Computers & Geosciences 105 (2017) 113–119

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Gaussian process emulators for quantifying uncertainty in CO_2 spreading predictions in heterogeneous media

Liang Tian[a,*], Richard Wilkinson[b], Zhibing Yang[a], Henry Power[c], Fritjof Fagerlund[a], Auli Niemi[a]

[a] *Air, Water and Landscape Sciences, Department of Earth Sciences, Villavägen 16, SE-752 36 Uppsala University, Sweden*
[b] *School of Mathematics and Statistics, University of Sheffield, UK*
[c] *Faculty of Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

ARTICLE INFO

Keywords:
Permeability heterogeneity
Karhunen-Loève expansion
Monte Carlo
Uncertainty analysis
Gaussian process emulator

ABSTRACT

We explore the use of Gaussian process emulators (GPE) in the numerical simulation of CO_2 injection into a deep heterogeneous aquifer. The model domain is a two-dimensional, log-normally distributed stochastic permeability field. We first estimate the cumulative distribution functions (CDFs) of the CO_2 breakthrough time and the total CO_2 mass using a computationally expensive Monte Carlo (MC) simulation. We then show that we can accurately reproduce these CDF estimates with a GPE, using only a small fraction of the computational cost required by traditional MC simulation. In order to build a GPE that can predict the simulator output from a permeability field consisting of 1000s of values, we use a truncated Karhunen-Loève (K-L) expansion of the permeability field, which enables the application of the Bayesian functional regression approach. We perform a cross-validation exercise to give an insight of the optimization of the experiment design for selected scenarios: we find that it is sufficient to use 100s values for the size of training set and that it is adequate to use as few as 15 K-L components. Our work demonstrates that GPE with truncated K-L expansion can be effectively applied to uncertainty analysis associated with modelling of multiphase flow and transport processes in heterogeneous media.

1. Introduction

Planning and operation of a carbon dioxide capture and storage (CCS) project requires reliable model predictions concerning the fate of the stored CO_2. Carefully conducted numerical simulations are critical for the understanding of the associated coupled physical and chemical processes (Pruess and García, 2002; Juanes et al., 2006; Doughty, 2007; Dai et al., 2016; Bacon et al., 2016; Xiao et al., 2016). An important additional complication arises from the geological heterogeneity of the target formation, such as stratigraphic architecture and facies distribution, which is difficult to estimate from the limited number of observations available (i.e., from the sparse networks of primarily vertical investigation wells) in a deterministic manner (Ambrose et al., 2007; Tsang et al., 2008; Gershenzon et al., 2015; Yang et al., 2015; Ritzi et al., 2016; Tian et al., 2016b; Ampomah et al., 2016). Therefore, robust and computationally effective methods for dealing with the uncertainty arising from the geological heterogeneity are in great need. In general, two components contribute to the modelling uncertainty for CO_2 geological storage: (1) input uncertainty, including the aforementioned parameter uncertainties (unknown geology), and (2) model uncertainty, or "structural uncertainty" according to the conventional hydrological modelling terminology (Renard et al., 2010), as modelling approaches are developed under different conceptual and methodological frameworks, involving various approximations and simplifications. An example on the latter is the work reported by Nordbotten et al. (2012), where a benchmark simulation case was run with various numerical codes and effort was made to evaluate the significance of deviated solutions from various modelling strategies and assumptions. In the present work, we focus on the input uncertainty.

