EI) or $E(I(\mathbf{x}^*))$ is a popular and efficient sampling criterion, also known as infill criterion or acquisition function, which quantifies the amount of improvement that is achieved by sampling from any location $\mathbf{x}^*$ [22, 52]. The improvement is measured with respect to the current best sample $\mathbf{x}_b$ having a minimum value of f_b. As it was shown above, the uncertainty about any unseen location $\mathbf{x}^*$ can be represented by the GP predictor which is a random variable Y with mean $\hat{y}(\mathbf{x}^*)$ and standard deviation $s(\mathbf{x}^*)$. The statistical improvement function is defined as $\max(f_b - Y, 0)$. The improvement is measured with respect to the current best value of the objective function (f_b). In the EI infill criterion, we try to maximize the expected value of the improvement as:

$$E(I(\mathbf{x}^*)) \equiv E\left[max(f_b - Y(\mathbf{x}^*), 0)\right] \tag{12}$$

The closed form of the expected improvement can be expressed as [22]:

$$E\left(I(\mathbf{x}^*)\right) = s(\mathbf{x}^*)u\phi_s(u) + s(\mathbf{x}^*)\psi_s(u) \tag{13}$$

where $u = (f_b - \hat{y}(\mathbf{x}^*))/s(\mathbf{x}^*)$, and ϕ_s and ψ_s are the standard normal cumulative distribution and density functions respectively, i.e.,

$$\phi_s\left(u = \frac{f_b - \hat{y}(\mathbf{x}^*)}{s(\mathbf{x}^*)}\right) = \left[\frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{f_b - \hat{y}(\mathbf{x}^*)}{\sqrt{2}s(\mathbf{x}^*)}\right)\right] \tag{14}$$

$$\psi_s\left(u = \frac{f_b - \hat{y}(\mathbf{x}^*)}{s(\mathbf{x}^*)}\right) = \frac{1}{s(\mathbf{x}^*)\sqrt{2\pi}}\exp\left[-\left(\frac{f_b - \hat{y}(\mathbf{x}^*)}{\sqrt{2}s(\mathbf{x}^*)}\right)^2\right] \tag{15}$$

The first term in Eq. 13 is larger than the second term when $\hat{y}(\mathbf{x}^*)$ is small, while the second term is larger than the first term when $s(\mathbf{x}^*)$ is large. In other words, the first term promotes the exploitation, and the second term promotes the exploration. Therefore, the EI has a way to balance exploration and exploitation for the optimization task which it turns to be very effective in practice.

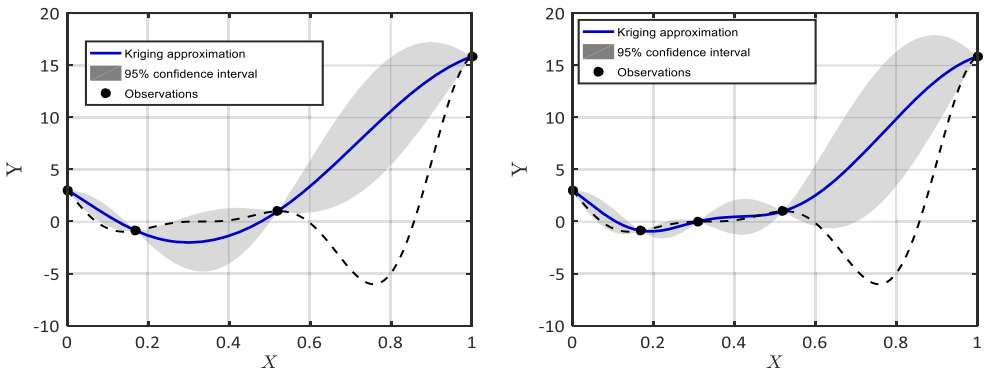

Fig. 1 The minimization of the fitted GP (Kriging) model without any infill criteria at iterations i (*left*) and $i + 1$ (*right*). Without any infill criteria, the uncertainty of the GP model is not taken into account for the sampling which consequently leads to a failure in finding the global minimum at $x = 0.75$. The *black dashed curve* is the truth function, and the *blue curve* is the fitted GP model. The figures were produced using UQLab [143, 144]

4 The GP-based optimization workflow

The application of GP for optimization (or reservoir history-matching) is a multi-stage workflow. Figure 2 shows a flowchart of the GP-based optimization algorithm that is used in this study. First, we generate a limited number of samples (model parameters) from the parameter space (initial design). The actual reservoir simulations are then performed to calculate the corresponding misfit values by comparing the reservoir simulation outputs with the measured data. Assuming a GP prior as a distribution over observed data ($\mathbf{D}_n$), we obtain a predictive (posterior) distribution $Y(\mathbf{x}^*) \sim N(\hat{y}(\mathbf{x}^*), s^2(\mathbf{x}^*))$ using Eqs. 3 and 4. The GP model parameters are estimated by the MLE approach as in Eqs. 5, 6 and 7 (a primary optimization stage). Secondly, using the predictive GP, the EI infill criterion (13), (14) and (15) is maximized to propose a new sample location ($\mathbf{x}^{**}$) which is likely to have a lower misfit value than the current best sample in our data $\mathbf{D}_n$ (a secondary optimization stage). Differential Evolution (DE) is used to maximize the EI criterion. This potential optimum point ($\mathbf{x}^{**}$) is passed to the actual simulator and the corresponding misfit value $f(\mathbf{x}^{**})$ is obtained. The new point ($\mathbf{x}^{**}$, $f(\mathbf{x}^{**})$) is then added to augment the available data ($\mathbf{D}_{n+1}$) which is also used to update the surrogate model's accuracy for the next iterations. This procedure is repeated until some stopping criteria are reached.

Figure 3 portrays the ability of the EI infill criterion in conjunction with the predictive GP to find the minimum of a 1D function in different iterations. In this figure, black circles represent a few observation points that are sampled from the truth function that is shown with a black dashed line. At iteration 0 ($i = 0$), the expected improvement (green curve) promotes the exploration of the parameter space as the highest pick of the EI function residing in an area where we have the largest uncertainty (the shaded region represents the uncertainty of the proxy model that is a 95 % confidence interval obtained from the Kriging error). Therefore, the next point is suggested at around $x = 1.5$. The process is repeated for a few iterations and, as a result, the Kriging model (blue curve) is sequentially updated. At iteration 8 ($i = 8$), although the EI is high at $x = -3$ (with a large prediction error), the expected improvement's pick is close to the points with lowest $\hat{y}$, which promotes the exploitation of the space. Eventually, the global minima is found at $x = 0$ in iteration 9.

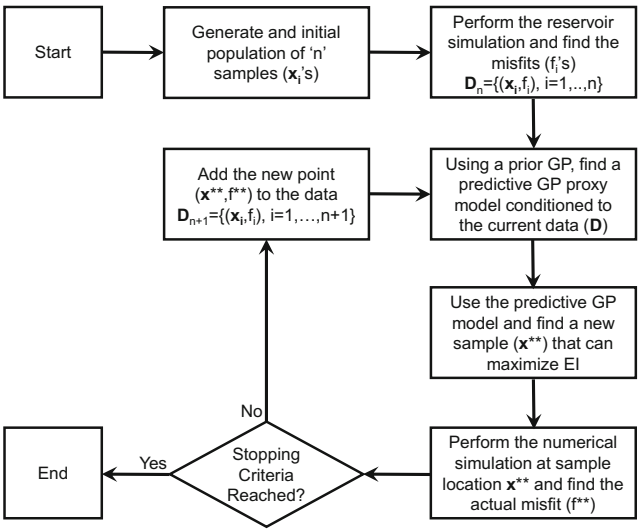

Fig. 2 A flowchart of the GP-based history-matching used in this study

4.1 A note on the dimensionality of the GP-based optimization method

Bull [53] gives explicit formulas for the convergence rate of GP-based optimization. Under some assumptions, he proves a convergence rate of $O(n\hat{}(-(v)/d))$ with n being the sample number, v the smoothness factor of the Matérn correlation function, and d the dimension of the problem. Using this, we can see the dimensionality scales inversely ($1/d$ in the exponent) with the number of required iterations. When solving a problem one may plug in a d and a v to get a function $f(n)$ for getting a rough idea on how fast a problem could converge. However, the current GP-based optimization approach that is presented in this paper should not be seen as a universal approach for all types of history-matching problems with large dimensions (>100). This is an optimization technique for handling low to medium dimension problems with expensive functions (simulations) where only limited number of function evaluations can be performed. As such, it can be combined with some methods such as gradual deformations [54] or probability perturbation [55] to formulate a history-matching problem by defining a geological process rather than large number of unknown cell values [56–58]. GP-based optimization methods, like most proxy modeling techniques, are limited by the curse of dimensionality. In practice, GPs have been applied to problems up to about 30 variables and most of the time to problems with only around 10 variables [59] for engineering design optimization, calibration, etc. [18]. It is important to note that, the number of variables is not necessarily a problem for GPs. The most difficult challenge is that introducing more variables indirectly increases the required size of the data. While recent GP techniques exist to scale up to more data, they often sacrifice accuracy and have not yet been tested for optimization purposes [60]. In addition, there are other challenges to overcome as well—such as optimizing the expected improvement criterion in thousand dimensions or optimizing the many hyperparameters of a high-dimensional GP model. However, the subject of GP

