An adaptive Kriging surrogate method for efficient uncertainty quantification with an application to geological carbon sequestration modeling

Shaoxing Mo[1], Xiaoqing Shi[1*], Dan Lu[2], Ming Ye[3], Jichun Wu[1]

[1]Key Laboratory of Surficial Geochemistry of Ministry of Education, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

[2]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

[3]Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, Florida, USA

*Corresponding author.

E-mail: shixq@nju.edu.cn (Xiaoqing Shi)

S.M. and X.S. conceived of the presented idea. S.M. developed the theory, performed the computations, and drafted the manuscript in consultation with X.S., D.L., M.Y., and J.W.. X.S., D.L., and M.Y. revised the manuscript critically. X.S. and J.W. supervised the project. All authors discussed the results, commented on the manuscript, and gave final approval of the version to be submitted and any revised version.

Abstract

In numerical modeling of geological carbon sequestration (GCS), uncertainty quantification (UQ) is usually needed to evaluate the impact of uncertain model parameters on model predictions caused by limited measurements and incomplete knowledge of the parameters. However, UQ for GCS is computationally expensive due to the large ensemble of complex and lengthy model simulations. In this study, we propose an adaptive Kriging method to build a fast-to-evaluate surrogate of the GCS model to alleviate the heavy computational burden. The surrogate model is efficiently generated using a Taylor expansion-based adaptive experimental design algorithm that combines a distance-based exploration criterion and an exploitation criterion to adaptively search for informative training samples. In addition, we analyze the uncertainty brought by substituting the surrogate for the actual simulation model and explore its influence on UQ results. Our method is demonstrated in a synthetic GCS model and its performance is evaluated in comparison with the conventional Monte Carlo sampling. Results indicate that our method can greatly improve the computational efficiency in UQ and provide an effective and reliable UQ solution with the consideration of surrogate uncertainty.

Key words: geological carbon sequestration, surrogate modeling, adaptive experimental design, Kriging, uncertainty quantification.

1. Introduction

Geological carbon sequestration (GCS) has been proposed as one of the most effective means to rapidly and greatly reduce the greenhouse gas emissions into the atmosphere (IPCC, 2005). In GCS, CO_2 is injected under supercritical conditions into the deep subsurface formations including depleted gas or oil reservoirs, unmineable coal beds, and saline aquifers (Holloway, 1997; IPCC, 2005). Substantial efforts have been spent to understand the subsurface processes involved in CO_2 sequestration (Schnaar and Digiulio, 2009; Harvey et al., 2012; Bandilla et al., 2015; De Silva et al., 2015; Jones et al., 2015; Wang et al., 2015). Among these efforts, numerical modeling is usually used for predication of CO_2 migration and evolution, evaluation of potential storage capacity, and design of injection strategies (Audigane et al., 2007; Schnaar and Digiulio, 2009). Model predictions are always associated with uncertainty due to limited measurements and incomplete knowledge of model parameters (Tartakovsky, 2013; Zhang et al., 2013; Zhang and Sahinidis, 2013; Shi et al., 2014; Kitanidis, 2015; Lu et al., 2016). Thus, uncertainty quantification (UQ) is required for facilitating science-informed decision making. Monte Carlo (MC) methods are widely used for UQ because they have no restriction in application to nonlinear models and their convergence is barely dependent on the problems' stochastic dimensionality (Ballio and Guadagnini, 2004, Lu et. al., 2018). However, the MC approaches involve repeated sampling, and a large ensemble of model simulations is usually required to ensure approximation accuracy, which makes the MC techniques computationally unaffordable or even infeasible for large-scale multi-phase flow models of CO_2 sequestration.

One solution for overcoming the computational burden is to build a statistical surrogate model for the simulation model. In this method, an approximation that maps the input-output relationship of the actual simulation model is constructed using a surrogate modeling technique based on a limited number of model evaluations, also called training samples. The surrogate model, which is usually computationally cheap,

is then used as a full replacement of the actual model in the uncertainty analysis. A variety of surrogate modeling techniques have been developed, including Kriging (Sacks et al., 1989), which is also known as Gaussian process (Rasmussen and Williams, 2006), neural networks (Demuth et al., 2014), polynomial chaos expansion (Sudret, 2008), and radial basis function (Buhmann, 2000). The surrogate methods have been widely used in hydrology (e.g., Crevillén-García et al., 2017; Espinet et al., 2013; Espinet and Shoemaker, 2013; Ju et al., 2018; Sun et al., 2018; Tian et al., 2017; Xu et al., 2017; Zeng et al., 2015, 2018; Zhang et al., 2013). More details on surrogate modeling can refer to review papers by Asher et al. (2015) and Razavi et al. (2012).

Depending on the purpose of surrogate modeling, the surrogates can be generally categorized as local surrogate models and global surrogate models (Eason and Cremaschi, 2014; Mo et al., 2017). The local surrogate models pursue to accurately approximate the actual model in certain interesting regions solely, e.g., the regions with the smallest objective functions (Espinet et al., 2013; Espinet and Shoemaker, 2013; Jones et al., 1998; Ju et al., 2018; Zhang et al. 2016) or where the parameters have high probability (Bliznyuk et al., 2008; Ju et al., 2018; Zhang et al., 2016). Whereas, a global surrogate model is devoted to provide an accurate approximation of the actual model over the entire parameter space. In comprehensive UQ where low probability regions matter (Crevillén-García et al., 2017; Mo et al., 2017; Pau et al., 2013; Tian et al., 2017; Zhang & Sahinidis, 2013), a global surrogate model which provides a globally accurate approximation is usually used.

Two requirements are needed to build an efficient and effective global surrogate model in UQ. The first requirement is to use the least amount of training data to build the most accurate surrogates. Construction of an accurate surrogate model requires evaluating the simulation model on the parameter samples drawn from the parameter space. Each sample collection involves one model evaluation, thus an efficient experimental design strategy is desired to locate the most informative samples so as to

reduce the computational costs in the surrogate construction. In this sense, an efficient experimental design should satisfy two criteria (Liu et al., 2018; Mo et al., 2017): (1) the exploration criterion, i.e., the samples should have a good coverage of the entire parameter space in order to reveal the main profile of the response surface of the actual model; and (2) the exploitation criterion, i.e., more samples should be distributed in the difficult-to-approximate nonlinear regions to reduce the approximation errors. Satisfying the exploration and exploitation criteria is vitally important in surrogate modeling for hydrological models, because the response surface of hydrological models are generally irregular that display a complex combination of flat regions and nonlinear regions (Shahsavani and Grimvall, 2011; Pau et al., 2013; Zhan et al., 2013; Zhang and Sahinidis, 2013; Jordan et al., 2015). The conventional one-shot space-filling design strategies (such as Latin hypercube sampling (LHS), McKay et al., 1979) that generate samples in a single stage satisfy the exploration criterion solely. Such approaches could cause oversampling in smooth regions and undersampling in nonlinear areas. An adaptive experimental design scheme that can adaptively place more samples in the nonlinear regions is desired to improve the computational efficiency in building accurate surrogate models (Shahsavani and Grimvall, 2011; Eason and Cremaschi, 2014; van der Herten, 2015; Mo et al., 2017; Liu et al., 2018).

The second requirement is to analyze the uncertainty in the UQ results brought by the surrogate. Surrogate models are approximation to the actual simulation models with errors. The approximation errors result in surrogate uncertainty in the UQ analysis and sometimes due to lack of training data, the surrogate uncertainty can be large and significantly affect the effectiveness of UQ analysis. Thus, it is important to analyze the surrogate uncertainty and explore its influence on and contribution to the UQ results, so as to improve the results explainability and reliability (Bilionis and Zabaras, 2016; Crevillén-García et al., 2017; Tian et al., 2017; Zhu and Zabaras, 2018).

In this study, we propose an efficient adaptive Kriging surrogate method for UQ. In this method, we use a Taylor expansion-based adaptive design (TEAD) strategy to locate informative samples so as to improve surrogate efficiency and use Kriging method to explicitly quantify the surrogate uncertainty so as to enhance the surrogate reliability. The TEAD strategy proposed in Mo et al. (2017) adaptively generates training samples by balancing exploration and exploitation. TEAD uses a hybrid score function to search for informative training samples. The hybrid function combines a distance-based exploration criterion for space exploration and a Taylor expansion-based exploitation criterion for sample densification. The exploration criterion promotes to add new samples to fill the sample-sparse areas, while the exploitation criterion promotes to densify the nonlinear regions with more samples. The two criteria are balanced by a dynamical weighting scheme to avoid oversampling and undersampling. TEAD has been demonstrated to be an efficient experiment design method for surrogate modeling in comparison with the space-filling and some widely used adaptive design methods (Mo et al., 2017). On the other hand, Kriging methods (Sacks et al., 1989) have been widely used in subsurface flow problems to build surrogate models (e.g., Crevillén-García et al., 2017; Ju et al., 2018; Sun et al., 2018; Tian et al., 2017; Xu et al., 2017; Zhang et al. 2016). They have been shown that can not only build an accurate surrogate but also quantify the surrogate uncertainty through calculating the prediction variance at unsampled points.