Standard geostatistical techniques are used to resolve the input uncertainty when evaluating reservoir CO_2 storage performance. For example, the Umbrella Point power plant model (based on the Frio formation) was created using TProGs program by Doughty and Pruess (2004) where multiple two-dimensional stochastic representations of fluvial depositional settings were picked deliberately to reproduce realistic three-dimensional geologic structures. A sequential indicator simulation approach was used by Flett et al. (2007) to create realistic shale facies distribution for 3-D notional marine sand system models with varying net-sand-to-gross-shale ratios. A sequential Bayesian simulation technology was used by Claprood et al. (2014) in construct-

[*] Corresponding author.
E-mail addresses: liang.tian@geo.uu.se (L. Tian), r.d.wilkinson@sheffield.ac.uk (R. Wilkinson), zhibing.yang@hyd.uu.se (Z. Yang), henry.power@nottingham.ac.uk (H. Power), fritjof.fagerlund@geo.uu.se (F. Fagerlund), auli.niemi@geo.uu.se (A. Niemi).

http://dx.doi.org/10.1016/j.cageo.2017.04.006
Received 16 September 2016; Accepted 23 April 2017
Available online 25 April 2017

ing a porosity distribution for a 3-D model of Beauharnois Formation to understand its CO_2 storage potential. In terms of the characterization of the spatial permeability distribution, Han et al. (2010) created multiple two-dimensional permeability fields with inclusion of low permeability lenses using a sequential Gaussian simulation approach. Discussions on effects of the permeability heterogeneity include the contributions from Jahangiri and Zhang (2011) with a focus on the plume distribution, and from Lengler et al. (2010) with a focus on small-scale heterogeneity (<100 m). Using a macroscopic invasion percolation model, Yang et al. (2013) performed a detailed parametric sensitivity study on upscaled capillary pressure-saturation-relative permeability relationships for CO_2 migration in multimodal heterogeneous media. A more recent sensitivity study was reported by Tian et al. (2016a) where the parameters controlling the spatial correlation structures of the permeability fields were systematically analysed so as to understand their effects on CO_2 storage performance.

A Monte Carlo simulation method is normally used when a deterministic description of the model input cannot be used (James, 1980). In this approach, multiple, mutually different but equiprobable realizations of the parameter field are generated, the model problem simulated for all of them, and the output analysed in terms of the statistics of the outputs. The method has been proved viable for the simulation of geological storage of CO_2 (Jahangiri and Zhang, 2011; Deng et al., 2012; Dai et al., 2014; Tian et al., 2016a). However, an obvious limitation for the method is the high computational cost, which limits the number of possible runs for large-scale, long-term simulations of CO_2 migration in 3-D heterogeneous medium. This in turn violates the underlying criteria of the Monte Carlo approach, which require the model to be run at many input configurations in order to accurately infer the uncertainty in the model predictions. Therefore, new reduced-order models that can capture the essential behaviour of the fully physically based models, yet avoiding the prohibitive computational cost of them are of great interest. A general overview on surrogate modelling in water resources was given by Razavi et al. (2012). More recently, Liu et al. (2013) developed geostatistical reduced order models (GROMs) in the parameter domain to solve under-determined inverse problems addressing subsurface multiphase transport.

In this paper, we propose a Bayesian approach for uncertainty analysis (UA), that is, the forward propagation of uncertainty through a model. We focus on simulators such as TOUGH2/ECO2N (Pruess et al., 1999; Pruess and Spycher, 2007), which are used for the numerical simulation of CO_2 injection into deep heterogeneous aquifers. These numerical models (called the *simulator*) are deterministic, meaning they will always produce the same output if the input is known exactly, and thus can be regarded as mathematical functions $f(\cdot)$. As we are uncertain about the input Z (i.e., the true permeability is unknown), this uncertainty is transferred to $f(Z)$, so that we are uncertain about the best prediction. The objective of uncertainty analysis is therefore to estimate the distribution of $f(Z)$, given a distribution for inputs Z.

2. Methodology

We present the modelling problem and describe the quantities of interest in Section 2.1. In Section 2.2, we present the method to simulate the random permeability field. In Section 2.3, we describe the Gaussian process emulation (GPE) methodology and it application to our problem. A complete procedure to our implementation of GPE is given in Section 2.4. In Section 2.5 we describe the use of GPE for uncertainty analysis.