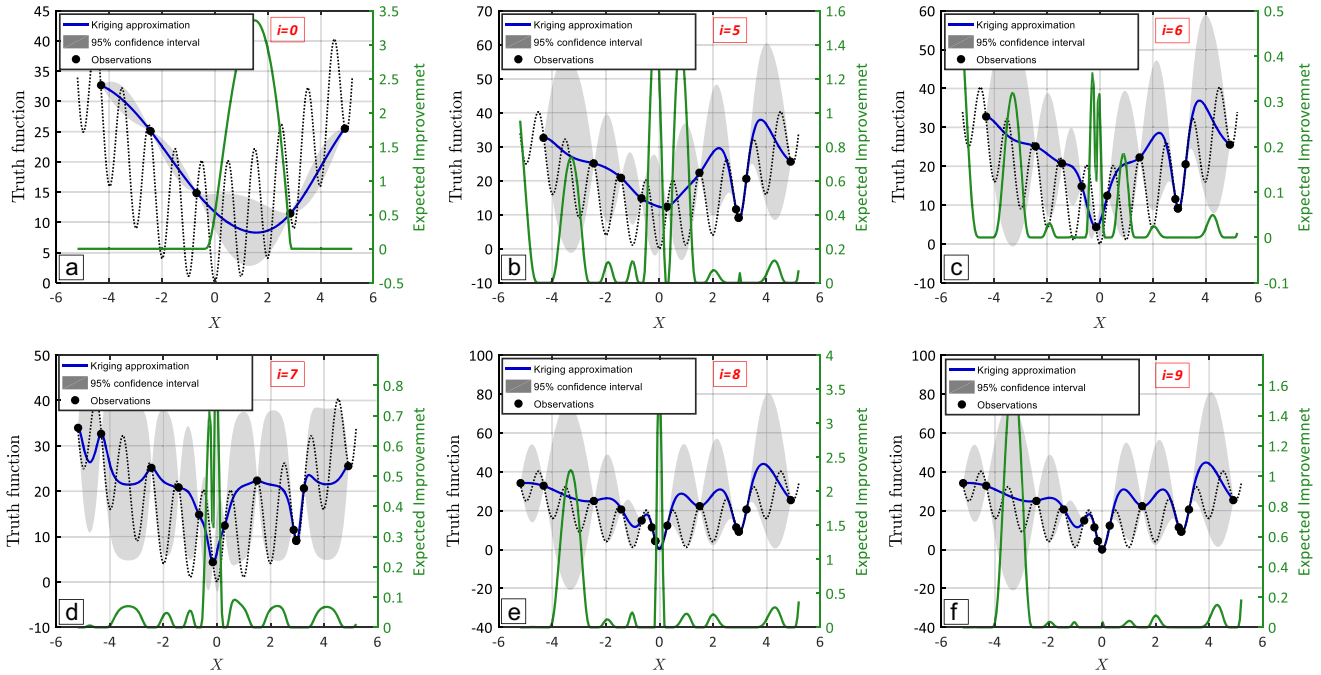

Fig. 3 The GP-based optimization using EI in different iterations (**a**–**f**) to find the minimum of a multi-modal function (*black dashed curve*). The *blue curve* is the fitted GP model with its associated prediction uncertainty (the *grey region*), the *green curve* is the EI function and the *circles* are the observed data at each iteration. The figures were produced using UQLab [143, 144]

is an active subject in global optimization and engineering designs and new variants of GP-based optimization are frequently introduced which seek to improve the performance of GP-based optimization to even ensure an *exponential* convergence rate without any internal optimization overhead [47]. Meanwhile, if the intrinsic dimensionality is low, sensitivity analysis and dimension reduction techniques can be used to reduce the number of variables in [61]. For instance, recently, Wang et al. [61] presented a method to use a variant of GP in a billion dimensions with low intrinsic dimensionality. Certainly, one of the most important challenges in optimization and history-matching problems is the large number of variables. However, finding the global optima efficiently for low-dimensional problems, which are often nonlinear and computationally expensive, are also equally important and challenging (see the example of IC fault model with 1 to 13 parameters in [62]). For large dimensional problems, a variety of other history-matching techniques exist including gradient- and ensemble-based methods [1, 5, 8, 63–68].

4.2 Possible routes for the extension of GP-based optimization method for multiple poro-perm realizations

If we use the GP-based optimization algorithm and perform the optimization for one realization with one set of porosity and permeability distributions, the final proxy model (i.e., at the end of optimization) which is obtained for this realization cannot be used to mimic the performance of a different realization with different porosity and permeability distributions. This is evident as different spatial distributions for the porosity and permeability can create different responses. However, this does not imply that the GP is inefficient. The purpose of GP-modeling in essence is not to replace a reservoir model but it is to help minimize an objective function with a minimum number of evaluations which is very important for many problems such as parameter estimation [69–71]. We should also note that as per the "no free lunch theorem" for optimization and search problems [72], finding a universal history-matching approach is theoretically impossible [8]. Nonetheless, we can think of numerous directions that the current workflow can be potentially extended for use in more complex optimization tasks involving geology:

1. The GP-based optimization is used to sequentially sample from the parameter space to minimize the objective function in limited number of simulation. However, there have been some work on parallel [48] hybrid and batch optimization using GPs [26], which can further improve the optimization cost.
2. The GP can be combined with any functional parameterization algorithm such as probability perturbation [73] to obtain the history-match models by changing the spatial distributions of the facies which are controlled

by a few global parameters to change the probabilities [58]. In addition, GP or in general, proxy-based optimization techniques might be used for different realizations if they can be combined with the other techniques such as Principal Component Analysis (PCA) [74] or Multi-Dimensional Scaling (MDS) [75]. These are used to project different geological models into a lower space to reduce the model dimensionality.

3. The ability of GP or proxy-based optimization methods to balance between the exploration and the exploitation can be used in the same way as the neighborhood algorithm [76] works, where a large quantity of initial models are produced and are projected to smaller dimensions [7]. Then the proxy-based approach can be used to explore the model space and find the regions (realizations) with lower misfits.

4.3 The efficiency of the GP-based optimization approach

The GP-based optimization approach is largely favorable for global optimization when we have limited resources and cannot afford running several costly numerical simulations. The gradient-based approaches are efficient in finding the local solutions particularly, and the global solutions when the objective function is relatively smooth. However, for situations that the objective functions have many local minima, the downhill trend of the gradient-based approaches would not necessary lead to a global solution. Besides, as mentioned earlier, the population-based methods generally require a large number of function evaluations to converge. In [77], a comparison is made between the GP-based optimization and the Sequential Quadratic Programming (SQP; based on Broyden-Fletcher-Goldfarb-Shanno (BFGS)) on several applications, and [78] compared the performance of a GP-based optimization algorithm with other global search methods such as genetic algorithms and pattern search. For those, the GP-based optimization showed superior performance. In this paper, we also try to compare the performance of the GP-based optimization with two famous members of the gradient-based and the derivative-free stochastic search methods.

(a) Quasi-Newton methods Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is a quasi-Newton approach for approximation of Hessian matrix in nonlinear optimization problem [79]. BFGS is one the most popular quasi-Newton approaches for solving the optimization problems [80, 81]. However, sometimes, when the initial Hessian is not properly chosen or the search direction is poorly defined, the standard BFGS might take a larger number of iterations to converge in some problems [82]. Self-scaling BFGS methods [83–87] are then developed to overcome these challenges and to improve the convergence rate in the unconstrained optimization context. In this paper, as suggested by Zhang and Reynolds [88] we use Oren and Spedicato's method [87] for scaling the Hessian in each iteration [80, 89].

(b) Differential Evolution In the context of derivative-free numerical optimization, DE is a simple, effective and robust population-based evolutionary algorithm for the global optimization [17, 90–92]. It has outperformed many other algorithms such as particle swarm optimization in numerous benchmark problems [93] and has been amongst frontier global optimization methods in numerous completions and contests [94–96]. As such, it has been successfully applied to many realworld problems such as financial markets modeling [97], data mining [98] and reservoir engineering [57, 99–101] among many others. DE has been implemented in major commercial reservoir simulation packages such as tNavigator.

Some studies showed that the performance of classical DE is sensitive to its setting parameters (crossover and scaling factor) and a special combination of the parameters are required for a favorable converge [102]. Hence, some empirical work focused on the selection of setting parameters to ensure the global convergence [103, 104]. Recently [105] showed that classical DE cannot guarantee the global convergence for a special category of functions where the global optimum is close to the boundaries with a larger region on the other side of the global optimum. In another study, Hu et al. [92] proved the sufficient conditions for the global convergence of classical DE and provided an algorithm that can satisfy such conditions. To overcome the challenges with parameter selection of classical DE algorithms, which might cause the DE to fall in local optima or have a premature convergence, many recent studies were conducted on developing improved and adaptive DE algorithms [90, 91, 106, 107] which resulted in promising results. In this study, we use the recent adaptive DE algorithm proposed by Gong et al. [107] which showed superior global convergence for a large category of benchmark functions. In this section, the need for finding the global minimum of a function with limited function evaluations is exemplified using the IC fault model [33].

The IC fault model is a 2D cross-sectional reservoir model of a layered reservoir with a total thickness of 60 ft and 100 × 200 cells. Each grid block is 10ft wide. The model has two segments that are separated with a single vertical fault and is comprised of six layers with alternating good and poor quality rocks. An injection well is placed in the left-hand edge and a production well is producing oil from the right-hand edge and both wells are operated by the bottom hole pressure This reservoir model is described by only three parameters: a high quality sand permeability

(k_{high}), a low quality permeability (k_{low}) and a fault throw (h) (Fig. 4). The fault throw controls the flow across the fault by changing the contact between the high and low quality sand layers. A truth case is generated by choosing $h = 10.4$ ft, $k_{\text{high}} = 131.7$ md and $k_{\text{low}} = 1.31$ md, where the values of these three parameters are drawn independently from uniform distributions as $h \in [0,60]$, $k_{\text{low}} \in [0,50]$ and $k_{\text{high}} \in [100,200]$. A full description of this model can be found in [33]. The misfit function for the IC fault model is defined by

$$M = 0.5 \sum_{p \in \{o,w\}} \sum_{t=1}^{T} \frac{\left(q_p^{\text{obs}} - q_p^{\text{sim}}\right)_t^2}{\sigma_t^2} \tag{16}$$

where T is the number of observations, q_p represents oil and water rates, sim and obs indicate the simulation and observed data respectively and $\sigma = 0.03 \times q_p^{\text{obs}}$ represents the measurement error. The IC fault is a difficult history-matching problem with several steep minima in the objective function. The full discussion of the objective function can be found in [33, 62]

For history-matching the IC fault model, the initial samples are generated using a Latin-Hypercube design with 20 members. In each individual optimization repetition, a similar set of initial samples is used for both DE- and GP-based optimization methods, the best of which (with minimum misfit value) is used as the starting point for the BFGS method. Figure 5 shows the performance of various optimization methods. The figure shows that the quasi-Newton method has a negligible improvement in the function value with an almost plateau around the misfit value of 300 after around 100 simulations (a possible local minimum). DE is converging slowly towards the global minimum with a larger number of evaluations, while the GP-based optimization method provides the lowest misfit of around 17 in less than 100 function evaluations.