Surrogate methods have been applied to solve the UQ problem in GCS modeling (e.g., Ashraf et al., 2013; Crevillén-García et al., 2017; Hou et al., 2014; Jia et al., 2016; Jordan et al., 2015; Oladyshkin et al., 2011; Pan et al., 2016; Pau et al., 2014; Sun et al., 2013; 2018; Tian et al., 2017; Zhang and Sahinidis, 2013). However, previous studies either used inefficient surrogate methods with space-filling experimental designs for training sample search or did not consider surrogate uncertainty. For example, Crevillén-García et al. (2017) and Tian et al. (2017) used

inefficient space-filling designs to generate training samples, although they used Kriging method to quantify the surrogate uncertainty. Hou et al. (2014) and Pau et al. (2014) did not consider surrogate uncertainty although they employed the adaptive design strategy for surrogate construction. In realization of the gaps in the previous studies, we, for the first time, apply the adaptive design strategy for globally accurate surrogate construction and analyze the surrogate uncertainty, so as to improve both the surrogate efficiency and reliability for UQ of GCS modeling. The proposed method is tested and demonstrated using a synthetic GCS model. The computational efficiency and approximation accuracy of our method is evaluated in comparison with the conventional MC sampling without using surrogates.

The paper is organized as follows. In Section 2, we describe the Kriging surrogate method and introduce the TEAD algorithm; besides, we analyze how the surrogate uncertainty is accounted and affects the UQ results. In Section 3, we apply the adaptive TEAD-based surrogate method to the GCS simulation and discuss the results. The major conclusions are presented in Section 4.

2. Methodology

The TEAD experimental design algorithm is independent of the surrogate techniques; here we combine TEAD with Kriging method to build the surrogate model because the Kriging method can provide an estimation of prediction variance, enabling our adaptive TEAD-based surrogate method to quantify the surrogate uncertainty. In this section, we first introduce the Kriging surrogate method and the TEAD experimental design. We then describe how the surrogate uncertainty is quantified. More details on TEAD can refer to Mo et al. (2017).

2.1 Kriging surrogate method

A Kriging model predicts the response value at an un-sampled point $\boldsymbol{x}$ as the sum of a regression model and a Gaussian process as follows (Sacks et al., 1989)

$$s(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta} + Z(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{g}(\boldsymbol{x}) = \{g_1(\boldsymbol{x}), \ldots, g_p(\boldsymbol{x})\}^{\mathrm{T}}$ is a set of regression basis functions and $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_p\}^{\mathrm{T}}$ are the corresponding regression coefficients. In general, $\boldsymbol{g}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta}$ is taken as either a constant or low-order polynomials. The random process $Z(\cdot)$ is assumed be a Gaussian process with zero mean and covariance

$$\mathrm{cov}[Z(\boldsymbol{x}), Z(\boldsymbol{x}')] = \sigma^2 R(\boldsymbol{x}, \boldsymbol{x}'), \tag{3}$$

that denotes the covariance between $Z(\boldsymbol{x})$ and $Z(\boldsymbol{x}')$, where σ^2 is the process variance and $R(\boldsymbol{x}, \boldsymbol{x}')$ is the correlation function.

Given a set of N design points $\mathcal{D}$, i.e. the input samples $\boldsymbol{S} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ and their corresponding responses $\boldsymbol{f} = \{f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)\}$ obtained from forward model executions, the Kriging predictor can be written as

$$s(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta}^* + \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}}\mathbf{R}^{-1}(\boldsymbol{f} - \mathbf{G}\boldsymbol{\beta}^*), \tag{4}$$

where $\boldsymbol{r}(\boldsymbol{\theta}) = \{R(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, R(\boldsymbol{x}, \boldsymbol{x}_N)\}^T$ denotes the correlation vector, $\boldsymbol{\beta}^*$ is the generalized least estimation of $\boldsymbol{\beta}$ and is given by

$$\boldsymbol{\beta}^* = (\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{f}, \tag{5}$$

where $\mathbf{G}$ is a $N \times p$ matrix with $G_{ij} = g_j(\boldsymbol{x}_i)$, $\mathbf{R}$ is a $N \times N$ correlation matrix with $R_{ij} = R(\boldsymbol{x}_i, \boldsymbol{x}_j)$. The prediction variance at a point $\boldsymbol{x}$ is given by

$$\sigma^2(\boldsymbol{x}) = \sigma^2\left\{1 - \boldsymbol{r}(\boldsymbol{x})^T\mathbf{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) + \left[\mathbf{G}^T\mathbf{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{x})\right]^T\left(\mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}\right)^{-1}\left[\mathbf{G}^T\mathbf{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{x})\right]\right\}, \tag{6}$$

where $\sigma^2 = \frac{1}{N}\left(\boldsymbol{f} - \mathbf{G}\boldsymbol{\beta}^*\right)^T\mathbf{R}^{-1}\left(\boldsymbol{f} - \mathbf{G}\boldsymbol{\beta}^*\right)$.

Various software packages for the Kriging model are available as introduced in Erickson et al. (2018). In this paper, the DACE toolbox developed by Lophaven et al. (2002) is used for construction of the Kriging surrogate models, and the first-order polynomial regression model and the frequently used Gaussian correlation function are chosen.

2.2 TEAD algorithm for adaptive experimental design

To construct an accurate surrogate model using the least amount of actual model runs, we employ an adaptive experimental design method TEAD to generate the training data $\mathcal{D}$ for surrogate construction. The TEAD experimental design algorithm (Mo et al., 2017) uses a score function to guide the collection of informative samples for surrogate construction. The score function is formulated as a combination of a distance measure to identify the sample-sparse regions (i.e., exploration) and a nonlinearity measure to find the nonlinear areas (i.e., exploitation). More specifically, the distance measure

$$D_{\min}(\boldsymbol{x}) = \min_{\boldsymbol{x}_{\mathrm{t}} \in \boldsymbol{S}} \left\| \boldsymbol{x} - \boldsymbol{x}_{\mathrm{t}} \right\|_2 , \tag{7}$$

quantifies the smallest Euclidean distance of a candidate point $\boldsymbol{x} \in \boldsymbol{\Gamma}$ from the sample $\boldsymbol{x}_{\mathrm{t}}$ in the existing training sample set $\boldsymbol{S}$.

The nonlinearity measure is defined based on the high-order remainders of the Taylor expansion, as represented by

$$R(\boldsymbol{x}) = \left| s(\boldsymbol{x}) - t(\boldsymbol{x}; \boldsymbol{a}) \right|, \tag{8}$$

where $s(\boldsymbol{x})$ is the Kriging surrogate prediction at $\boldsymbol{x}$ as defined in Eq. (3) and $t(\boldsymbol{x}; \boldsymbol{a})$ is the first-order Taylor expansion of $s(\boldsymbol{x})$ expanded at $\boldsymbol{a} \in \boldsymbol{S}$, which can be represented by

$$t(\boldsymbol{x}; \boldsymbol{a}) = s(\boldsymbol{a}) + \nabla s(\boldsymbol{a})^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{a}), \tag{9}$$

where $\boldsymbol{a} \in \boldsymbol{S}$ is the sample point having the smallest distance to $\boldsymbol{x}$, and $\nabla s(\boldsymbol{a})$ is the gradient of $s(\boldsymbol{x})$ at $\boldsymbol{a}$. The intuition behind using the high-order remainders of

Taylor expansion to quantify the function nonlinearity is that the approximation accuracy of the first-order Taylor expansion decreases as the nonlinearity increases (Mo et al., 2017).