2.1. Modelling of CO_2 migration in a heterogeneous aquifer

We consider supercritical CO_2 injection from a vertical borehole, and we simulate CO_2 migration until the CO_2 plume front reaches the monitoring well at the far end of the domain (Fig. 1). The simulations are performed using the TOUGH2/ECO2N code (Pruess et al., 1999; Pruess and Spycher, 2007). The quantities of interest are the breakthrough time (BT) and the total mass (TM) of the injected CO_2. For the numerical experiments where we want to address the uncertainty caused by heterogeneity, we vary the correlation length of the randomly generated permeability fields, but use a fixed standard deviation (see Section 2.2). A more detailed description is given in the *Supporting Information*(SI).

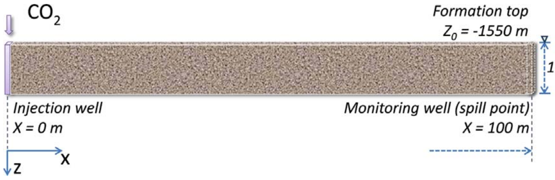

Fig. 1. Conceptual model of the simulation domain (Tian et al., 2016a).

In this work, we use the notation Z to denote the permeability spatial field and want to find the distribution of $f(Z)$ given the distribution of Z, where $f(\cdot)$ represents the simulator output (e.g., either the total mass or the breakthrough time of the CO_2). In other words, our objective is to estimate the cumulative distribution functions (CDFs)

$$F(y) = \mathbb{P}(f(Z) \le y). \tag{1}$$

The CDFs can be estimated using a Monte Carlo (MC) approach if sufficient computer power is available. If $Z_1, \ldots, Z_n$ is a large sample from log-Gaussian random field (log-GRF) we are using to model the heterogeneous permeability field, then the empirical CDF (ECDF),

$$\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{f(Z_i)\le y}, \tag{2}$$

is an unbiased estimator of the CDF. Here, $\mathbb{1}_A$ is an indicator function taking value 1 if event A occurred and 0 otherwise.

2.2. Modelling the heterogeneous permeability field

We consider a representation of Z on a two-dimensional mesh grid with a finite resolution 100×20. The x in the notation $Z(x)$ is the location coordinate vector, emphasizing that Z is location dependent. Our prior model for Z is

$$\log Z \sim N(\mu, \Sigma), \tag{3}$$

where we specify Σ through a covariance function that describes the permeability covariance between any two locations in the domain, i.e., $\Sigma_{ij} = c(x_i, x_j)$ for some covariance function c, and spatial locations x_i and x_j. Several techniques exist to simulate realizations from this distribution, including circulant embeddings, Karhunen-Loève expansions and stochastic collocation (Graham et al., 2011). The method of Karhunen-Loève (K-L) decomposition is used in our work. The Karhunen-Loève theorem says that $Z(x)$ admits a representation of the form

$$Z(x) = \sum_{i=1}^{\infty} \xi_i \lambda_i \phi_i(x) \tag{4}$$

where the λ_i and $\phi_i(x)$ are the ordered eigenvalues and eigenfunctions of the covariance function respectively, and the ξ_i are independent $N(0, 1)$ random variables. Note that if interest lies solely in the value of Z on a finite grid of n values (as in our case), then this reduces to a finite sum of n terms, and the K-L decomposition provides an exact decomposition of the correlation function on the discrete grid (Crevillén-García et al., 2017). To reconstruct $Z(x)$, only the $\{\xi_i\}_{i=1}^{n}$ need to be saved, since λ_i and ϕ_i are determined by the covariance function and thus remain the same throughout the uncertainty analysis. The simulator is then considered as a function of $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^\top$ instead of Z, i.e., $f(Z) \equiv f(\boldsymbol{\xi})$.

In order to calculate the CDFs of the target quantities and evaluate the performance of the GP emulator, two datasets are generated for

Table 1
Case specifications and results for model selection.