The faster convergence rate of the misfit function suggest the potential effectiveness of this surrogate-based method in reservoir history-matching for the cases where the misfit evaluation in a single simulation run could perhaps take several minutes to hours consequently and hence the number of affordable numerical simulation runs is drastically low. However, it is noted that if the objective function is computationally cheap, the GP-based algorithm may not be very efficient as the computational time from the optimization itself can be more than doing function evaluations. For exact inference using Cholesky decomposition [20], the computational complexity of a GP depends directly on the size of the data set and it scales as $O(n^3)$ [108], with n being the number of data points. For a limited number of data points, such as when dealing with expensive simulations, constructing a GP model takes well under 1 s for smaller samples or within 1 min (around 500 samples). As mentioned above, the domain of GP-based optimization is being rapidly developed and many new algorithms are introduced. We have tested a GP algorithm [47] without the need for an internal optimization which could tremendously reduce the optimization cost of a 20-dimensional problem for less than a few minutes. Using isotropic correlation function for some problems can also substantially reduce the computation associated with the hyperparameters estimations. Increasing the size of the training data points, can cause the

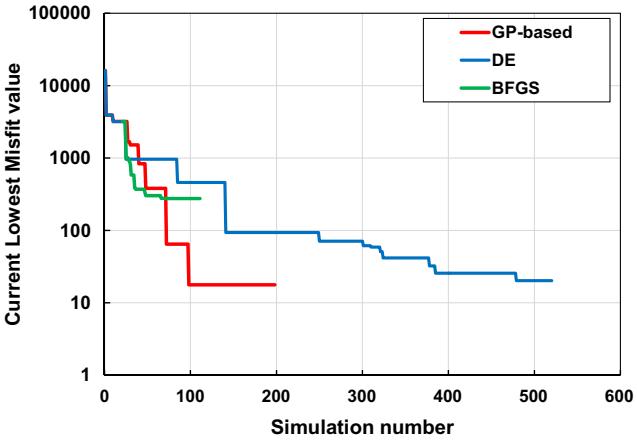

Fig. 5 The average performance of DE, quasi-Newton, and GP-based methods in minimizing the IC fault model

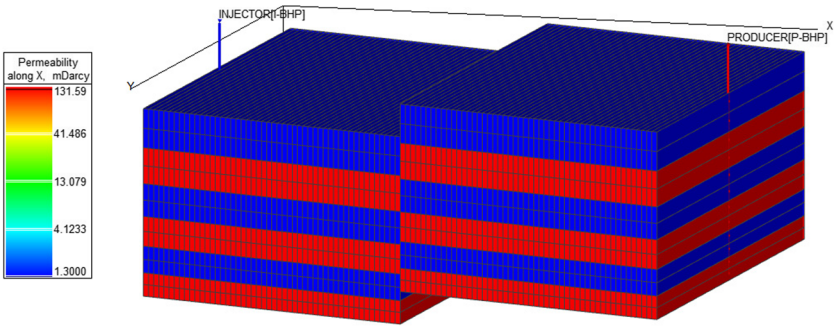

Fig. 4 IC fault model

GP models to become slower. In such cases, some approximation methods exist (on the CPU as well as the GPU) which can trade off the accuracy versus the computational cost. These can reduce the computation complexity with $O(n^2)$ [109] or even $O(n)$ [110, 111]. Nevertheless, in the context of expensive simulations, the number of data points is inherently low and the speed is not often a big issue for many applications.

Although in this study we were interested in global optimization, finding local optima by the gradient-based methods can be a useful characteristic for uncertainty quantification. For instance, using multiple initial points based on a prior Gaussian, one might be able to identify all the modes for some problems, which can be the motivation for some methods such as the gradient-based randomized likelihood; see Oliver et al. [64]. Depending on the statistical criterion used, there is also some work on the convergence properties of GP-based optimization methods [112], which can prove that, given enough samples, it will cover the search space uniformly by finding all global and local optima. The convergence rates are also given in [53]. Nevertheless, it should be noted that such work is theoretical in nature, and depends on how good the expected improvement criterion can be optimized as the required number of samples can grow large.

5 Application example

Triassic Montney Formation is a low-energy hydrocarbon-rich Formation located within the Western Canadian Sedimentary Basin across Alberta and British Columbia, Canada [12]. The Formation consists of various facies ranging from conventional sandstone reservoirs in the east to tight siltstones and shales in the west (Fig. 6, left) [113]. The fluid distribution trend map (Fig. 6; right) also shows a great

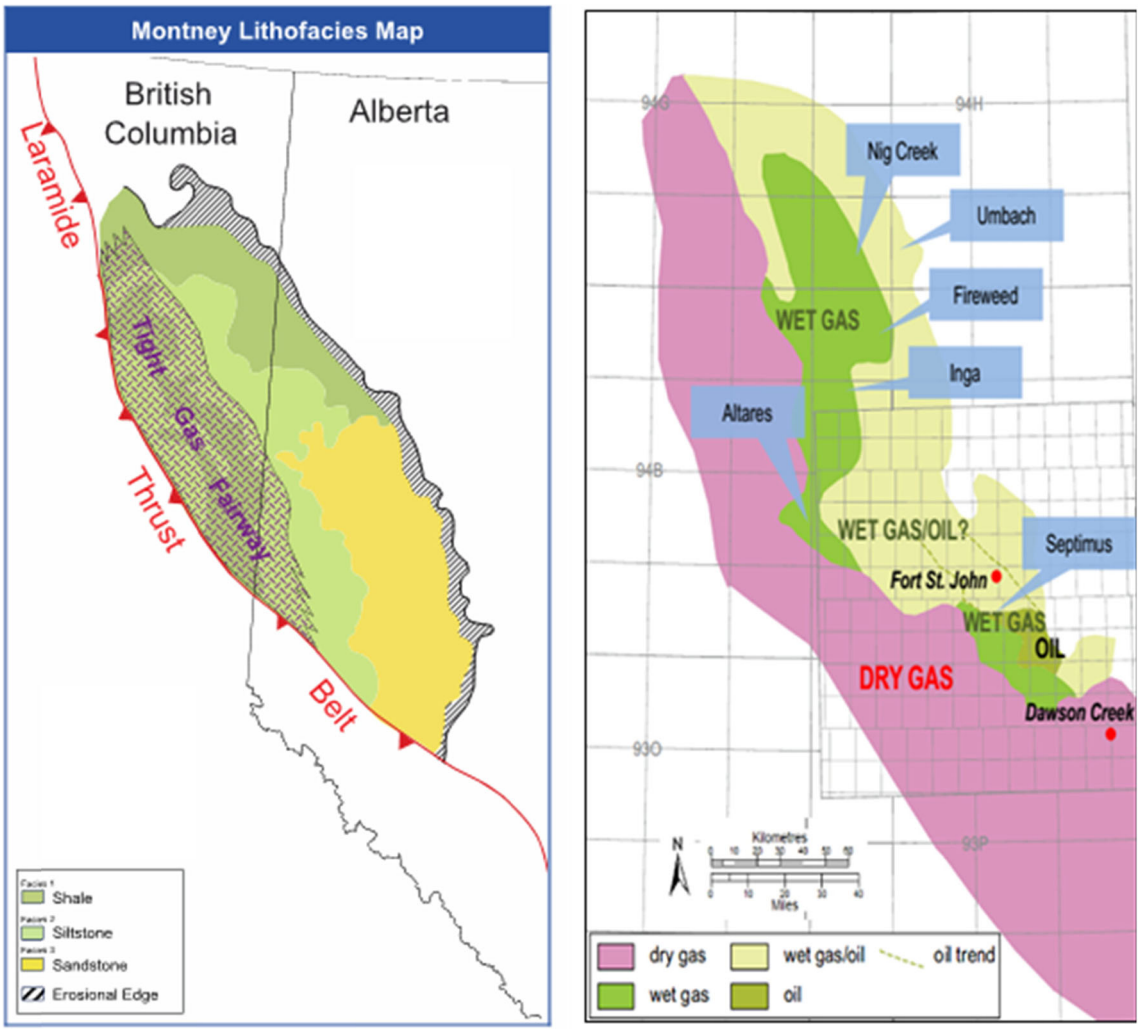

Fig. 6 *Left* The expected facies distribution map of the Montney Formation across Alberta and British Columbia's areas [113]. *Right* The expected fluid distribution within the British Columbia's region [114]

variability in expected fluid distribution from dry gas in the western flank to more liquid rich fluid in the eastern areas [114]. We consider a 1-year production data from a fractured tight gas condensate well in this Formation. The subject well is located within the British Columbia's areas, and is producing liquid-rich gas from a tight siltstone reservoir. This is a horizontal 8-stage fractured well with a length of 1520 m that is drilled at an average depth of 1984.2 m Kelly Bushing (KB) True Vertical Depth (TVD). The subject well is producing a mixture of hydrocarbon and water. The initial reservoir pressure is estimated 4850 psi from the Diagnostic Fracture Injection Test (DFIT) [115]. The Condensate Gas Ratio (CGR) varies from around 35 bbl/MMscf to a constant low value of 4 bbl/MMscf while the well pressure is almost stabilized at around 430 psi after 1 year of production.

Initial PVT analysis considered an 11-component fluid with a composition that was approximately known from one surface sampling after around 100 days of production. However, there was a high concern that if the recombination of the sampled fluid would be a good representation of the initial fluid. This is of course a valid concern as for the tight and shale reservoirs, large pressure drawdowns happen around the wellbore, which can dramatically change the composition of the produced fluid [116, 117]. This issue in turn questions the application of analytical methods [118, 119] for the use of standard production data analysis to achieve a confident well/reservoir characterization. Therefore, the main objective of this well history-matching problem was to characterize the reservoir and fluid models from numerical (compositional) simulations.

The fluid behavior is modeled using a Peng-Robinson Equation-of-State (PR-EOS) [120] with an unknown *initial* fluid composition. The gas condensate fluid components are N_2, CO_2, C_1, C_2, C_3, i-C_4, n-C_4, i-C_5, n-C_5, n-C_6, and C_{7+} that are similar to those determined from the separator samples. N_2 and CO_2 had very small contributions and were kept constant, while the composition of the other eight components were considered unknown. The C_{7+} properties are estimated from specific gravity and molecular weights from some relevant correlations. Therefore, the *in situ* fluid composition (the mole fractions of the components) is used as a matching parameter vector and the misfit function is formulated to also include the mole fraction of the produced fluid at the separator.

The numerical model for this problem is an effective 1D-gridded model suitable for simulating the linear flow, with an assumption of uniform symmetrical hydraulic fractures. This 1D model has one well connection and tends to simulate the flow towards one face of a fracture instead of simulating a horizontal well with eight symmetrical fractures. The compositional simulations are performed using tNavigator [121]. tNavigator is a powerful reservoir simulator designed to run parallel acceleration algorithms on multicore and manycore shared and distributed memory computing systems. For this 1D model, logarithmic gridding with a geometric ratio of 1.28 is used for a grid size distribution away from the fracture face. A large lateral distance is used to ensure that the well is in transient flow for the entire producing period. We assumed the fracture behaves as an infinite conductivity fracture. The fracture is modeled using a single grid with a width of 1 ft and large permeability (comparing to the matrix permeability) to give a large conductivity of 50 md ft (see Fig. 7). It was previously shown by Hamdi et al. [39] that for the duration of the production data in this well, the fractures interference did not start as the Palacio-Blasingame type curve [122] showed a dominant linear flow (a negative one half slope trend) for duration of the test. By assuming symmetric and similar hydraulic fractures as shown in Fig. 7, the production flow rates of the well were divided by 16, which amounts to flow toward one side of a fracture plane. In this model, the reservoir thickness (h) is 328 ft, (assumed) fracture half-length (x_f) is 158.5 ft, porosity (φ) is 5 %, lateral length (l) is 5000 ft for simulating a pure transient flow, and the assumed propped fracture conductivity ($k \times w$) is 50 md ft. An enhanced permeability region with a variable size is considered around the wellbore to simulate the impact of the Stimulated Reservoir Volume (SRV). The lateral extent of the SRV is determined from a simplified stepwise water saturation profile and the remaining water in the SRV after injection (Q_{wr} = volume of injected water − volume of produced water = 2245 bbl or 12,606 ft^3), that is:

$$W_{\text{SRV}} = \frac{Q_{w_r}}{2x_f h\varphi\,(S_{\text{w\_SRV}} - S_{\text{w\_init}})} \tag{17}$$

where W_{SRV} is the width of the SRV (ft), Q_{wr} is the remaining water in the SRV (ft^3), $S_{\text{w\_SRV}}$ is the average water saturation within the SRV, and $S_{\text{w\_init}}$ is the lower average water saturation in the matrix reservoir volume. The latter two saturation values construct an initial water profile in the

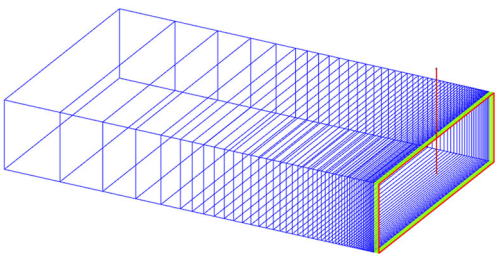

Fig. 7 The equivalent 1D fracture model that is used for the simulation of the linear flow towards the fracture face. The *warmer color* is the larger permeability and indicates the permeability contrast in the fracture, SRV and the matrix for one realization

reservoir model. Introducing this water profile is necessary to simulate an initially high volume of water that was being produced at the beginning of the production.