Note that $R(\boldsymbol{x})$ is the Taylor remainders of $s(\boldsymbol{x})$ that is only an approximation to the exact Taylor remainders of $f(\boldsymbol{x})$, so its performance in identifying nonlinear regions will be affected by the approximation accuracy of $s(\boldsymbol{x})$ for $f(\boldsymbol{x})$. In TEAD, this issues is addressed by assigning a weight that is dependent on $D_{\min}(\boldsymbol{x})$ to $R(\boldsymbol{x})$ in the final hybrid score function. More specifically, the score function is a combination of the normalized exploration and exploitation criteria defined above together with a dynamical weight ω, i.e.,

$$J(\boldsymbol{x}) = \frac{D_{\min}(\boldsymbol{x})}{\max_{\boldsymbol{x}\in\Gamma} D_{\min}(\boldsymbol{x})} + \omega(\boldsymbol{x}) \frac{R(\boldsymbol{x})}{\max_{\boldsymbol{x}\in\Gamma} R(\boldsymbol{x})}. \tag{10}$$

The $\omega(\boldsymbol{x})$ is defined by

$$\omega(\boldsymbol{x}) = 1 - \frac{D_{\min}(\boldsymbol{x})}{L_{\max}} \in [0,\ 1], \tag{11}$$

where $L_{\max}$ denotes the maximum distance of any two points in the parameter space. The intuition behind this weighting scheme is that the accuracy of $R(\boldsymbol{x})$ decreases as the distance to the expansion point (i.e., the $D_{\min}(\boldsymbol{x})$ in Eq. (6)) increases, in this case less weight should be assigned to the nonlinearity measure. This adaptive weighting strategy turns out to be an effective solution for the aforementioned issue and makes a good compromise between exploration and exploitation (Mo et al., 2017). In the search of new points, the candidates with large exploration and exploitation score values are treated as the potential points that sit in sample-sparse and nonlinear regions, thus are more likely chosen as new samples.

Apart from the score function for informative training sample search, TEAD also provides a stopping criterion for automatic termination of the search process when desired approximation accuracy is achieved. The stopping criterion uses an error

indicator based on the root mean squared error (RMSE) evaluated on the newly selected points at one iteration to terminate the design process. The error indicator is written as

$$\mathrm{RMSE}_{\mathrm{new}} = \sqrt{\frac{1}{N_{\mathrm{new}}} \sum_{i=1}^{N_{\mathrm{new}}} \left[f(\boldsymbol{x}_{\mathrm{new}}^{(i)}) - s(\boldsymbol{x}_{\mathrm{new}}^{(i)}) \right]^2}, \tag{12}$$

where $\boldsymbol{x}_{\mathrm{new}}^{(i)}$, $i = 1, \ldots, N_{\mathrm{new}}$, are the newly selected points, f and s denote the actual and surrogate model, respectively. The $\mathrm{RMSE}_{\mathrm{new}}$ is calculated at each iteration before updating the surrogate model. As the potential new points suggested by the score function generally locate in un-sampled and/or highly nonlinear regions, where the approximation errors are generally larger than other points in the parameter space, using $\mathrm{RMSE}_{\mathrm{new}}$ as the error indicator can ensure satisfaction of the desired RMSE accuracy as shown in Mo et al. (2017).

Putting the ingredients introduced above together, the procedure of TEAD-based adaptive surrogate modeling is illustrated in Fig. 1. Briefly speaking, it includes the following five steps. First, the algorithm starts with generating a set of initial samples and evaluating the forward model at these samples in Step (1). In Step (2) the surrogate model is constructed based on the samples in the current training data set. Then in Step (3) we select new samples from a set of candidate points according to their score values defined in Eq. (9). Next in Step (4) we evaluate the newly selected samples in the forward model. If the stopping criterion is not met in Step (5), we repeat above Steps (2-5) to add new samples and update the surrogate model till the stopping criterion is met.

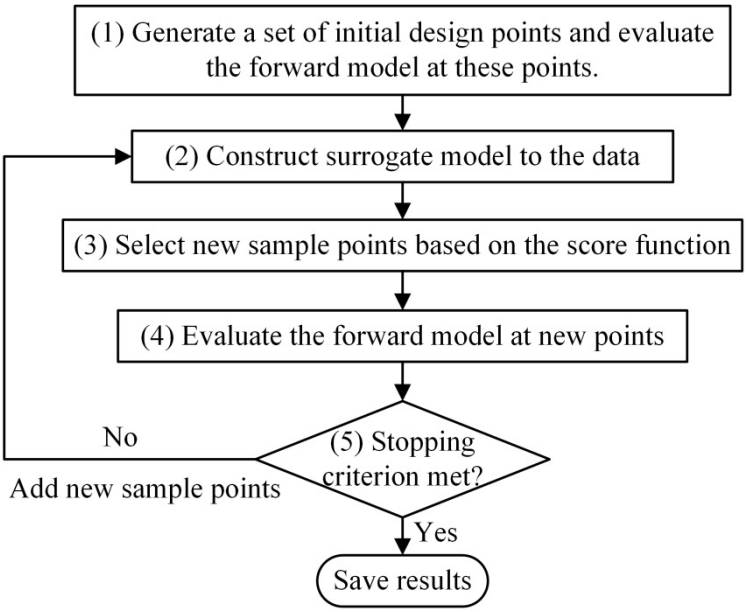

Fig. 1 Procedure of the TEAD-based adaptive global surrogate modeling.

Here we consider a simple two-dimensional analytical function:

$$f(\boldsymbol{x}) = x_1 \exp(-x_1^2 - x_2^2),\ \boldsymbol{x} \in [-2,\ 4]^2. \tag{13}$$

to illustrate TEAD's performance in collecting informative training samples. This function is nonlinear in the lower left quarter of the domain but rather smooth in the remaining regions as shown in labeled black isolines in Fig. 2. We build two surrogate models for this function: one is built using 5×5 uniformly distributed training samples; the other is built using 77 training samples generated by TEAD. The former is used to illustrate that the large approximation errors are easy to occur in highly nonlinear regions as shown Fig. 2(a); thus more samples should be placed in these regions to greatly improve the approximation accuracy. The latter is used to illustrate that TEAD is able to successfully identify and densify those poorly approximated regions. Starting with five initial samples (four samples at the corners and one in the center), Fig. 2(b) shows that TEAD successfully identifies the nonlinear regions and densifies them with more samples, greatly reducing the approximation errors.

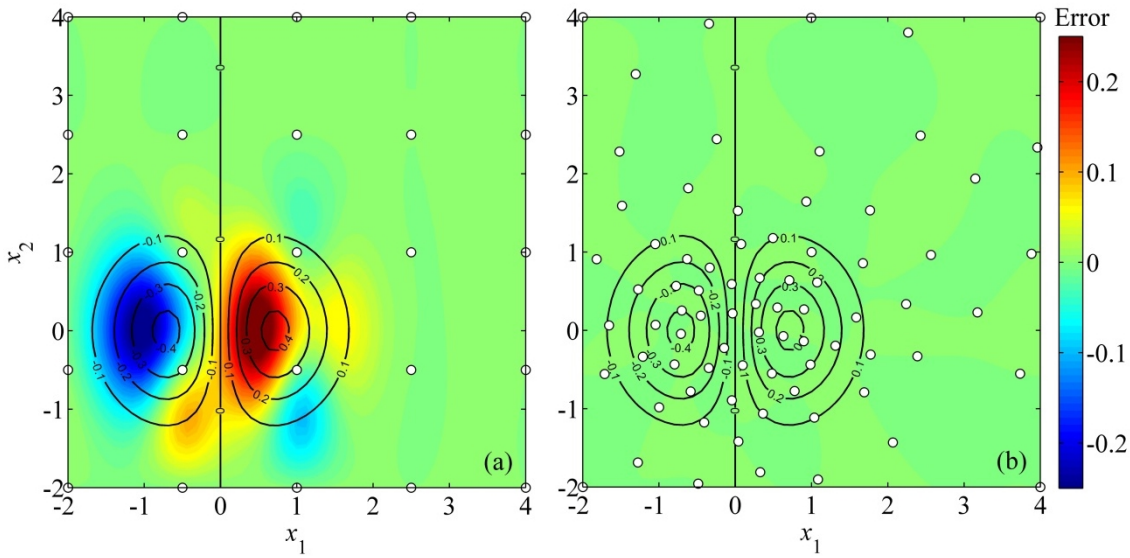

Fig. 2 Approximation errors of surrogate models for an analytical function (depicted with labeled isolines) based on (a) uniformly distributed 5×5 and (b) 77 training samples (white dots) collected by TEAD, respectively.

2.3 Quantification of uncertainty in surrogate-based UQ results

In surrogate-based UQ, the surrogate model $s(\boldsymbol{x})$ is used as a full replacement of the actual model $f(\boldsymbol{x})$. The cumulative distribution function (CDF) of quantity of interest (QoI) Q is approximated based on $s(\boldsymbol{x})$ as follows:

$$F(q) = P(Q \le q) = \int_{\mathrm{R}^d} \mathrm{I}[s(\boldsymbol{x}) \mid D \le q] dp(\boldsymbol{x}), \tag{14}$$

where $\mathcal{D} = [\boldsymbol{S}^{\mathrm{T}}, \boldsymbol{f}^{\mathrm{T}}]$ is the training sample set of the surrogate model generated by TEAD, $p(\cdot)$ is the probability density function of the random input vector $\boldsymbol{x} \in \mathbb{R}^d$, and $\mathrm{I}[\cdot]$ denotes the indicator function. The MC methods are commonly used to obtain the numerical approximation of $F(\cdot)$.