Case No.		1	2	3
Correlation length		0.075	0.15	0.30
size of MC set	N_{MC}	10,000	10,000	10,000
size of training set	n_{train}	800	400	400
dimension of the training set	d_{train}	30	20	20
$CRPS_{BT,\text{Matérn}}$		0.00640	0.00193	0.00153
$CRPS_{BT,\ \text{SE}}\ (d_{train}=20)$		0.00108	0.00187	0.00135
$CRPS_{TM,\text{Matérn}}$		0.00490	0.00766	0.00975
$CRPS_{TM,\ \text{SE}}\ (d_{train}=20)$		0.02489	0.02508	0.02534

each of three selected scenarios where we vary the correlation-length of the unknown permeability fields (Table 1, first three rows). The first dataset consists of 10^4 input-output pairs and is used to produce a MC estimate of the CDF; the second dataset consists of a smaller number of numerical simulations and is used for training the emulator. The overall procedure is illustrated in Fig. 2 and is further explained in the following section.

2.3. Gaussian process emulation

An emulator (Kennedy and O'Hagan, 2000) is a statistical model that closely mirrors a simulator. It is built using an ensemble of input-output pairs $\{X_i, y_i\}_{i=1}^{N}$ and can be used to predict the simulator output for any new input. The most popular approach to building emulators is to use a Gaussian process (GP) (Rasmussen and Williams, 2006), which are equivalent to the *kriging* models used in geostatistics (Stein, 1999). Gaussian processes describe an infinite collection of random variables, and can be thought of as distributions over functions (Rasmussen and Williams, 2006; Crevillén-García et al., 2017). A GP is fully specified by its mean and covariance functions (Rasmussen and Williams, 2006).

In our case, direct application of GP would be computationally costly for that a 2000 dimensional input space would require thousands of training samples (as the *hyperparameters* associated with each input component are estimated from the simulator data by solving an optimization problem, e.g., Crevillén-García et al., 2017). Instead, we can construct a GP emulator by exploiting the spatial structure in Z provided by the exact decomposition of Z on a discrete grid. If we order the eigenvalues in Eq. (4) so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, then we can achieve a form of data compression by truncating the expansion to the first d terms

$$\widetilde{Z}(x) = \sum_{i=1}^{d} \xi_i \lambda_i \phi_i(x), \tag{5}$$

and thus representing the permeability in a lower dimensional space. This truncation explains the most variance and achieves the minimum mean square error amongst all such approximations. We exploit this truncation in order to build a reduced order emulator from $\widetilde{Z}$ rather than Z, which is equivalent to building an emulator with input $\boldsymbol{\xi} = (\xi_1, \ldots \xi_d)^{\top}$.

The emulator requires the simulator to be run a small number of times (n_{train}) at carefully selected inputs (design points) to create a set of *training inputs* (See Fig. 2). Because the simulation of Z is based on a truncated K-L expansion, the training ensemble is a set $\{\boldsymbol{\xi}_i, y_i\}_{i=1}^{n_{train}}$ where each $\boldsymbol{\xi}_i \in \mathbb{R}^d$. Space-filling designs (McKay et al., 1979; Morris and Mitchell, 1995) are recommended for GP models, as GP predictions essentially interpolate based on the distance to a few of the nearest training points. We use the maximin Latin hypercube designs which maximise the minimum distance between any two points in the training set. We will examine the optimal value of d and n_{train} using predictive performance measures in Section 4.

The implementation of GPs require that we specify prior mean and covariance functions. We use a constant mean function and choose between the squared exponential and Matérn covariance functions. The *hyperparameters* involved in these two terms are estimated through *training* using type II maximum likelihood (Rasmussen and Williams, 2006). We use the GPstuff implementation of Gaussian processes (Vanhatalo et al., 2012), which are a set of MATLAB codes integrating Gaussian process models for Bayesian analysis. Notice that the GP covariance function (also called the kernel) should be distinguished from the one mentioned earlier in describing the spatial correlation of the permeability field.