Other uncertain model parameters includ, the permeabilities and initial saturations in the SRV and in the bulk of the reservoir. The geomechanical effects are only considered in the main hydraulic fracture with a simplified model to account for the permeability variation with pressure [123], that is:

$$k_f = k_i \exp\left(\gamma\left(p - p_i\right)\right) \tag{18}$$

in which k_f and k_i are the current and the original fracture permeabilities at p and p_i respectively. p_i is the initial reservoir pressure and γ is the permeability reduction factor. Using pressure-dependent permeability, instead of a is fully geomechanical approach, was suggested by Dinh et al. [124] to be a good approximation for hard competent linear elastic rocks that are found in this field. The selected ranges of γ allow the simulation of the extreme cases when the propped fracture stays unchanged or when it is immediately closed by the early drawdowns.

Corey's equations [125] are used to generate the relative permeability curves from some fixed end-points for relative permeabilities and saturations, by only varying some exponents (N_w, N_{ow}, N_g, N_{og}) to shape the water-oil and oil-gas relative permeability curves [125]. In this work, the irreducible water saturation is assumed to be 0.15. Other phenomena, including sorption and non-Darcy flow are not considered in this study. Table 1 presents the model parameters with their prior variation ranges for this problem.

6 Results

This well is operated by gas production for the first 50 days and is then switched to the bottomhole pressure production control afterwards. For the single-objective history-matching case, a total (or scalarized) misfit function is defined to combine the misfits of the daily production rates, as well as pressure and surface hydrocarbon compositions (δ). Hence, this is the "total misfit" that is modeled by the GP in the single-objective optimization trials. However, because the individual misfits have different magnitudes, they are normalized by their average values of corresponding measured data over the entire production period (shown with a bar sign "‾") so that the contribution of each misfit is comparable to the other misfits. The average values are constant and act as the normalization constants [126]. Although, in general, it is recommended to use the formal misfit formulation to include the measurement errors, in this case, we did not have any issue to simply combine our

Table 1 Parameters adjusted for history-matching [39]

		Index	Parameter description	Min	Max
Initialization	Water	$X1$	Movable initial water in virgin matrix	0.2	0.4
		$X2$	Movable initial water in the SRV	0.4	0.7
	Rock	$X3$	Bulk permeability	0.00001 md	0.001 md
		$X4$	SRV permeability	0.01 md	10 md
	Fluid	$X11$	C1	0.75	0.9
		$X12$	C2	0.05	0.08
		$X13$	C3	0.04	0.08
		$X14$	i-C4	0.005	0.02
		$X15$	n-C4	0.005	0.02
		$X16$	i-C5	0.002	0.01
		$X17$	n-C5	0.002	0.01
		$X18$	C6	0.002	0.01
		$X19$	C7+	0.015	0.06
PVT(C7+ characterization)		$X5$	MW of C7+	100	140
		$X6$	SG of C7+	0.75	0.9
Relative permeability (Corey's)		$X7$	N_w	2	4
		$X8$	N_{ow}	2	4
		$X9$	N_g	2	4
		$X10$	N_{og}	2	4
Fracture permeability reduction		$X20$	γ	1E-6	0.01

objectives using the average values. The sum of the squared normalized difference or total misfit is defined as follows:

$$\begin{aligned}\text{Total misfit} = &\sum_{t=1}^{T=368}\left(\frac{q_{o_{\text{obs}}}(\text{t})-q_{o_{\text{sim}}}(t)}{\bar{q}_{o\_obs}}\right)^2\\ &+\sum_{t=1}^{T=368}\left(\frac{q_{g_{\text{obs}}}(\text{t})-q_{g_{\text{sim}}}(t)}{\bar{q}_{g\_obs}}\right)^2\\ &+\sum_{t=1}^{T=368}\left(\frac{q_{w_{\text{obs}}}(\text{t})-q_{w_{\text{sim}}}(t)}{\bar{q}_{w\_obs}}\right)^2\\ &+\sum_{t=1}^{T=368}\left(\frac{p_{\text{obs}}(t)-p_{\text{sim}}(t)}{\bar{p}_{obs}}\right)^2\\ &+10\times\delta_{\text{misfit}|@\,t=100}\end{aligned} \tag{19}$$

where $\delta_{\text{misfit}|@t\,=\,100}$ can be expanded as:

$$\begin{aligned}\delta_{\text{misfit}|@\,t=100} = &\sum_{j=1}^{9}\left(\frac{\delta_{j,\text{obs}}-\delta_{j,\,\text{sim}}}{\delta_{j,\text{obs}}}\right)_o^2\\ &+\sum_{j=1}^{9}\left(\frac{\delta_{j,\text{obs}}-\delta_{j,\,\text{sim}}}{\delta_{j,\text{obs}}}\right)_g^2\\ =&\left(\frac{\delta_{1,\text{obs}}-\delta_{1,\,\text{sim}}}{\delta_{1,\text{obs}}}\right)_o^2+\left(\frac{\delta_{2,\text{obs}}-\delta_{2,\,\text{sim}}}{\delta_{2,\text{obs}}}\right)_o^2\\ &+\ldots+\left(\frac{\delta_{1,\text{obs}}-\delta_{1,\,\text{sim}}}{\delta_{1,\text{obs}}}\right)_g^2\\ &+\left(\frac{\delta_{2,\text{obs}}-\delta_{2,\,\text{sim}}}{\delta_{2,\text{obs}}}\right)_g^2+\ldots\end{aligned} \tag{20}$$

in which $t = 1, 2, 3, \ldots, 368$ are the simulation time steps (in days), and $j = 1, 2, 3, ..., 9$ iterates over the hydrocarbon fluid components. $\delta_1, \delta_2, \delta_3 \ldots, \delta_9$ represent the mole fractions of C_1, C_2, C_3, …, C_{7+}. In above equations, p is the well pressure, q is the surface production rate, the subscripts obs and sim represent the observed and the simulation data, and the subscripts o, w, and g indicate oil, water, and gas phases. As such, $(\delta_{1,\text{obs}})_o$ represent the measured (observed) mole fraction of C_1 in the oil phase at the separator. $\delta_{\text{misfit}|@t=100}$ is the summation of composition misfits for all components in the separator fluid (oil and gas) that is calculated from the measured separator sample and the simulation output after 100 days of production.

As it was noted in the manuscript, the *in situ* fluid composition is also an unknown vector to be estimated by the GP-based optimization algorithm (Table 1, parameters X11 to X19). The initial ranges of the *in situ* fluid mole fractions for each component in the history-matching are determined by considering various published data for gas condensate fluids and by consulting the operating company. Therefore, at each iteration of the history-matching process an *in situ* fluid vector is generated that is used to perform the compositional simulation. To calculate $\delta_{\text{misfitt}|@t=100}$, we run the simulation using the generated composition (and other parameters) at each iteration, and then after performing the simulations we extract the mole fractions of the *produced* hydrocarbon fluid (oil and gas) at the separator condition at time step 100 (days). The extracted mole fractions from the simulations are compared with the available separator sample data, which are subsequently used to compute the overall composition misfits from Eq. 20. Note that we only have one reliable measured separator sample that was taken after 100 days of production. The overall composition misfit ($\delta_{\text{misfitt}|@t=100}$) is multiplied by 10 to be in the same range as the other misfit components. This number has been determined by comparing the individual misfits calculated from many sensitivity runs before starting the history-matching process.

6.1 Convergence

The performance of GP-based optimization method is measured using the current best samples that have the lowest misfit at each iteration (i.e., a regret value). An initial population with 10 random members is uniformly generated from the parameter space and the reservoir simulations are performed to calculate the total misfits for this initial population. We set a budget of 200 affordable simulations for any single optimization run and we repeat our procedure 20 times. For this problem, the efficiency of GP-based optimization algorithm for history-matching is reported by taking an average over the current best misfits for all GP-based optimization runs and presenting the associated standard deviations. This is to minimize the effect of randomness in the initial population that might mislead the performance outlook.

As the iterations progress, more samples become available and the predictive performance of the GP model increases. This could lead to a misfit reduction over almost two orders of magnitude in around 100 simulations. The computational efficiency of the GP-based optimization is

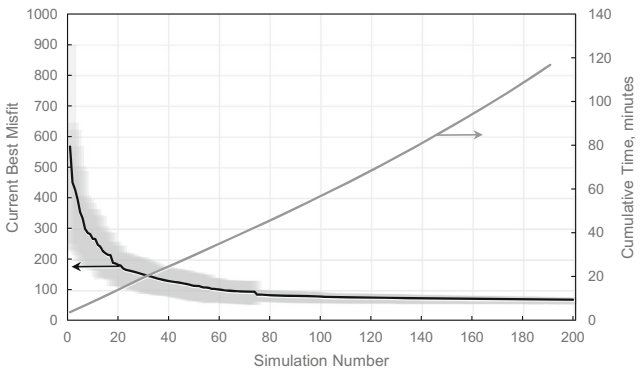

Fig. 8 The average convergence of the GP in terms of current best sample resulting from 20 different history-matching tries. In almost all trials, the convergence is attained after around 100 simulations

relatively cheap for small number of samples (<500) and moderate dimensions (<20). Figure 8 displays the average performance of the best models and the overall spent time resulting from the convergence of 20 individual optimization trials. The error bars only represent the standard deviations of the regret values at each iteration that are obtained from 20 optimization runs. The overall match quality for such a limited number of simulations in each optimization trial is relatively efficient. The corresponding history-matched curves for the rates and compositions are shown in Fig. 9. In this problem, we could not further improve the quality of the match particularly for the seperator oil compositions. Although, this could be likely due to the existence of some errors in the sample data or the use of a simplified model in our simulations, the limited data available to this study does not allow us to conduct a robust conclusion. However, for this particular exercise, the rather poor match for the lighter components would not jeopardize the expectations as other production data are satisfactorily matched.