In surrogate modeling, the substitution of surrogate model for the actual model causes the surrogate uncertainty mainly caused by insufficient training data. This uncertainty will then propagate to the surrogate-based UQ results. Thus quantifying the uncertainty of $F(\cdot)$ with respect to the surrogate uncertainty is necessary to assess the reliability of the estimated CDF. To this end, we need to draw multiple realizations of $s(\boldsymbol{x})$ from its posterior prediction space being consistent with the given training data $\mathcal{D}$. We refer to the approach proposed in Oakley and O'hagan

(2002) to generate M posterior realizations of Kriging surrogates by leveraging Kriging's capability in providing an estimation of the prediction uncertainty. This involves drawing posterior sample functions $\{s_{(m)}\}_{m=1}^{M}$ from the Kriging model that are consistent with training data $\mathcal{D}$ by adding in fake training samples $\{\tilde{\boldsymbol{x}}_j, \tilde{f}(\tilde{\boldsymbol{x}}_j)\}_{j=1}^{\tilde{N}}$ which are evaluated using the Kriging surrogate instead of the actual model.

The procedure of generating one posterior realization of Kriging surrogate $s^{(m)}(\cdot)$ is summarized in Algorithm 1. One can then repeat Algorithm 1 for M times to obtain M posterior realizations of the Kriging surrogate. Briefly speaking, we first randomly generate $\tilde{N}$ input samples $\tilde{\mathcal{S}} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{\tilde{N}}\}$ from the input space. In each iteration, the input in $\tilde{\mathcal{S}}$ maximizing the prediction variance, i.e., $\tilde{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \tilde{\mathcal{S}}} \sigma^2_{(m)}(\boldsymbol{x})$, is selected as a new fake sample. The fake response value is the surrogate prediction at $\tilde{\boldsymbol{x}}$ perturbed with a Gaussian white noise having a variance that equals to the Kriging prediction variance $\tilde{\boldsymbol{x}}$, i.e.,

$$\tilde{f}(\tilde{\boldsymbol{x}}) = s_{(m)}(\tilde{\boldsymbol{x}}) + e(\tilde{\boldsymbol{x}}), \tag{15}$$

where $e(\tilde{\boldsymbol{x}}) \sim N\left(0, \sigma^2_{(m)}(\tilde{\boldsymbol{x}})\right)$ and $\sigma^2_{(m)}(\tilde{\boldsymbol{x}})$ is the Kriging prediction variance at $\tilde{\boldsymbol{x}}$, and $s^{(m)}(\tilde{\boldsymbol{x}})$ is the Kriging prediction mean as defined in Eq. (3). The selected fake training sample is then added to the real training data set $\mathcal{D}$ to update the surrogate model $s^{(m)}(\tilde{\boldsymbol{x}})$. By repeating these steps one can eventually obtain a training data set $\tilde{\mathcal{D}}_m$ consisting of N real samples and $\tilde{N}$ fake samples, based on which one posterior realization of Kriging model $s^{(m)}(\cdot)$ is constructed. Note that the above procedure is computationally cheap as it does not involve any actual model evaluations.

The sequential selection of the fake training samples and update of the surrogate guarantees the fake response values at adjacent points are close to each other and thus

ensures the achievement of smooth posterior surrogate realizations. For example, assume $\tilde{\boldsymbol{x}}_1$ is the point maximizing the prediction variance, $\tilde{\boldsymbol{x}}_2$ is a point close to $\tilde{\boldsymbol{x}}_1$ (i.e., $s_{(m)}(\tilde{\boldsymbol{x}}_1) \approx s_{(m)}(\tilde{\boldsymbol{x}}_2)$), the prediction variance at $\tilde{\boldsymbol{x}}_2$, $\sigma^2_{(m)}(\tilde{\boldsymbol{x}}_2)$, is thus also large. If we evaluate the fake response values at $\tilde{\boldsymbol{x}}_1$ and $\tilde{\boldsymbol{x}}_2$ simultaneously in one single step, their fake response values (i.e., $\tilde{f}(\tilde{\boldsymbol{x}}_1)$ and $\tilde{f}(\tilde{\boldsymbol{x}}_2)$) obtained by Eq. (14) may be substantially different since the variance of the Gaussian error $e(\tilde{\boldsymbol{x}})$ is large at the two points,. As a result, the posterior surrogate realization based on the actual training samples $\mathcal{D} = \{\boldsymbol{x}_i, f(\boldsymbol{x}_i)\}_{i=1}^N$ and the two fake training samples $\{\tilde{\boldsymbol{x}}_i, \tilde{f}(\tilde{\boldsymbol{x}}_i)\}_{i=1}^2$ will be oscillatory around $\tilde{\boldsymbol{x}}_1$ and $\tilde{\boldsymbol{x}}_2$. Fig. 3(a) illustrates the oscillatory results for an one-dimensional function, $y = \cos(x) + x + 5$, conditioning on 5 training samples. On the contrary, if we evaluate the fake response value at $\tilde{\boldsymbol{x}}_1$ first and then update the surrogate using the actual training samples $\mathcal{D} = \{\boldsymbol{x}_i, f(\boldsymbol{x}_i)\}_{i=1}^N$ and the fake training sample $\left(\tilde{\boldsymbol{x}}_1, \tilde{f}(\tilde{\boldsymbol{x}}_1)\right)$, the prediction variance of the updated surrogate at $\tilde{\boldsymbol{x}}_2$ will be small as $\tilde{\boldsymbol{x}}_2$ is close to the evaluated point $\tilde{\boldsymbol{x}}_1$. As a result, the fake response value $\tilde{f}(\tilde{\boldsymbol{x}}_2)$ obtained by Eq. (14) is close to $\tilde{f}(\tilde{\boldsymbol{x}}_1)$ as $e(\tilde{\boldsymbol{x}}_2)$ is small, resulting in a smooth posterior surrogate realization as shown in Fig. 3 (b).

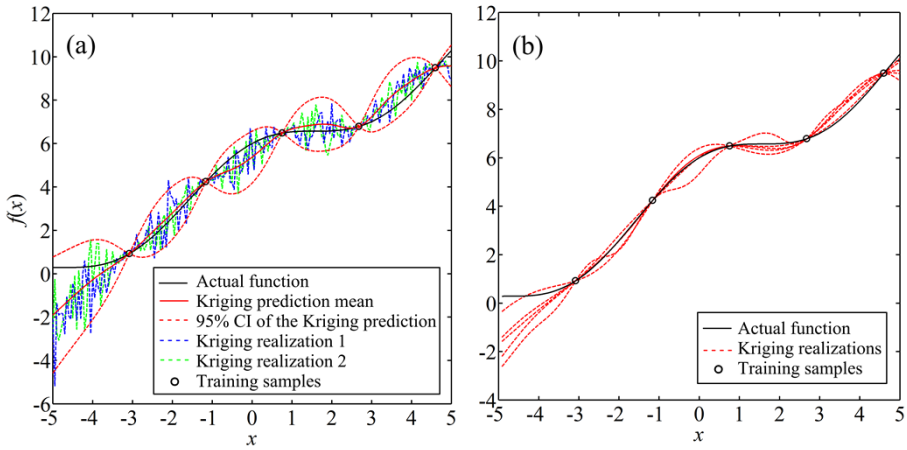

Fig. 3 Illustration of posterior Kriging surrogate realizations for a function conditioning on five training samples. (a) The fake training samples are generated in one single step. The red solid line represents the Kriging prediction mean. CI: confidence interval. (b) The fake training samples are generated through a sequential manner summarized in Algorithm 1.

Algorithm 1: Procedure of generating one posterior Kriging surrogate realization $s_{(m)}(\cdot)$ given the training data set $\mathcal{D}$.