2.4. GP emulation with K-L truncation

We summarize the procedure as follows:

1. Choose design $\boldsymbol{\xi}_{i=1}^{n}$ using a maximin Latin hypercube design where $\boldsymbol{\xi} \in R^N$.
2. Run simulator to obtain training set $\{\boldsymbol{\xi}_i, y_i\}_{i=1}^{n}$. We then truncate each $\boldsymbol{\xi}$ to the first d elements. The value of d will be optimized in Step 6.
3. Pick a prior mean function $m(\boldsymbol{\xi}) = \mathbb{E}[\hat{f}(\boldsymbol{\xi})]$ and covariance function $k(\boldsymbol{\xi}, \boldsymbol{\xi}') = \mathbb{C}\text{ov}(\hat{f}(\boldsymbol{\xi}), \hat{f}(\boldsymbol{\xi}'))$ where $\hat{f}(\cdot)$ is the emulator. For example, the *square exponential* (SE) covariance function is

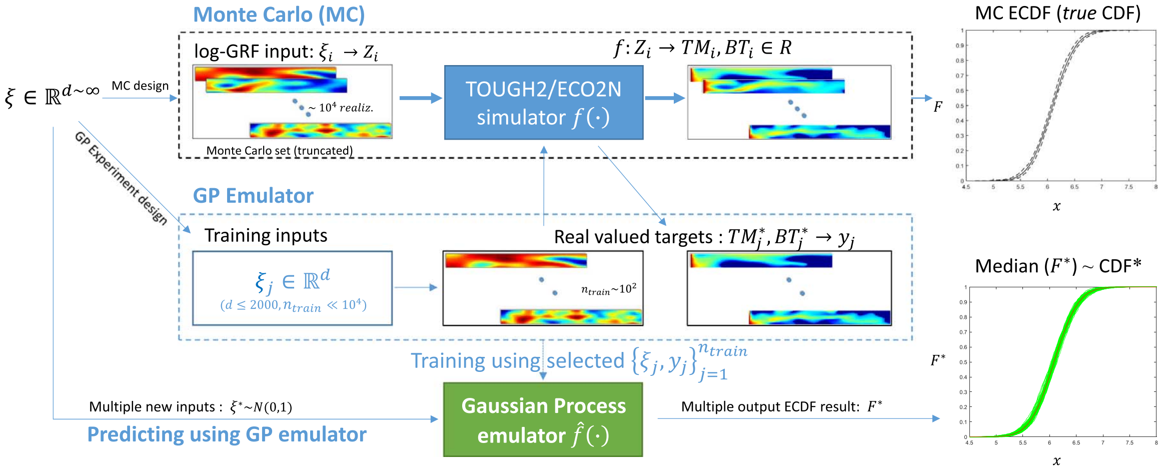
Monte Carlo (MC)
log-GRF input: $\xi_i \to Z_i$
$f: Z_i \to TM_i, BT_i \in R$
$\xi \in \mathbb{R}^{d\sim\infty}$
MC design
~ 10^4 realiz.
Monte Carlo set (truncated)
TOUGH2/ECO2N simulator $f(\cdot)$
F
MC ECDF (*true* CDF)
x
GP Experiment design
GP Emulator
Training inputs
$\xi_j \in \mathbb{R}^d$
$(d \leq 2000, n_{train} \ll 10^4)$
n_{train}~10^2
Real valued targets : $TM_j^*, BT_j^* \to y_j$
Median (F^*) ~ CDF*
F^*
Training using selected $\{\xi_j, y_j\}_{j=1}^{n_{train}}$
Multiple new inputs : ξ^*~$N(0,1)$
Predicting using GP emulator
Gaussian Process emulator $\hat{f}(\cdot)$
Multiple output ECDF result: F^*
4.5 5 5.5 6 6.5 7 7.5 8
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Fig. 2. Comparing procedures for estimating CDFs using Monte Carlo simulation (TOUGH2/ECO2N) and Gaussian process emulation. The thickness of the arrow illustrates the relative computational cost.