7 Discussion

The matched parameters for the best model, out of all the GP runs with the lowest misfit, are listed in Table 2. These results fully describe a simplified 1D model that is used to approximate the well and reservoir performance. Despite its simplicity, the average response of the subject well is adequately simulated. This is in good agreement with the results of [39] obtained from the DE algorithm after 1300 simulation runs. This indicates an almost 10-time improvement in the misfit convergence for the similar problem when the GP is used. This point is indicated in Fig. 10, in which the

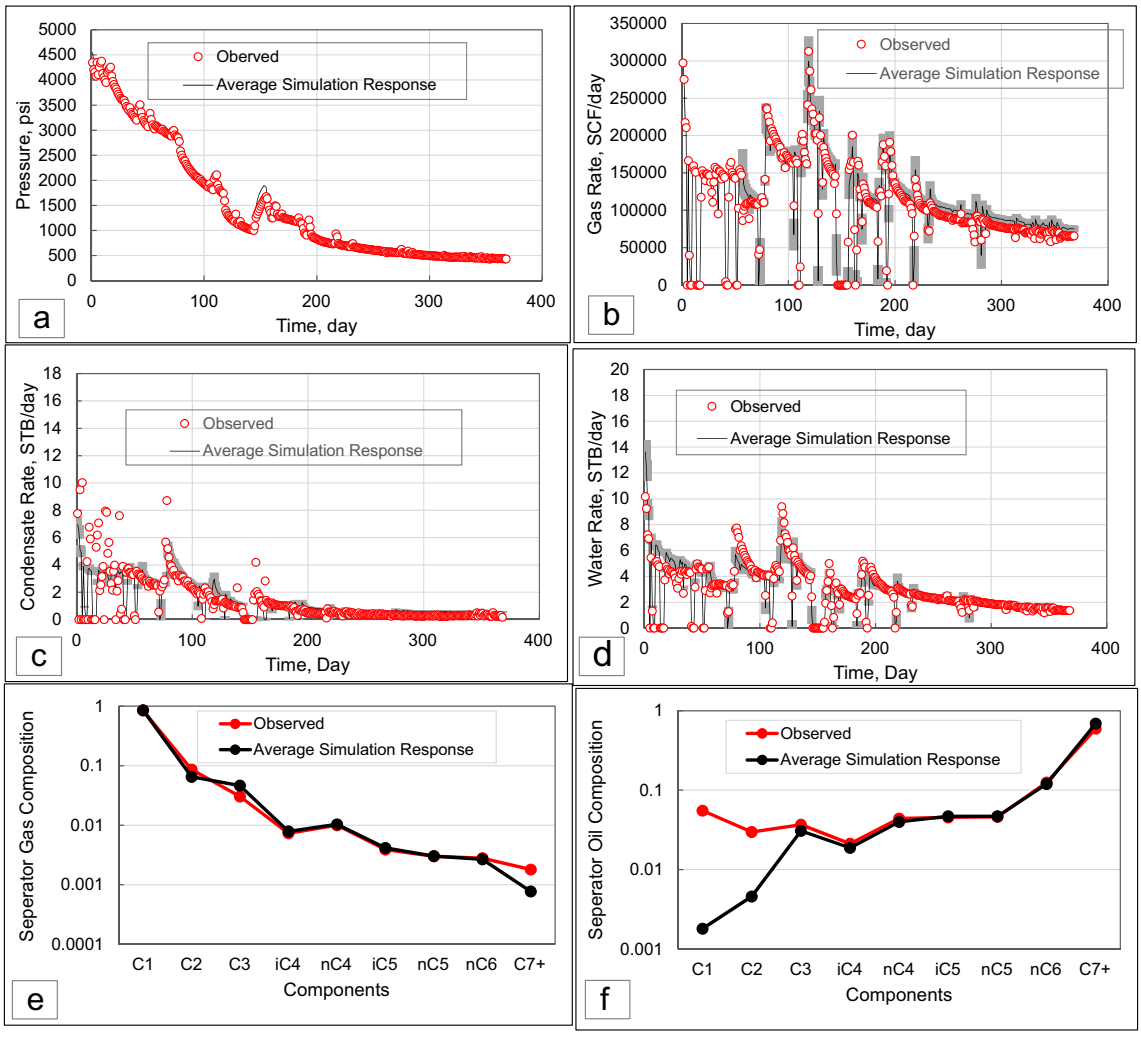

Fig. 9 The observed and the average history-matching results for pressure (**a**); oil, gas and water rates (**b**–**d**); and separator fluid composition (**e**, **f**). The *grey shaded area* around the average response (*black curves*) represents the standard deviation of the best model responses from a course of 20 different single-objective GP runs

Table 2 The best case solution vector obtained from 20 different GP runs

$X1$	0.27	$X11$	0.84
$X2$	0.45	$X12$	0.072
$X3$	0.000265	$X13$	0.038
$X4$	0.010	$X14$	0.0071
$X5$	128.59	$X15$	0.0108
$X6$	0.76	$X16$	0.0045
$X7$	2	$X17$	0.0038
$X8$	3.43	$X18$	0.0046
$X9$	3.98	$X19$	0.0142
$X10$	3.86	$X20$	0.00002
		Total misfit	50

current best misfits for the two DE and GP runs are displayed. For reservoir history-matching problems with larger models, this result is particularly appealing because running thousands of simulations is not practically feasible for many industry practitioners.

According to this solution, the fracture propagation has created an improved permeability in the SRV, which has 37 times better conductivity than the matrix. The initial movable water saturation is 0.27 in bulk and 0.45 in the SRV. Parameters X5 to X19 construct a PR-EOS model for a lean gas condensate fluid with 1.4 % of C_{7+} that has a maximum liquid drop out of 1.16 % and a dew point pressure of 2437 psi at the reservoir temperature of 162 °F from the Constant Composition Experiment (CCE) tests. Finally, X20 imposes a function with a negligible reduction of propped permeability for the entire production period, which is in line with the other observations in the nearby areas of the subject well in this field [124].

We should also note that the GP models are used to approximate the misfit functions and are not suited to directly predict the future production trends. However, it should be possible to propagate the cross validation score (or, e.g., the posterior variance at the calibrated point) of the GP model to the calibrated simulation model to add confidence bounds. In addition, for some problems, the final GP model might be used as a surrogate to perform a Markov chain Monte Carlo sampling to quantify the approximate importance of the model parameters [7, 57]. In this context, the posterior distributions of the physical model parameters can be estimated to perform forward modeling and predictions. This process, can be strengthened if the generated samples from various repetitions can be collected for a better exploration of the space in the case we have multiple solutions. For these situations, we have to re-build a one-time GP model using all samples. GP models, and other proxy models, can also be applied to time series data and provide predictive capabilities [127] to determine the future trends (forecasting) [128]. This could be an additional step after the history-matching problem was solved in this problem. However this would require generating more data to train the GP time series model.

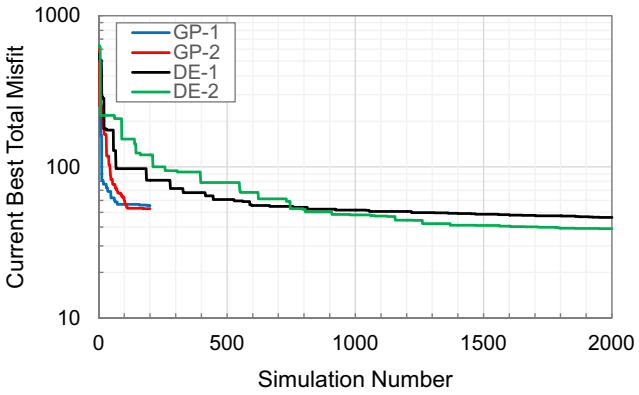

Fig. 10 The convergence of the two GP runs in comparison with the DE's ones. An almost 10-times quicker convergence is obtained using the GP for the problem studied in this paper

7.1 The impact of correlation function

The correlation function is the heart of GP modeling as it determines the smoothness of the samples which are drawn from it. In geostatistical problems, we can have a rather clear sense of the variogram type from the measured well data with fitting a suitable model to the experimental variogram. However, in the GP-based optimization methods, we usually do not know *a prior* the information about the types of the correlation functions. Nevertheless, the stationary Gaussian and Matérn functions are amongst the popular choices for optimization problems [129]. While the Gaussian correlation function works really well for many benchmark functions, it imposes unrealistically smoothness assumptions (it assumes the underlying response surface is infinite differentiable) [130]. The Matérn class of correlation functions is often advised over the more (historically) popular Gaussian correlation function for many real problems [131, 132]. Interestingly, the Gaussian kernel is a special case of the Matérn correlation function when $\nu \rightarrow$ infinity [133]. There is no straightforward procedure to choose the optimal smoothness parameter ν and this issue is considered as an open question [53]. Van der Vaart and van Zanten [134] presented an approach where the GP models could automatically adjust to the smoothness parameter. However, in their approach, the estimated correlation lengths have a tendency to zero, leading to practical and theoretical challenges [53]. Nevertheless, the Matérn correlation functions with $\nu = 3/2$ and $\nu = 5/2$ are the most commonly used variants of this family which are analytically tractable

and have been successfully used in many studies. Similar to many other studies (e.g. [135, 136]), in our experience for many optimization and history-matching problems, the Matérn class with $\nu = 3/2$ provides good results with favorable convergence rate. In this work, some sensitivities are performed to understand the performance of GP-based optimization algorithm using different correlation functions. Figure 11 shows the impact of Matérn (with $\nu = 3/2$), Exponential, Gaussian, and spherical functions on the regret values for a similar problem as studied in this paper. A similar initial population with 20 members from a space-filling Latin-Hypercube design is used in all optimization runs. Figure 11 indicates that the Matérn and Gaussian correlation functions can provide a better performance with lower misfits. A consistent reduction of the regret value is observed for the Matérn function, while the spherical correlation does not show a substantial systematic improvement within the affordable evaluation budget of 250 simulations. Although, the spherical covariance can be indeed considered as an intermediate class in smoothness between the exponential and Gaussian, it carries some characteristics that cannot be readily interpreted in terms of the convergence rate. In particular, the spherical covariance function has a finite range (compact support) meaning the covariance is reduced to zero if the distance between two points is greater than some threshold, which depends on its hyperparameters. Covariance functions with compact support often lead to ill-defined covariance matrices for certain hyperparameters combination which can result in a more difficult hyperparameter optimization process. More importantly, as in our work, the spherical covariance function does not guarantee positive-definiteness anymore above three dimensions [137, 138]. Consequently, the optimization may get stuck in local optima which are suboptimal with regards to the other covariance functions.