Require: Available training data set $\mathcal{D}$, probability density function of the input vector $p(\boldsymbol{x})$, number of fake training samples $\tilde{N}$

1. $\tilde{\mathcal{D}}_m \leftarrow \mathcal{D}$ ▷ Initialization
2. Draw $\tilde{N}$ random sample of inputs $\tilde{S} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{\tilde{N}}\}$ from $p(\boldsymbol{x})$
3. **for** $j = 1, \ldots, \tilde{N}$ **do**
4. Construct the Kriging surrogate based on $\tilde{\mathcal{D}}_m$, i.e., $s_{(m)}(\boldsymbol{x}) | \tilde{\mathcal{D}}_m$
5. $\tilde{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \tilde{S}} \sigma^2_{(m)}(\boldsymbol{x})$ ▷ Select the input for a new fake sample
6. $\tilde{f}(\tilde{\boldsymbol{x}}) = s_{(m)}(\tilde{\boldsymbol{x}}) + e, \quad e \sim N\left(0, \sigma^2_{(m)}(\tilde{\boldsymbol{x}})\right)$ ▷ Fake response value
7. $\tilde{\mathcal{D}}_m \leftarrow \tilde{\mathcal{D}}_m \cup [\boldsymbol{x}, \tilde{f}(\boldsymbol{x})]$
8. **end for**
9. Construct the Kriging surrogate based on $\tilde{\mathcal{D}}_m$, i.e., $s_{(m)}(\boldsymbol{x}) | \tilde{\mathcal{D}}_m$
10. **return** $s_{(m)}(\cdot)$ ▷ The mth realization of Kriging surrogate

---

Base on each posterior surrogate realization in $\{s_{(m)}\}_{m=1}^{M}$, we can obtain a CDF

estimate of the QoI using the MC in the usual manner, resulting in M distributions $\{F_{(m)}(\cdot)\}_{m=1}^{M}$. From the M distributions we can obtain the required statistics, for instance the sample mean:

$$\overline{F}(q) = \frac{1}{M}\sum_{m=1}^{M} F_{(\mathrm{m})}(q), \tag{16}$$

and can quantify uncertainty about the CDF estimate using the ensemble of distributions.

3. Application to geological CO_2 sequestration simulation

In this section, we apply the TEAD-based Kriging surrogate method to perform an uncertainty analysis of the evolution and leakage of CO_2 in a synthetic GCS model. The accuracy and computational efficiency of our method is evaluated by comparing with the conventional MC simulation without surrogate modeling.

3.1 Problem statement

We consider a zonal homogenized system, i.e., the spatial heterogeneity is zoned according to different geological media, for illustrative purpose. However, the method can be easily extended to heterogeneous cases where the spatial correlated heterogeneity of parameter fields can be parameterized using Karhunen-Loève expansion (Zhang and Lu, 2004) to reduce the number of parameters. Fig. 4 illustrates the conceptual model. In a 2D vertical section, the geological formations are composed of two aquifers separated by an aquitard. The top of the upper aquifer is not the ground surface but assumed to be an impermeable layer. CO_2 is injected into the deep aquifer through a vertical well at a constant rate of 1.5 kg/s for a period of 150 days. The goal of the simulation is to quantify the uncertainty of the total mass of injected CO_2 leaked to the overlying aquifer through a leaky well 45 m away from the injection well. Thus the mass fraction of CO_2 in the upper aquifer after 150 days is chosen as the QoI. The migration of injected CO_2 in the aquifers is simulated using the TOUGH2/ECO2N simulator (Pruess et al., 1999; Pruess, 2005).

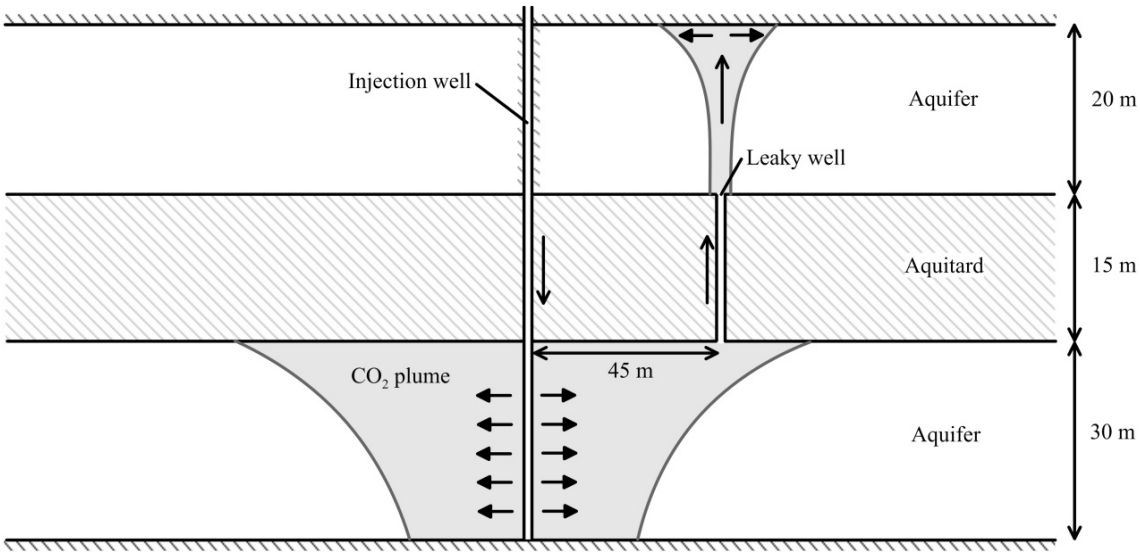

Fig. 4 Conceptual model for CO_2 leakage through an abandoned well.

Due to the high computational demand of MC simulation, a relatively simple GCS model is considered here for the illustration purpose. In the numerical model, we consider a domain with a radius of 300 m. It is discretized into 31 grid blocks with a local refinement near the injection well using a logarithmically increasing radial distance to accurately characterize the near well processes. In the vertical direction, the domain is discretized into 20 layers with the aquitard densified to capture the large pressure gradient between the aquifers and aquitard. The average of initial pressure is assumed to be 15 MPa (corresponds to a depth of 1500 m) with a constant temperature of 60 ℃. The van Genuchten-Mualem (VGM) model (Mualem, 1976; Van Genuchten, 1980) and van Genuchten (VG) function (Van Genuchten, 1980) are chosen as the relative permeability and capillary pressure functions, respectively. The lateral boundary conditions are assumed to be constant and equals to the initial conditions. The top and bottom boundaries are assumed to be no-flow boundaries. More information about the model parameters is listed in Table 1.

Table 1 Hydrogeologic parameters for the GCS model.

Parameter	Aquifer	Aquitard	Leaky well
Temperature (℃)		60	
Average pressure (MPa)		15	
Brine		0.03	
Compressibility (Pa^{-1})		4.5e-10	
Porosity	0.25	0.05	0.3
Permeability (m^2)	uncertain†	1e-18	1e-12
Irreducible gas saturation		0.05	
Air entry pressure (kPa)	19.61	19.61	uncertain
Irreducible water saturation			
VGM model	0.2	0.3	0.15
VG function	0.0	0.0	0.0
Pore size distribution index			
VGM model		0.457	
VG function	0.457	0.457	uncertain

†The anisotropy ratio between the vertical and horizontal permeability is also considered uncertain.

To illustrate the performance of the TEAD-based surrogate method, we consider four uncertain parameters that may greatly influence the QoI, including the aquifer permeability (k) and its anisotropy ratio (r_a) between the vertical and horizontal directions, and two parameters in the VG function for the leaky well, i.e., the pore size distribution index (λ) and the air entry pressure (P_0). The leaky well is modeled as porous media with a high permeability and the aquitard is thought to be difficult to permeate thus its permeability is assumed to be known. The ranges of the four uncertain parameters are listed in Table 2. We assume that the parameters are uniformly distributed within ranges listed in Table 2.

Table 2 Ranges, means, and standard deviations for the three uncertain parameters.

Parameter	$\log_{10}(k)$	r_a	λ	$\log_{10}(1/P_0)$
Range	[-13, -12]	[0.1, 1.0]	[0.2, 0.6]	[-5, -3]

Note: k denotes the reservoir permeability; r_a is the anisotropy ratio of the vertical and horizontal permeability; λ and P_0 represent the pore size distribution index and air entry pressure in the VG function for the leaky well, respectively.

To demonstrate TEAD's performance in efficiently generating accurate surrogate models, we also use the space-filling design LHS to generate the training samples for the surrogate model. We randomly generate 2000 MC realizations by direct GCS model runs to compute the reference statistics. We also use these 2000 samples to compute the RMSE accuracy of the surrogate model, although the TEAD algorithm does not require validation points for termination.

3.2 Results and discussions

The TEAD algorithm starts from an initial sample set consisting of 17 sample points at the corners and center of the domain. In each iteration, additional 16 new samples are selected based on the score function defined in Eq. (9). The sampling process is terminated when the error indicator $\mathrm{RMSE}_{\mathrm{new}}$ is less than 0.02 in two successive iterations. Fig. 5 shows the RMSE decay evaluated on the 2000 samples as the sample size increases; the design process terminates after 10 iterations as the stopping criterion is met and a total of 177 samples are collected.