$$k(\boldsymbol{\xi}, \boldsymbol{\xi}') = \sigma^2 \exp\left(-\frac{1}{2}\frac{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2}{\lambda}\right)$$

where λ is a length scale hyper parameter, and σ^2 a variance parameter. We denote the GP prior by:

$$\hat{f}(\boldsymbol{\xi}) \sim \mathcal{GP}(m(\boldsymbol{\xi}), k(\boldsymbol{\xi}, \boldsymbol{\xi}')).$$

4. Update the GP to find the posterior mean (m*) and covariance functions (k*) using equations:

$$m^*(\boldsymbol{\xi}) = m(\boldsymbol{\xi}) + t(\boldsymbol{\xi})^\top K^{-1}(\mathbf{y} - \mathbf{m}),$$
$$k^*(\xi^*, \xi^*) = k(\boldsymbol{\xi}, \boldsymbol{\xi}) - t(\boldsymbol{\xi})^\top K^{-1} t(\boldsymbol{\xi})$$

where $K_{ij} = k(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)$ is the Gram matrix, $t(\boldsymbol{\xi})^\top = (k(\boldsymbol{\xi}_1, \boldsymbol{\xi}), \ldots, k(\boldsymbol{\xi}_n, \boldsymbol{\xi}))$, and $\mathbf{m}$ and $\mathbf{y}$ are the vectors of simulator responses and their prior mean for the emulator. Note that the posterior is a GP conditioned on the training set.

5. Optimize the hyperparameters, such as λ, σ^2 in SE, by maximising the type II maximum likelihood (see Rasmussen and Williams, 2006).
6. Optimize the choice of d, the covariance function, etc, using cross-validation to estimate a measure of the predictive performance.

2.5. Using GP for UA

Once we have a GP emulator of the simulator, we can use it to predict the simulator CDF and to quantify the uncertainty in our estimate. To estimate the CDFs, we use the procedure suggested in Oakley and O'Hagan (2002). This involves drawing sample functions $\{f_j\}_{j=1}^{L}$ from the GP that are consistent with the training data by adding in new design points $\{\boldsymbol{\xi}_i^*\}_{i=1}^{1000}$, and simulating a value for the response from the GP emulator. We then update the emulator to take into account the fake simulated data. The placement and number of additional design points is chosen so as to make the uncertainty in the simulated functions f_j essentially zero. We then estimate the CDF for each simulated function using Monte Carlo in the usual manner, giving us L realizations $F_1^*, \ldots F_L^*$. From this we use the median of the CDFs as a point estimate, and can calculate uncertainty about our estimates using the ensemble of CDFs.

3. Results

3.1. Estimating the CDF

Each quantity of interest (total mass (TM) or breakthrough-time (BT)) from each of the three cases (three different models for the unknown permeability field) is considered as a standalone problem. As the training set is based on a Latin hypercube design, we use a fixed number of training points (Table 1) to construct each of the three GP structures. For each emulated ECDF curve, 1000 random sample points are first generated using a pseudorandom number (vector) generator in Matlab assuming a dimension corresponding to $d_{train} = 30$ (Case 1) or $d_{train} = 20$ (Cases 2 and 3). Then, this set of random inputs, together with the corresponding training pairs, were used to feed the designated GP structure in order to produce/draw one sample from the posterior distribution. For each quantity of interest, 100 posterior samples (L=100) were used to calculate the median ECDF. Note that this is computationally cheap as it does not involve running the TOUGH2/ECO2N simulator.