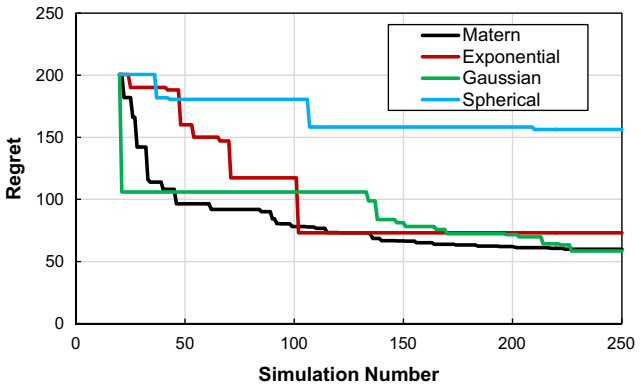

Fig. 11 The impact of various correlation functions on the performance of a GP-based optimization method by reducing the regret. Similar initial samples have been used for all cases

7.2 The impact of initial design

Finding the optimal size of the initial training set is an open problem; one rule-of-thumb is to choose it as $10 \times d + 1$ where d is the dimension. However, this depends on the complexity and the dimension of the problem. The distribution of points is important for the GP models, and is preferred to be relatively uniform and space filling. In literature, often, an optimized maximin Latin Hypercube is used [139, 140]. However, generating such a distribution of points becomes infeasible in higher dimensions such as in this paper. One should note that if the number of initial points is too low, the GP model is prone to deceptive data [23], where data is distributed in such a way that the GP model can fit it very well (e.g., the data might resemble a Gaussian), discouraging further exploration. It is known that the posterior variance of the Kriging model is an underestimation of the true error [141]. However, this can be solved in various ways such as applying a full Bayesian analysis to the Kriging model. A popular approach is to use slice sampling over the hyperparameters [142], to effectively construct multiple Kriging models. The final posterior variance is then an average of the posterior variance of all Kriging models (similarly for the prediction).

In summary, the quality of the GP models and their abilities to effectively search for the global solutions are naturally dependent on the quality and the location of the initial samples. Intuitively, if the initial samples can better cover the sampling space, a more reliable GP model can be fitted to the data and the search for the global solutions would be easier.

8 Conclusions

In this work, we applied the GP surrogate modeling approach for the history-matching purposes. We examined the efficiency of the GP-based optimization methods compared to the DE algorithm and a quasi-Newton optimization method using the IC fault model for global optimization. The results showed that a larger reduction in the misfit is achieved for less function evaluations when the GP-based method is used. This suggested the effectiveness of the GP modeling for reservoir history-matching problems when running thousands of numerical simulations is not feasible. We then demonstrated the efficiency of this approach for a real 20-dimensional problem in a real tight gas condensate well. The compositional simulations were used to simulate the hydrocarbon and water rates, as well as the fluid composition sample at the surface. The main motivation for this exercise was to quickly and efficiently characterize the *in situ* fluid and well/reservoir parameters using

numerical methods and to handle the nonlinearities that are the stumbling blocks to the application of analytical models. Previous work used the power of population-based methods for such tasks, which required a large number of unnecessary simulation runs. The results showed that by using a GP, one could significantly reduce the number of simulations by around 10 times. Moreover, the effect of initial designs was discussed and some sensitivities were performed to understand the effect of the correlation functions on the performance of the GP-based optimization method. The results showed that a better performance is expected if the Matérn correlation function is used in conjunction with a space-filling initial design.

Acknowledgments Hamidreza Hamdi would like to thank Mitacs and Rock Flow Dynamics for supporting his postdoctoral fellowship.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Funding Ivo Couckuyt is a postdoctoral research fellow of FWO-Vlaanderen. He has received partial funding from Interuniversity Attraction Poles Program BESTCOM initiated by the Belgian Science Policy Office.

References

1. Evensen, G.: Data assimilation-the ensemble Kalman filter, 2nd edn, vol. XXIV, p. 308. Springer (2009)
2. Skjervheim, J.-A., Evensen, G.: An Ensemble Smoother for Assisted History Matching. SPE-141929-MS, Paper Presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA (2011)
3. Leeuwen, P.J.v., Evensen, G.: Data assimilation and inverse methods in terms of a probabilistic formulation. Mon. Weather Rev. **124**(12), 2898–2913 (1996). doi:10.1175/1520-0493(1996)1242898:DAAIMI2.0.CO;2
4. Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math Geosci **44**(1), 1–26 (2012). doi:10.1007/s11004-011-9376-z
5. Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. Comput. Geosci. **55**, 3–15 (2013). doi:10.1016/j.cageo.2012.03.011
6. Chen, Y., Oliver, D.S., Zhang, D.: Efficient ensemble-based closed-loop production optimization. SPE J. **14**(4), 634–645 (2009). doi:10.2118/112873-PA
7. Christie, M., Cliffe, A., Dawid, P., Senn, S.S.: Simplicity, complexity and modelling, p. 220. Wiley
8. Oliver, D., Chen, Y.: Recent progress on reservoir history matching: a review. Comput Geosci **15**(1), 185–221 (2011). doi:10.1007/s10596-010-9194-2
9. Bard, Y.: Nonlinear parameter estimation, p. 341. Academic Press, NY (1974)
10. Gill, P.E., Murray, W., Wright, M.H.: Practical optimization, p. 401. Academic Press (1981)
11. Regis, R.G., Shoemaker, C.A.: Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization. Eng. Optim. **45**(5), 529–555 (2012). doi:10.1080/0305215X.2012.687731
12. Ciaurri, D., Mukerji, T., Durlofsky, L.: Derivative-Free Optimization for Oil Field Operations. In: Yang, X.-S., Koziel, S. (eds.) Computational Optimization and Applications in Engineering and Industry, Vol. 359. Studies in Computational Intelligence, pp. 19–55. Springer, Berlin Heidelberg (2011)
13. Wang, Y., Shoemaker, C.A.: A general stochastic algorithmic framework for minimizing expensive black box objective functions based on surrogate models and sensitivity analysis. arXiv:2014arXiv1410.6271W (2014)
14. Hong, H., Mahajan, A., Nekipelov, D.: Extremum estimation and numerical derivatives. J. Econ. **188**(1), 250–263 (2015). doi:10.1016/j.jeconom.2014.05.019
15. Gilman, J.R., Ozgen, C.: Reservoir simulation: history matching and forecasting. Society of Petroleum Engineers, Richardson, TX (2013)
16. Landa, J.L.: Integration of Well Testing into Reservoir Characterization. In: Kamal (ed.) TransientWell Testing-Mongraph Series, vol. 23. Society of petroleum Engineers, USA (2009)
17. Price, K., Storn, R.M., Lampinen, J.: Differential evolution: a practical approach to global optimization, 538 (2005)
18. Forrester, A., Sobester, A., Keane, A.: Engineering design via surrogate modelling: a practical guide. Wiley (2008)
19. Ciaurri, D.E., Isebor, O.J., Durlofsky, L.J.: Application of derivative-free methodologies to generally constrained oil production optimization problems. Procedia Computer Science **1**(1), 1301–1310 (2010). doi:10.1016/j.procs.2010.04.145
20. Rasmussen, C.E., Williams, C.K.I.: Gaussian processes for machine learning (adaptive computation and machine learning). The MIT Press (2005)
21. Couckuyt, I., Dhaene, T., Demeester, P.: ooDACE toolbox: a flexible object-oriented Kriging implementation. J. Mach. Learn. Res. **15**(1), 3183–3186 (2014)
22. Jones, D.: A taxonomy of global optimization methods based on response surfaces. J. Global Optim. **21**(4), 345–383 (2001). doi:10.1023/A:1012771025575
23. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. J. Glob. Optim. **13**(4), 455–492 (1998). doi:10.1023/a:1008306431147
24. Rasmussen, C.E.: Gaussian Processes in Machine Learning. In: Bousquet, O., von Luxburg, U., Rätsch, G. (eds.) Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4–16, 2003, Revised Lectures, pp. 63–71. Springer, Berlin, Heidelberg (2004)
25. Chiles, J.P., Delfiner, P.: Geostatistics: modeling spatial uncertainty, Volume 713 of Wiley Series in Probability and Statistics, 576 (2012)
26. Azimi, J., Fern, A., Fern, X.: Batch Bayesian optimization via simulation matching Paper presented at the NIPS (2010)
27. Bo, L., Qingfu, Z., Gielen, G.G.E.: A Gaussian process surrogate model assisted evolutionary algorithm for medium scale expensive optimization problems. IEEE Transactions on Evolutionary Computation **18**(2), 180–192 (2014). doi:10.1109/TEVC.2013.2248012
28. Lu, J., Li, B., Jin, Y.: An evolution strategy assisted by an ensemble of local Gaussian process models, Paper presented at the Proceedings of the 15th annual conference on Genetic and evolutionary computation, Amsterdam The Netherlands (2013)
29. Lizotte, D.J., Wang, T., Bowling, M., Schuurmans, D.: Automatic gait optimization with Gaussian process regression, Paper presented at the IJCAI (2007)