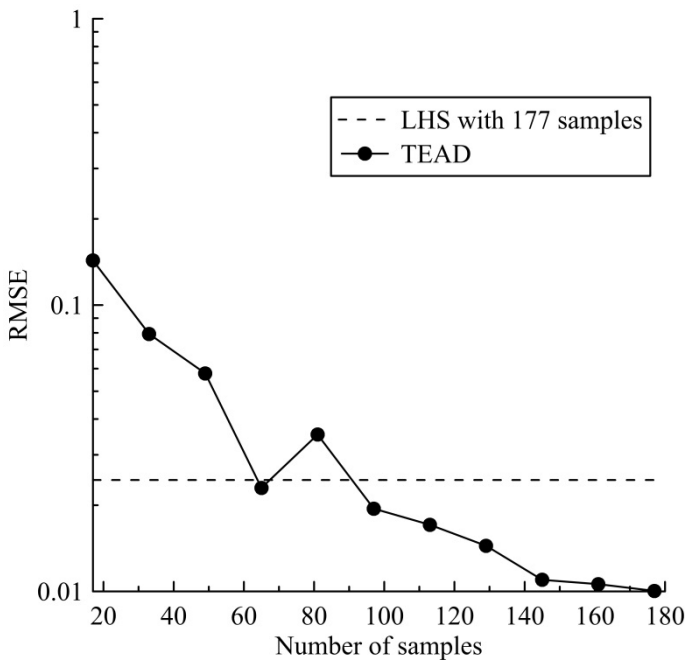

Fig. 5 RMSE decay as the sample size increases for the TEAD algorithm. The lateral dash line represents the RMSE accuracy of the surrogate model based on 177 samples generated by the LHS design. At each iteration of TEAD, 16 new samples are selected.

In addition, we calculate the RMSE of the surrogate model based on 177 samples generated by the LHS method. The result is also shown in Fig. 5 in a dashed line. Note that the LHS method is a one-shot space-filling design so it generates samples at once not iteratively. For a fair comparison, we compare the final RMSE achieved by both the TEAD and the LHS designs using 177 samples. It is observed that LHS achieves a RMSE of only 0.025, much higher than that of TEAD (0.010). To reach the RMSE of 0.025, the TEAD only needs about 90 samples, resulting in almost 50% reduction in the number of GCS model evaluations. This indicates that, by adaptively selecting samples according to the nonlinear features of the response surface rather than uniformly sampling, TEAD can greatly reduce the number of GCS model evaluations required to construct accurate surrogate models. In practice, the large-scale multi-phase flow models of CO_2 sequestration are computationally intensive models, whose simulations can take up to hours or even days. Therefore, the reduction in the number of model evaluations can lead to significant computational savings.

We then test the TEAD-based surrogate method's performance in UQ of the GCS model. The CDF is used to quantify the uncertainty of the leakage of CO_2 based on the MC sampling. In the MC simulation, we execute the constructed Kriging surrogate model instead of the actual GCS model, which means that the surrogate model is used as a full replacement of the GCS model and thus no extra model evaluations are required. Thus the computational cost of the surrogate method is mainly from the evaluation of the 177 training samples for surrogate construction. The accuracy of the estimated CDF from the surrogate method is evaluated by comparing with the reference CDF from the conventional MC simulation with direct 2000 GCS model runs. The surrogate uncertainty in the estimated CDF from the surrogate model is quantified using the method presented in Section 2.3. Fig. 6 shows the changes of mean (μ_{QoI}) and standard deviation (σ_{QoI}) of the QoI along the number of MC samples. It is observed that the μ_{QoI} and σ_{QoI} approach a steady value after

1000 model runs and are stabilized with the sample size of 2000. Thus CDF from MC simulation with 2000 GCS model runs are used as the reference solution.

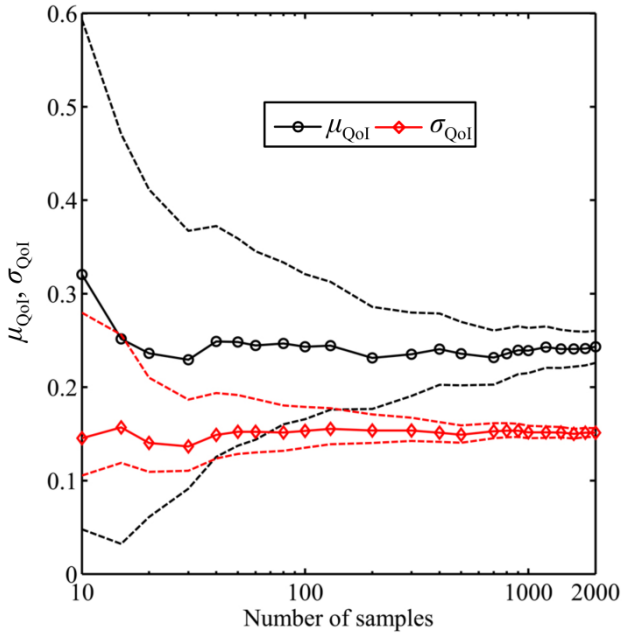

Fig. 6 Changes of mean (μ_{QoI}) and standard deviation (σ_{QoI}) of the fraction of the mass CO_2 in the upper aquifer after 150 days along the number of MC samples, where the dash lines represent the 95% confidence interval of the estimations.

We assess the accuracy of our TEAD-based method from two aspects. First we compare the QoI values from the surrogate model and the GCS model simulated at the 2000 MC samples; and then we compare the CDFs approximated from the surrogate-based MC simulation and the conventional MC simulation basing on the same 2000 samples to rule out the randomness inherent in the sampling process. As a comparison, we also summarize the results of LHS-based surrogate. Fig. 7 summarizes the results of the point-to-point comparison. Fig. 7(a) shows the pairwise comparison of the QoI values from the LHS-based surrogate model and the GCS model. It is observed that although the QoI values have a relatively strong correlation with a correlation coefficient (R^2) of 0.9887, as a whole, the QoI values from the LHS-based surrogate are larger than the actual values in the range [0, 0.3], resulting in an inaccurate surrogate model. On the contrary, it can be seen in Fig. 7(b) that the QoI values from the TEAD-based surrogate model and the GCS model have a stronger

correlation with a R^2 value of 0.9972. Moreover, the data points in Fig. 7(b) are very closely distributed around the 1:1 agreement line, suggesting the high accuracy of the TEAD-based surrogate model in approximation of the GCS model.

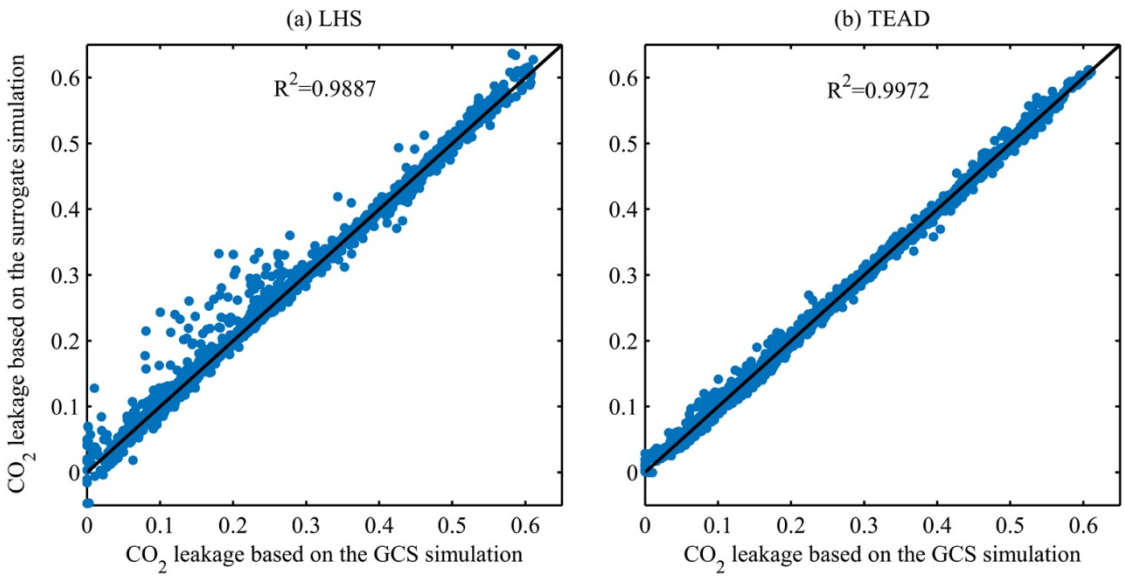

Fig. 7 Pairwise comparison of the CO_2 leakage based on the GCS and the surrogate model evaluated at the 2000 parameter samples. (a) LHS-based surrogate and (b) TEAD-based surrogate with 177 training samples.