Fig. 3 shows the breakthrough time for *Case 1*. The GP curve is the median CDF calculated from the 100 posterior samples. The confidence intervals of the MC CDF are omitted for visual clarity. The dashed lines (posterior credible intervals) indicate that the MC CDF is enveloped within the emulator confidence intervals. Excellent matches are observed: for all cases examined, the median GP curves replicate the MC ones almost exactly. The mean CRPS (Continuous Rank Probability Score, see the SI) for the three correlation length cases are 0.00640, 0.00193 and 0.00153, respectively. A similar procedure was used for the total CO_2 mass (TM) at the breakthrough time. The TM ECDF curves from the MC are also well predicted by the median GP results. The TM result exhibits a slightly less good match in comparison to the observation from the BT, especially for the lower and upper tail of the ECDF. However, the 5th to the 95th percentiles of the GP prediction agree closely with the MC results. The CRPSs for three tested cases are, respectively, 0.00490, 0.00766 and 0.00975.

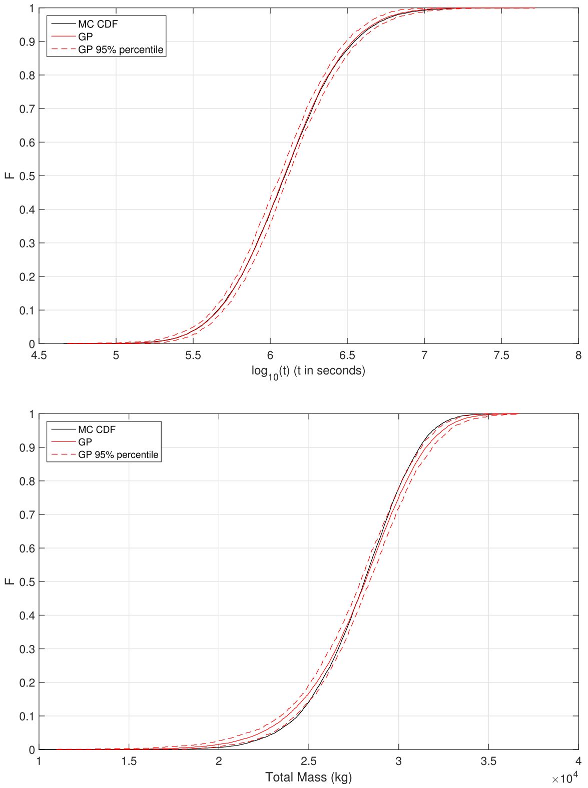

Fig. 3. Comparison of GP emulation vs. Monte Carlo simulations. Top: breakthrough time (BT) recorded in seconds; bottom: the total mass of CO_2 (TM).

Note that for TM smaller CRPSs are observed for Case 1 in comparison to the other cases (Table 1) due to a larger number of training points ($n_{train,case1} = 800$) and the higher dimension of the training inputs ($d_{case1} = 30$ KL components). Note also that the CRPSs for BT are noticeably smaller in comparison to the TM ones (one order of magnitude). Excellent agreement is observed for BT results (Fig. 3). For Case 2 and Case 3, the results are visually similar to Case 1 and are therefore not included for space considerations.

3.2. Cross validation

At the initial stage of the experimental design, two key factors are very difficult to determine beforehand, namely the size of the training set (n_{train}) and its dimension (d_{train}, the number of K-L components retained for the prediction). Using leave-one-out cross validation (LOO-CV, see also SI) can guide us in tackling these issues. For each GP, LOO-CV has been performed to estimate the predictive accuracy of the emulator in two steps: Step 1, a training set with fixed size is selected and the predictive performance measured using the Dawid score (DS), which can be thought of as being similar to the log-likelihood (see Wilkinson et al. (2011), and the SI). This score is then plotted as a function of the number of K-L components; Step 2, the number of K-L components is now fixed and the predictive performance is plotted as a function of the size of the training set.