30. Chan, L.L.T., Liu, Y., Chen, J.: Nonlinear system identification with selective recursive gaussian process models. Ind. Eng. Chem. Res. **52**(51), 18276–18286 (2013). doi:10.1021/ie4031538
31. Gorissen, D., Couckuyt, I., Demeester, P., Dhaene, T., Crombecq, K.: A surrogate modeling and adaptive sampling toolbox for computer based design. J. Mach. Learn. Res. **11**, 2051–2055 (2010)
32. Hamdi, H., Hajizadeh, Y., Azimi, J., Sousa, M.C.: Sequential Bayesian optimization coupled with differential evolution for geological well testing, Paper presented at the 76th EAGE Conference and Exhibition 2014 Amsterdam the Netherlands (2014)
33. Tavassoli, Z., Carter, J.N., King, P.R.: An analysis of history matching errors. Comput. Geosci. **9**(2), 99–123 (2005). doi:10.1007/s10596-005-9001-7
34. Elahi, S.H., Jafarpour, B.: Characterization of fracture length and conductivity from tracer test and production data with ensemble Kalman Filter, Paper presented at the Unconventional Resources Technology Conference, San Antonio, Texas USA (2015)
35. Anderson, D.M., Nobakht, M., Moghadam, S., Mattar, L.: Analysis of Production Data from Fractured Shale Gas Wells. SPE-131787-MS, Paper Presented at the SPE Unconventional Gas Conference, Pittsburgh, Pennsylvania, USA (2010)
36. Orangi, A., Nagarajan, N.R., Honarpour, M.M., Rosenzweig, J.J.: Unconventional Shale Oil and Gas-Condensate Reservoir Production, Impact of Rock, Fluid, and Hydraulic Fractures. 140536-MS, Paper Presented at the SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, USA (2011)
37. Zhang, X., Du, C.M.: Sensitivity Analysis of Hydraulically Fractured Shale Gas Reservoirs. In: Ma, Y.Z., Pointe, P.R. (eds.) Uncertainty Analysis and Reservoir Modeling: AAPG Memoir 96. The American Association of Petroleum Geologists, USA (2011)
38. Storn, R., Price, K.: Differential Evolution—A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. In: Technical Report TR-95-012. Berkeley (1995)
39. Hamdi, H., Behmanesh, H., Clarkson, C.R., Costa Sousa, M.: Using differential evolution for compositional history-matching of a tight gas condensate well in the Montney Formation in western Canada. J. Nat. Gas Sci. Eng. **26**, 1317–1331 (2015). doi:10.1016/j.jngse.2015.08.015
40. Sacks, J., Welch, W., Mitchell, T., Wynn, H.: Design and analysis of computer experiments. Stat. Sci. **4**(4), 409–423 (1989). doi:10.2307/2245858
41. Mockus, J.: Application of Bayesian approach to numerical methods of global and stochastic optimization. J. Glob. Optim. **4**(4), 347–365 (1994). doi:10.1007/bf01099263
42. Handcock, M.S., Stein, M.L.: A Bayesian analysis of kriging. Technometrics **35**(4), 403–410 (1993). doi:10.2307/1270273
43. Helbert, C., Dupuy, D., Carraro, L.: Assessment of uncertainty in computer experiments from universal to Bayesian kriging. Appl. Stoch. Model. Bus. Ind. **25**(2), 99–113 (2009). doi:10.1002/asmb.743
44. Kleijnen, J.P.C.: Design and analysis of simulation experiments. Springer Publishing Company, Incorporated p. 218 (2007)
45. Waller, L.A., Gotway, C.A.: Spatial Exposure Data. In: Applied Spatial Statistics for Public Health Data, pp. 272?324. Wiley (2004)
46. MathWorks: Matlab Optimization Toolboxuser's guide, Version 2014a (2014)
47. Kawaguchi, K., Kaelbling, L.P., Lozano-Perez, T.: Bayesian Optimization with Exponential Convergence, Paper Presented at the Advances in Neural Information Processing Systems. arXiv:1604.01348, Montreal, Canada (2015)
48. Ginsbourger, D., Le Riche, R., Carraro, L.: Kriging is Well-Suited to Parallelize Optimization. In: Tenne, Y., Goh, C. (eds.) Computational Intelligence in Expensive Optimization Problems, pp. 131–162. Springer, Berlin (2010)
49. DUBOURG, V.: Adaptive surrogate models for reliability analysis and reliability-based design optimization. PhD thesis, université Blaise Pascal, Clermont-Ferrand, France, p. 282 (2011)
50. Jeffreys, H.: An invariant form for the prior probability in estimation problems. Proc. R. Soc. Lond. A Math. Phys. Sci. **186**(1007), 453–461 (1946)
51. Zellner, A.: An introduction to Bayesian inference in econometrics. J. Wiley, p. 431 (1971)
52. Moèkus, J.: On Bayesian Methods for Seeking the Extremum. In: Marchuk, G.I. (ed.) Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974, Vol. 27. Lecture Notes in Computer Science, Pp. 400–404. Springer, Berlin Heidelberg (1975)
53. Bull, A.D.: Convergence rates of efficient global optimization algorithms. J. Mach. Learn. Res. **12**, 2879–2904 (2011). arXiv:1101.3501
54. Hu, L.Y., Blanc, G., Noetinger, B.: Gradual deformation and iterative calibration of sequential stochastic simulations. Math. Geol. **33**(4), 475–489 (2001). doi:10.1023/a:1011088913233
55. Caers, J.: Geostatistical History Matching under Training-Image Based Geological Model Constraints. 00077429, Paper Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas (2002)
56. Hoffman, B.T., Caers, J.: Regional probability perturbations for history matching. J. Pet. Sci. Eng. **46**(1–2), 53–71 (2005). doi:10.1016/j.petrol.2004.11.001
57. Hamdi, H., Hajizadeh, Y., Costa Sousa, M.: Population based sampling methods for geological well testing. Comput. Geosci. **19**(5), 1089–1107 (2015). doi:10.1007/s10596-015-9522-7
58. Hamdi, H., Sousa, M.C.: Calibrating Multi-Point Geostatistical Models Using Pressure Transient Data. SPE-180163-MS, Paper Presented at the SPE Europec Featured at 78Th EAGE Conference and Exhibition, Vienna, Austria (2016)
59. Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M.M.A., Prabhat, R.P.A.: Scalable Bayesian optimization using deep neural networks. arXiv:1502.05700 (2015)
60. Hensman, J., Fusi, N., Lawrence, N.D.: Gaussian Processes for Big Data. arXiv:1309.6835 (2013)
61. Wang, Z., Hutter, F., Zoghi, M., Matheson, D., de Feitas, N.: Bayesian optimization in a billion dimensions via random embeddings. J. Artif. Intell. Res. **55**, 361–387 (2013)
62. Carter, J.N., White, D.A.: History matching on the imperial college fault model using parallel tempering. Comput. Geosci. **17**(1), 43–65 (2013). doi:10.1007/s10596-012-9313-3
63. Li, R.: Conditioning geostatistical models to three-dimensional three-phase flow production data by automatic history matching. Ph.D thesis (2001)
64. Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse theory for petroleum reservoir characterization and history matching. Cambridge University Press, Cambridge (2008)
65. Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation. EEE Control Systems Magazine, pp. 83–104 (2009)
66. Jansen, J.D.: Adjoint-based optimization of multi-phase flow through porous media. Comput. Fluids **46**, 40–51 (2010)
67. Emerick, A.A., Reynolds, A.C.: History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. Comput. Geosci. **16**(3), 639–659 (2012)
68. Li, R., Reynolds, A.C., Oliver, D.S.: History matching of three-phase flow production data. SPEJ **8**(04), 328–340 (2003). doi:10.2118/87336-PA

69. Zhang, F., Reynolds, A.C., Oliver, D.S.: The impact of upscaling errors on conditioning a stochastic channel to pressure data. SPE J. **8**(01), 13–21 (2003). doi:10.2118/83679-PA
70. Huguet, F., Lange, A., Egermann, P., Schaaf, T.: Automated History-Matching of Radial Models Using Self Organizing Classification Method for Underground Gas Storage Reservoirs Characterization. SPE-180177-MS, Paper Presented at the SPE Europec Featured at 78Th EAGE Conference and Exhibition, Vienna, Austria (2016)
71. Shahverdi, H., Sohrabi, M.: Relative permeability characterization for water-alternating-gas injection in oil reservoirs. SPE J. **21**(03), 799–808 (2016). doi:10.2118/166650-PA
72. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. Trans. Evol. Comp. **1**(1), 67–82 (1997). doi:10.1109/4235.585893
73. Caers, J.: The probability perturbation method—an alternative to a traditional bayesian approach for solving inverse problems, Paper presented at the ECMOR IX - 9th European Conference on the Mathematics of Oil Recovery Cannes France (2004)
74. Mohamed, L., Christie, M.A., Demyanov, V., Robert, E., Kachuma, D.: Application of Particle Swarms for History Matching in the Brugge Reservoir. SPE-135264-MS, Paper Presented at the SPE Annual Technical Conference and Exhibition, Florence, Italy (2010)
75. Honarkhah, M., Caers, J.: Stochastic simulation of patterns using distance-based pattern modeling. Math. Geosci. **42**(5), 487–517 (2010). doi:10.1007/s11004-010-9276-7
76. Sambridge, M.: Geophysical inversion with a neighbourhood algorithm-I. Searching a parameter space. Geophys. J. Int. **138**(2), 479–494 (1999). doi:10.1046/j.1365-246X.1999.00876.x
77. Eng Swee, S., Sasena, M., Volakis, J.L., Papalambros, P.Y., Wiese, R.W.: Fast parameter optimization of large-scale electromagnetic objects using DIRECT with Kriging metamodeling. IEEE Trans. Microwave Theory Tech. **52**(1), 276–285 (2004). doi:10.1109/TMTT.2003.820891
78. Couckuyt, I., Declercq, F., Dhaene, T., Rogier, H., Knockaert, L.: Surrogate-based infill optimization applied to electromagnetic problems. Int. J. RF Microwave Comput. Aided Eng. **20**(5), 492–501 (2010). doi:10.1002/mmce.20455
79. Luenberger, D.G.: Linear and nonlinear programming (2nd edition). Addison-wesley, p. 491 (2003)
80. Nocedal, J., Wright, S.: Numerical optimization, 664 (2006)
81. Al-Baali, M.: On the behaviour of a combined extra-updating/self-scaling BFGS method. J. Comput. Appl. Math. **134**(1–2), 269–281 (2001). doi:10.1016/S0377-0427(00)00554-9
82. Cheng, W.Y., Li, D.H.: Spectral scaling BFGS Method. J. Optim. Theory Appl. **146**(2), 305–319 (2010). doi:10.1007/s10957-010-9652-y
83. Oren, S.S., Luenberger, D.G.: Self-scaling variable metric (SSVM) algorithms. Manag. Sci. **20**(5), 845–862 (1974). doi:10.1287/mnsc.20.5.845
84. Oren, S.S.: Self-scaling variable metric (SSVM) algorithms. Manag. Sci. **20**(5), 863–874 (1974). doi:10.1287/mnsc.20.5.863
85. Al-Baali, M.: Numerical experience with a class of self-scaling quasi-Newton algorithms. J. Optim. Theory Appl. **96**(3), 533–553 (1998). doi:10.1023/a:1022608410710
86. Nocedal, J., Yuan, Y.-X.: Analysis of a self-scaling quasi-Newton method. Math. Program. **61**(1), 19–37 (1993). doi:10.1007/bf01582136
87. Oren, S.S., Spedicato, E.: Optimal conditioning of self-scaling variable metric algorithms. Math. Program. **10**(1), 70–90 (1976). doi:10.1007/bf01580654
88. Zhang, F., Reynolds, A.C.: Optimization Algorithms for Automatic History Matching of Production Data. In: ECMOR VIII-8Th European Conf. on the Mathematics of Oil Recovery, Freiberg, Germany, 3-6 September (2002)
89. Schmidt, M.: minFunc: unconstrained differentiable multivariate optimization in Matlab. http://www.cs.ubc.ca/schmidtm/Software/minFunc.html (2005)
90. Deng, W., Yang, X., Zou, L., Wang, M., Liu, Y., Li, Y.: An improved self-adaptive differential evolution algorithm and its application. Chemom. Intell. Lab. Syst. **128**, 66–76 (2013). doi:10.1016/j.chemolab.2013.07.004
91. Mohamed, A.W.: An improved differential evolution algorithm with triangular mutation for global numerical optimization. Comput. Ind. Eng. **85**, 359–375 (2015). doi:10.1016/j.cie.2015.04.012
92. Hu, Z., Xiong, S., Su, Q., Zhang, X.: Sufficient conditions for global convergence of differential evolution algorithm. J. Appl. Math. **2013**, 14 (2013). doi:10.1155/2013/193196
93. Vesterstrom, J., Thomsen, R.: A Comparative Study of Differential Evolution, Particle Swarm Optimization, and Evolutionary Algorithms on Numerical Benchmark Problems. In: Evolutionary Computation, 2004. CEC2004. Congress On, 19-23 June 2004, Pp. 1980-1987 Vol.1982, 19-23 June 2004 (2004)
94. Das, S., Suganthan, P.N., Coello, C.A.C.: Guest editorial special issue on differential evolution. IEEE Trans. Evol. Comput. **15**(1), 1–3 (2011). doi:10.1109/TEVC.2011.2108970
95. Das, S., Suganthan, P.N.: Differential evolution: a survey of the state-of-the-art. IEEE Trans. Evol. Comput. **15**(1), 4–31 (2011). doi:10.1109/TEVC.2010.2059031
96. Storn, R., Price, K.: Minimizing the Real Functions of the ICEC'96 Contest by Differential Evolution. In: Evolutionary Computation, 1996., Proceedings of IEEE International Conference On, 20-22 May 1996, pp. 842-844 (1996)
97. Hachicha, N., Jarboui, B., Siarry, P.: A fuzzy logic control using a differential evolution algorithm aimed at modelling the financial market dynamics. Inf. Sci. **181**(1), 79–91 (2011). doi:10.1016/j.ins.2010.09.010
98. Das, S., Sil, S.: Kernel-induced fuzzy clustering of image pixels with an improved differential evolution algorithm. Inf. Sci. **180**(8), 1237–1256 (2010). doi:10.1016/j.ins.2009.11.041
99. Hajizadeh, Y.: Population-based algorithms for improved history matching and uncertainty quantification of Petroleum reservoirs. PhD thesis, Heriot-Watt University, p. 315 (2011)
100. Mirzabozorg, A., Nghiem, L., Yang, C., Chen, Z.: Differential Evolution for Assisted History Matching Process: SAGD Case Study. 165491-MS, Paper Presented at the SPE Heavy Oil Conference, Calgary, Alberta, Canada (2013)
101. Hajizadeh, Y., Christie, M.A., Demyanov, V.: Application of Differential Evolution as a New Method for Automatic History Matching. SPE-127251-MS, Paper Presented at the Kuwait International Petroleum Conference and Exhibition, Kuwait City, Kuwait (2009)
102. Gamperle, R., Müller, S.D., Koumoutsakos, P.: A Parameter Study for Differential Evolution. In: Proceedings of the 2002 Advances in Intelligent Systems, Fuzzy Systems, Evolutionary Computation, pp. 293?298. WSEAS Press, Interlaken (2002)
103. Storn, R.: On the Usage of Differential Evolution for Function Optimization. In: Fuzzy Information Processing Society, 1996. NAFIPS., 1996 Biennial Conference of the North American, 19-22 Jun 1996, Pp. 519-523, 19-22 Jun 1996 (1996)
104. Rönkkönen, J., Kukkonen, S., Price, K.V.: Real-Parameter Optimization with Differential Evolution. In: 2005 IEEE Congress on Evolutionary Computation, IEEE CEC 2005. Proceedings, pp. 506-513 (2005)
105. Hu, Z., Su, Q., Yang, X., Xiong, Z.: Not guaranteeing convergence of differential evolution on a class of multimodal functions. Appl. Soft Comput. **41**, 479–487 (2016). doi:10.1016/j.asoc.2016.01.001
106. Yi, W., Zhou, Y., Gao, L., Li, X., Mou, J.: An improved adaptive differential evolution algorithm for continuous optimization. Expert Syst. Appl. **44**, 1–12 (2016). doi:10.1016/j.eswa.2015.09.031