Fig. 8 shows the approximated CDFs obtained from the surrogate-based MC simulation and the GCS model-based MC simulation. The surrogate-based curves (red solid lines) are the mean CDF calculated from 400 posterior Kriging surrogate realizations. The corresponding 95% confidence interval calculated from the 400 CDF estimates from the surrogate realizations is depicted with red dashed lines. It is observed that the CDF estimated from the LHS-based surrogate model is far from accurate when the QoI in the range [0.15, 0.3] due to inaccurate surrogate approximation in this range. The LHS-based surrogate model tends to overestimate the QoI in this range (Fig. 7a), thus the reference line is above the 97.5th percentile of the surrogate estimated line (the zoom-in plot of Fig. 8a). On the contrary, the mean CDF estimated from the TEAD-based surrogate model is almost identical to the reference line, which is enveloped within the small 95% confidence interval of the surrogate prediction. Besides, the narrow confidence interval suggests a high

reliability for this estimate. It is noted that to achieve the nearly identical results based on the 2000 GCS model simulations, our adaptive surrogate-based UQ method only uses 177 GCS model runs in the surrogate construction, resulting in a substantial computational gain. On the other hand, with the same 177 GCS model runs, the conventional MC simulation produces a very poor CDF approximation as shown by blue lines in Fig. 8. This once again suggests the advantage of using the adaptive surrogate method in the UQ.

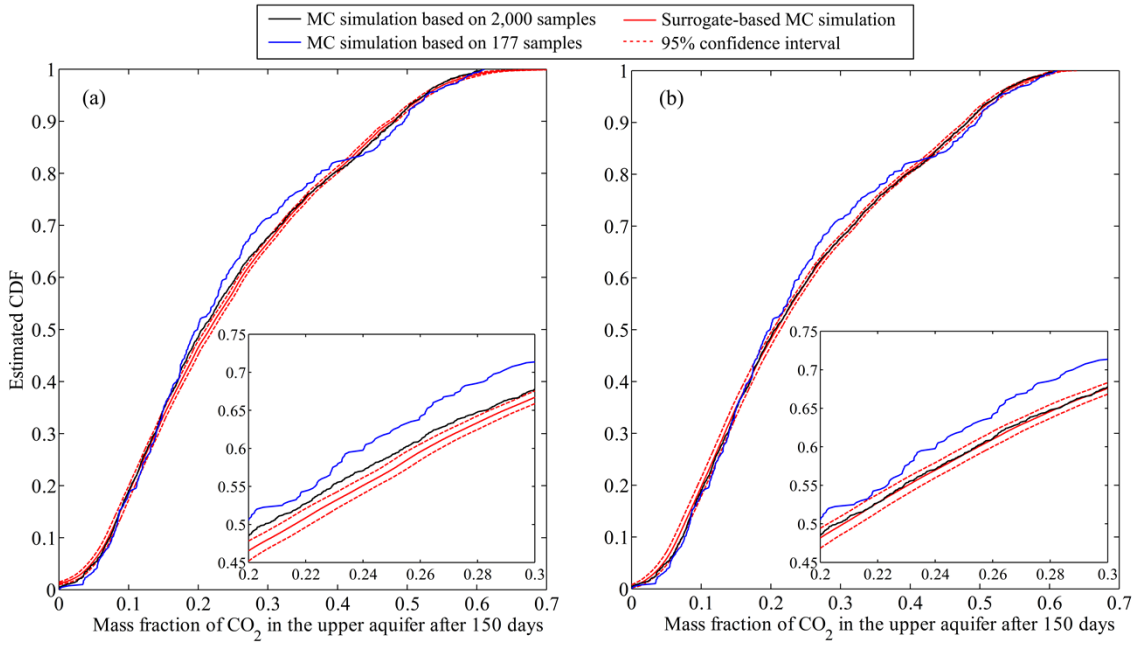

Fig. 8 Comparison of estimated CDFs of the fraction of the mass CO_2 in the upper aquifer after 150 days from (a) LHS-based and (b) TEAD-based surrogate methods with the conventional MC simulation. The surrogate models are constructed using 177 samples. The red dashed lines represent the 2.5th and 97.5th percentiles of CDF estimations from 400 posterior Kriging surrogate realizations.

The results indicate that the combination of the TEAD experimental design with the Kriging surrogate method provides an effective and efficient way to perform an uncertainty analysis in GCS modeling. First, the employment of adaptive design TEAD reduces the computational cost in actual model evaluations for surrogate construction. Second, in the MC sampling process, the generated surrogate model is used as a full replacement of the actual model, which means no more actual model simulation is needed in this process and it can be performed rather fast. As a result,

the computational cost of UQ is significantly reduced. In addition, basing on the prediction variance of the Kriging surrogate, we are able to efficiently quantify the surrogate uncertainty in the UQ results and provide a reliability estimation of the solution.

4. Conclusions

In this work, we propose an adaptive Kriging surrogate method to efficiently solve the uncertainty quantification (UQ) problems in multi-phase flow models of geological carbon sequestration (GCS). The Kriging surrogate model is efficiently constructed using a Taylor expansion-based adaptive design (TEAD) algorithm that combines a distance-based exploration criterion and a Taylor expansion-based exploitation criterion to search for informative training samples. The TEAD algorithm can adaptively place more samples in poorly approximated highly nonlinear areas and in sample-sparse regions, thus greatly reducing the number of actual model evaluations required for surrogate construction compared to conventional space-filling design methods. In addition, our Kriging surrogate method can effectively quantify the surrogate uncertainty. Presenting the surrogate-based UQ analysis with surrogate uncertainty improves the UQ results explanablility and reliability.

The performance of the proposed method is demonstrated using a synthetic GCS model. In evaluation, we use the conventional MC sampling without surrogate modeling as a reference. Results indicate that the cumulative density function estimated by our method is very close to the reference and the surrogate uncertainty bound calculated by our method can enclose the reference. However, in achievement of this accurate UQ results, our method used only 10% of the GCS model runs required by the conventional MC. These results suggest that our method greatly improves the computational efficiency in UQ and provides reliable UQ results with accurately estimated surrogate uncertainty. The GCS model considered in this work is relatively fast in order to quickly test the proposed method in a reasonable time. When applying our method to computationally more expensive models, larger computational

savings can be expected. As our method is general purpose, the application to other environmental and geological models should be straightforward.

Acknowledgements

We are grateful to the High Performance Computing Center of Nanjing University for doing the numerical calculations in this manuscript on its blade cluster system. This work was financially supported by the National Key Research and Development Program of China (No. 2018YFC0406402) and the National Nature Science Foundation of China grants (No. 41672229 and U1503282).

Computer code availability

The Matlab source codes of TEAD are available on GitHub at https://github.com/njujinchun/Codes-of-TEAD. More information about the program can be found in the file "README.md." The data used can be reproduced using TOUGH2 (http://esd1.lbl.gov/research/projects/tough/) and the needed scripts are available upon request from the corresponding author.

References

Asher, M., Croke, B., Jakeman, A., Peeters, L., 2015. A review of surrogate models and their application to groundwater modeling. Water Resources Research 51, 5957-5973.

Ashraf, M., Oladyshkin, S., Nowak, W., 2013. Geological storage of CO_2: Application, feasibility and efficiency of global sensitivity analysis and risk assessment using the arbitrary polynomial chaos. International Journal of Greenhouse Gas Control 19, 704-719.

Audigane, P., Gaus, I., Czernichowski-Lauriol, I., Pruess, K., Xu, T., 2007. Two-dimensional reactive transport modeling of CO_2 injection in a saline aquifer at the Sleipner site, North Sea. American Journal of Science 307, 974-1008.

Ballio, F., Guadagnini, A., 2004. Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. Water Resources Research, 40.

Bandilla, K.W., Celia, M.A., Birkholzer, J.T., Cihan, A., Leister, E.C., 2015. Multiphase modeling of geologic carbon sequestration in saline aquifers. Groundwater 53, 362-377.

Bilionis, I., Zabaras, N., 2016. Bayesian Uncertainty Propagation Using Gaussian Processes. Handbook of Uncertainty Quantification, 1-45.

Bliznyuk, N., Ruppert, D., Shoemaker, C., Regis, R., Wild, S., Mugunthan, P., 2008. Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation. Journal of Computational and Graphical Statistics 17(2), 270–294.

Buhmann, M.D., 2000. Radial basis functions. Acta Numerica 9, 1-38.

Crevillén-García, D., Wilkinson, R., Shah, A., Power, H., 2017. Gaussian process modelling for uncertainty quantification in convectively-enhanced dissolution processes in porous media. Advances in Water Resources 99, 1-14.

De Silva, G.P.D., Ranjith, P.G., Perera, M.S.A., 2015. Geochemical aspects of CO_2 sequestration in deep saline aquifers: A review. Fuel 155, 128-143.

Eason, J., Cremaschi, S., 2014. Adaptive sequential sampling for surrogate model generation with artificial neural networks. Computers & Chemical Engineering 68, 220-232.

Erickson, C.B., Ankenman, B.E., Sanchez, S.M., 2018. Comparison of Gaussian process modeling software. European Journal of Operational Research 266(1), 179-192.

Espinet, A., Shoemaker, C., & Doughty, C., 2013. Estimation of plume distribution

for carbon sequestration using parameter estimation with limited monitoring data. Water Resources Research 49, 4442–4464.

Espinet, A.J., Shoemaker, C. A., 2013. Comparison of optimization algorithms for parameter estimation of multi - phase flow models with application to geological carbon sequestration. Advances in Water Resources 54, 133–148.

Demuth, H.B., Beale, M.H., De Jess, O., Hagan, M.T., 2014. Neural network design. Martin Hagan.

Harvey, O.R., Qafoku, N.P., Cantrell, K.J., Lee, G., Amonette, J.E., Brown, C.F., 2012. Geochemical implications of gas leakage associated with geologic CO_2 storage-A qualitative review. Environmental Science & Technology 47, 23-36.

Holloway, S., 1997. An overview of the underground disposal of carbon dioxide. Energy Conversion and Management 38, S193-S198.

Hou, Z., Bacon, D.H., Engel, D.W., Lin, G., Fang, Y., Ren, H., Fang, Z., 2014. Uncertainty analyses of CO_2 plume expansion subsequent to wellbore CO_2 leakage into aquifers. International Journal of Greenhouse Gas Control 27, 69-80.

IPCC, 2005. Special report on carbon dioxide capture and storage. Intergovernmental Panel on Climate Change, Working Group III. Cambridge University Press, Cambridge, UK and New York, USA.

Jia, W., McPherson, B.J., Pan, F., Xiao, T., Bromhal, G., 2016. Probabilistic analysis of CO_2 storage mechanisms in a CO_2-EOR field using polynomial chaos expansion. International Journal of Greenhouse Gas Control 51, 218-229.

Jones, D.G., Beaubien, S.E., Blackford, J.C., Foekema, E.M., Lions, J., De Vittor, C., West, J.M., Widdicombe, S., Hauton, C., Queirós, A.M., 2015. Developments since 2005 in understanding potential environmental impacts of CO_2 leakage from geological storage. International Journal of Greenhouse Gas Control 40, 350-377.

Jordan, A.B., Stauffer, P.H., Harp, D., Carey, J.W., Pawar, R.J., 2015. A response surface model to predict CO_2 and brine leakage along cemented wellbores. International Journal of Greenhouse Gas Control 33, 27-39.

Ju, L., Zhang, J., Meng, L., Wu, L., Zeng, L., 2018. An adaptive Gaussian process-based iterative ensemble smoother for data assimilation. Advances in Water Resources 115, 125-135.

Kitanidis, P.K., 2015. Persistent questions of heterogeneity, uncertainty, and scale in subsurface flow and transport. Water Resources Research 51, 5888-5904.

Liu, H., Ong, Y.S., Cai, J., 2018. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. Structural and Multidisciplinary Optimization, 57, 393–416.

Lophaven, S.N., Nielsen, H.B., Søndergaard, J., 2002. DACE-A MATLAB kriging toolbox, version 2.0.

Lu, D., Ricciuto, D., Evans, K., 2018. An efficient Bayesian data-worth analysis using a multilevel Monte Carlo method. Advances in Water Resources 113, 223-235.

Lu, D., Zhang G., Webster, C., Barbier, C., 2016. An improved multilevel Monte Carlo method for estimating probability distribution functions in stochastic oil reservoir simulations. Water Resources Research, 52, 9642-9660.

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

Mo, S., Lu, D., Shi, X., Zhang, G., Ye, M., Wu, J., Wu, J., 2017. A Taylor expansion-based adaptive design strategy for global surrogate modeling with applications in groundwater modeling. Water Resources Research, 53, 10802-10823.

Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resources Research 12, 513-522.

Oakley, J., O'hagan, A., 2002. Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika, 89, 769-784.

Oladyshkin, S., Class, H., Helmig, R., Nowak, W., 2011. A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. Advances in Water Resources 34, 1508-1518.

Pan, F., McPherson, B. J., Dai, Z., Jia, W., Lee, S. Y., Ampomah, W., Viswanathan, H., Esser, R., 2016. Uncertainty analysis of carbon sequestration in an active CO_2-EOR field. International Journal of Greenhouse Gas Control 51, 18-28.

Pau, G.S.H., Zhang, Y., Finsterle, S., 2013. Reduced order models for many-query subsurface flow applications. Computational Geosciences 17, 705-721.

Pau, G.S.H., Zhang, Y., Finsterle, S., Wainwright, H., Birkholzer, J., 2014. Reduced order modeling in iTOUGH2. Computers & Geosciences 65, 118-126.

Pruess, K., 2005. ECO2N: A TOUGH2 fluid property module for mixtures of water, NaCl, and CO_2. Lawrence Berkeley National Laboratory, Berkley, CA.

Pruess, K., Oldenburg, C., Moridis, G., 1999. TOUGH2 user's guide version 2. Lawrence Berkeley National Laboratory, Berkley, CA.

Rasmussen, C.E., Williams, C.K., 2006. Gaussian processes for machine learning. MIT Press Cambridge.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water

resources. Water Resources Research 48.

Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989. Design and analysis of computer experiments. Statistical Science, 409-423.

Schnaar, G., Digiulio, D.C., 2009. Computational modeling of the geologic sequestration of carbon dioxide. Vadose Zone Journal 8, 389-403.

Shahsavani, D., Grimvall, A., 2011. Variance-based sensitivity analysis of model outputs using surrogate models. Environmental Modelling & Software 26, 723-730.

Shi, X., Ye, M., Curtis, G.P., Miller, G.L., Meyer, P.D., Kohler, M., Yabusaki, S., Wu, J., 2014. Assessment of parametric uncertainty for groundwater reactive transport modeling. Water Resources Research 50, 4416-4439.

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering & System Safety 93, 964-979.

Sun, A.Y., Jeong, H., González-Nicolás, A., Templeton, T.C., 2018. Metamodeling-based approach for risk assessment and cost estimation: Application to geological carbon sequestration planning. Computers & Geosciences 113, 70-80.

Sun, A.Y., Zeidouni, M., Nicot, J.P., Lu, Z., Zhang, D., 2013. Assessing leakage detectability at geologic CO_2 sequestration sites using the probabilistic collocation method. Advances in Water Resources 56, 49-60.

Tartakovsky, D.M., 2013. Assessment and management of risk in subsurface hydrology: A review and perspective. Advances in Water Resources 51, 247-260.

Tian, L., Wilkinson, R., Yang, Z., Power, H., Fagerlund, F., Niemi, A., 2017. Gaussian process emulators for quantifying uncertainty in spreading predictions in heterogeneous media. Computers & Geosciences 105, 113-119.

van der Herten, J., Couckuyt, I., Deschrijver, D., Dhaene, T., 2015. A fuzzy hybrid sequential design strategy for global surrogate modeling of high-dimensional computer experiments. SIAM Journal on Scientific Computing 37, A1020-A1039.

Van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44, 892-898.

Wang, D., Dong, B., Breen, S., Zhao, M., Qiao, J., Liu, Y., Zhang, Y., Song, Y., 2015. Review: Approaches to research on CO_2/brine two-phase migration in saline aquifers. Hydrogeology Journal 23, 1-18.

Xu, T., Valocchi, A.J., Ye, M., Liang, F., 2017. Quantifying model structural error: Efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model. Water Resources Research, 53,

4084-4105.

Zeng, X., Ye, M., Burkardt, J., Wu, J., Wang, D., Zhu, X., 2016. Evaluating two sparse grid surrogates and two adaptation criteria for groundwater Bayesian uncertainty quantification. Journal of Hydrology 535, 120-134.

Zeng, X., Ye, M., Wu, J., Wang, D., Zhu, X., 2018. Improved nested sampling and surrogate‐enabled comparison with other marginal likelihood estimators. Water Resources Research 54, 797-826.

Zhan, C.-s., Song, X.-m., Xia, J., Tong, C., 2013. An efficient integrated approach for global sensitivity analysis of hydrological model parameters. Environmental Modelling & Software 41, 39-52.

Zhang, D., Lu, Z., 2004. An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loève and polynomial expansions. Journal of Computational Physics 194, 773-794.

Zhang, G., Lu, D., Ye, M., Gunzburger, M., Webster, C., 2013. An adaptive sparse‐grid high‐order stochastic collocation method for Bayesian inference in groundwater reactive transport modeling. Water Resources Research 49, 6871-6892.

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

Zhang, Y., Sahinidis, N.V., 2013. Uncertainty quantification in CO_2 sequestration using surrogate models from polynomial chaos expansion. Industrial & Engineering Chemistry Research 52, 3121-3132.

Zhu, Y., Zabaras, N., 2018. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. Journal of Computational Physics 366, 415-447.