107. Gong, W., Cai, Z., Wang, Y.: Repairing the crossover rate in adaptive differential evolution. Appl. Soft Comput. **15**, 149–168 (2014). doi:10.1016/j.asoc.2013.11.005
108. Braham, H., Jemaa, S.B., Sayrac, B., Fort, G., Moulines, E.: Low Complexity Spatial Interpolation for Cellular Coverage Analysis. In: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (Wiopt), 2014 12Th International Symposium On, 12-16 May 2014, Pp. 188-195, 12-16 May 2014 (2014)
109. Memarsadeghi, N., Raykar, V.C., Duraiswami, R., Mount, D.: Efficient Kriging via Fast Matrix-Vector Products. In: Aerospace Conference, 2008 IEEE (2008)
110. Srinivasan, B.R.D., Murtugudde, R.: Efficient Kriging for Real-Time Spatio-Temporal Interpolation. In: Proceedings of 20Th Conference on Probability and Statistics in the Atmospheric Sciences, Pp. 228–235. American Meteorological Society (2010)
111. Yang, C., Duraiswami, R., Davis, L.S.: Efficient kernel machines using the improved fast Gauss transform. Advances in neural information processing systems 17 (2005)
112. Vazquez, E., Bect, J.: Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. J. Stat. Planning Inference **140**(11), 3088–3095 (2010). doi:10.1016/j.jspi.2010.04.018
113. Cardoso, M.A.: Development and application of reduced-order modeling procedures for reservoir simulation. Stanford University (2009)
114. Adams, C.: The Status of Drilling and Production in BC'S Shale Gas Plays. In: 5Th Northeast B.C. Natural Gas Summit, Vancouver, BC, Canada, September 23–24, September 23–24 (2013)
115. Barree, R.D.: Overview of Current DFIT Analysis Methodology. http://eo2.commpartners.com/users/spe/session.php?id=10329 (2013)
116. Whitson, C.H., Sunjerga, S.: PVT in Liquid-Rich Shale Reservoirs. 155499-MS, Paper Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA (2012)
117. Peres, A.M.M., Macias-Chapa, L., Serra, K.V., Reynolds, A.C.: Well-conditioning effects on bubblepoint pressure of fluid samples from solution-gas-drive reservoirs. SPE Form. Eval., 389–398 (1990) doi:10.2118/18530-PA
118. Behmanesh, H., Hamdi, H., Clarkson, C.R.: Production data analysis of tight gas condensate reservoirs. J. Nat. Gas Sci. Eng. **22**(0), 22–34 (2015). doi:10.1016/j.jngse.2014.11.005
119. Boe, A., Skjaeveland, S.M., Whitson, C.H.: Two-phase pressure test analysis. SPE Form. Eval. **4**(4), 604–610 (1989). doi:10.2118/10224-pa
120. Robinson, D.B., Peng, D.-Y.: The characterization of the heptanes and heavier fractions for the GPA Peng-Robinson programs. Gas Processors Association, Tulsa, Okla (1978)
121. Rock Flow Dynamics: tNavigator reservoir simulator's user manual (v.4.2.4) (2016)
122. Palacio, J.C., Blasingame, T.A.: Decline-curve analysis with type curves—analysis of gas well production data. 25909-MS, Paper presented at the Low Permeability Reservoirs Symposium, Denver, Colorado USA (1993)
123. Yilmaz, O., Nur, A., Nolen-Hoeksema, R.: Pore pressure profiles in fractured and compliant rooks. in society of petroleum engineers (1991)
124. Dinh, V.P., Gouge, B.A., White, A.J.: Estimating Long Term Well Performance in the Montney Shale Gas Reservoir. 169843-MS, Paper Presented at the SPE Hydrocarbon Economics and Evaluation Symposium, Houston, Texas, USA (2014)
125. Corey, A.T.: The interrelation between gas and oil relative permeabilities. Producers Monthly **19**, 38–41 (1954)
126. Juell, A.O., Whitson, C.H.: Optimized Well Modeling of Liquid-Rich Shale Reservoirs. 166380-MS, Paper Presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, USA (2013)
127. Frigola-Alcalde, R.: Bayesian time series learning with Gaussian processes. PhD thesis, University of Cambridge, p. 97 (2015)
128. Brahim-Belhouari, S., Bermak, A.: Gaussian process for nonstationary time series prediction. Comput. Stat. Data Anal. **47**(4), 705–712 (2004). doi:10.1016/j.csda.2004.02.006
129. Wilson, A.G., Adams, R.P.: Gaussian process kernels for pattern discovery and extrapolation. arXiv:1302.4245 (2013)
130. Mohammad, Y., Nishida, T.: Data mining for social robotics: toward autonomously social robots. Springer International Publishing, p. 328 (2016)
131. Stein, M.: Interpolation of spatial data: some theory for kriging. Springer Series in Statistics. Springer Science & Business Media, N. Y., p. 247 (1999)
132. Snoek, J., Larochelle, H., Adams, R.: Practical Bayesian optimization of machine learning algorithms. In (2012)
133. Brochu, E., Cora, V.M., de Freitas, N.: A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. CoRR 1012.2599 (2010)
134. Vaart, A.W.V.D., Zanten, J.H.V.: Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth, arXiv:0908.3556. Ann. Stat. **37**(5B), 2655–2675 (2009)
135. Marmin, S., Chevalier, C., Ginsbourger, D.: Differentiating the Multipoint Expected Improvement for Optimal Batch Design. In: Pardalos, P., Pavone, M., Farinella, G.M., Cutello, V. (eds.) Machine Learning, Optimization, and Big Data: First International Workshop, MOD 2015, Taormina, Sicily, Italy, July 21-23, 2015, Revised Selected Papers, pp. 37–48. Springer International Publishing, Cham (2015)
136. Picheny, V., Wagner, T., Ginsbourger, D.: A benchmark of kriging-based infill criteria for noisy optimization. Struct. Multidiscip. Optim. **48**(3), 607–626 (2013). doi:10.1007/s00158-013-0919-4
137. Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M.: Handbook of spatial statistics. CRC Press, p. 619 (2010)
138. Kao, Y.: ST733 – Applied Spatial statistics – Spring 2013, http://www4.ncsu.edu/ykao/docs/lab%203/estimation%20and%20modeling%20of%20sptial%20correlation.pdf, NC State University (2013)
139. Pistone, G., Vicario, G.: Comparing and generating Latin hypercube designs in kriging models. AStA Adv. Stat. Anal. **94**(4), 353–366 (2010). doi:10.1007/s10182-010-0142-1
140. Morris, M.D., Mitchell, T.J.: Exploratory designs for computational experiments. J. Stat. Planning and Inference **43**(3), 381–402 (1995). doi:10.1016/0378-3758(94)00035-T
141. Kleijnen, J.P.C., van Beers, W., van Nieuwenhuyse, I.: Expected improvement in efficient global optimization through bootstrapped kriging. J. Glob. Optim. **54**(1), 59–73 (2012). doi:10.1007/s10898-011-9741-y
142. Murray, I., Adams, R.P.: Slice sampling covariance hyperparameters of latent Gaussian models. arXiv:1006.0868 (2010)
143. Marelli, S., Sudret, B.: UQLab: A framework for uncertainty quantification in Matlab, Vulnerability, Uncertainty, and Risk (Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014)), pp. 2554-2563 ETH-zürich (2014)
144. Lataniotis, C., Marelli, S., Sudret, B.: UQLab user manual—kriging, chair of risk, safety & uncertainty quantification, ETH Zurich report# UQLab-v0.9-105 (2015)