The DS estimated using LOO-CV are plotted as a function of the number of K-L components in Fig. 4. It is found that by using a fixed size of the training set for all cases, the DS score becomes stabilized when using more than 15 K-L components ($d_{train} \geq 15$). When using exactly 15 K-L components for each case to fit the GPs, the DS score appears to become stabilized when using a training set with more than 100 design points ($n_{train} \geq 100$, see Fig. 5).

4. Discussion

The investigated two dimensional model domain has 2000 elements representing a spatially correlated heterogeneous permeability field. Uncertainty analysis using the classical MC method requires that the already computational demanding simulator to be run for as many as 10^4 times.

For the GP emulator approach to UA, the main part of computational cost comes from the simulator runs needed for the training inputs. GP posterior sampling has in comparison virtually no computational cost. In this section we discuss the design and the construction of the GP emulator.

4.1. Model configuration

One very important aspect of using GP emulation is the choice of the covariance function that defines the nearness or similarity in the input space (Rasmussen and Williams, 2006). In other words, how similar $f(\mathbf{x})$ is likely to be to $f(\mathbf{x}')$ when $\mathbf{x}$ is close to $\mathbf{x}'$. The covariance function can be any positive definite function, so that it generates a valid covariance matrix for any set of inputs. Some of the commonly used functions are the squared exponential covariance function (SE) and the Matérn class of covariance functions. The SE covariance function generates samples that are infinitely differentiable, whereas the Matérn covariance function (with $\nu = \frac{3}{2}$ degrees of freedom) generates samples that are only once differentiable. It can be hard to judge in advance what the more appropriate model might be, but we can use CV scores to guide the choice. We constructed alternative GPs using both for each of the cases examined in Section 3 (see Table 1). The ECDFs calculated using the Matérn covariance function ($\nu = \frac{3}{2}$) exhibit smaller CRPS values in comparison to the ones calculated using SE. For the emulation of BT, there is no noticeable difference between using the SE or Matérn covariance functions. However, for TM the Matérn exhibits much better predictive performance. Notice that the choice of d_{train} (the dimension of training points, in our case equivalent to the number of K-L components) will affect the performance of the GP emulator, depending on the number of training points (n_{train}). We note that the choice of covariance function can affect the performance of the GPE, and that more complex covariance functions can be obtained by combining covariance functions (see Rasmussen and Williams (2006), for example). A detailed discussion is beyond the scope of the current work, but can be found in Crevillén-García (2016).

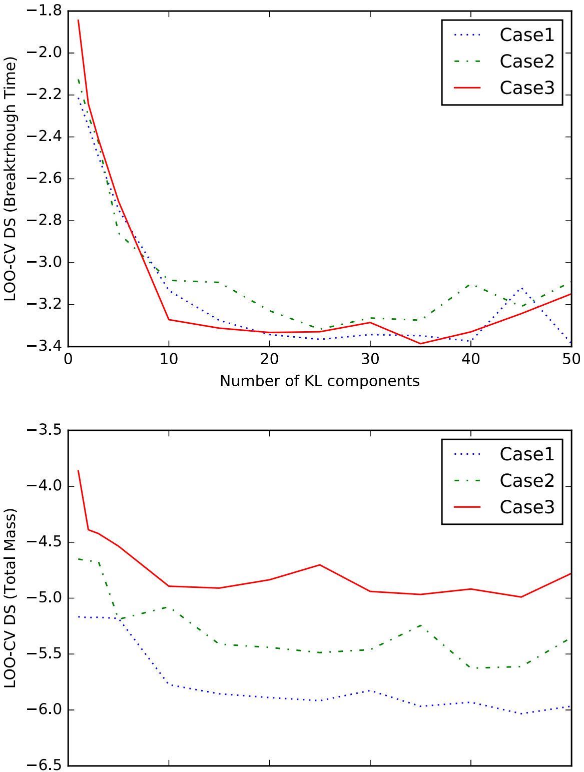

Fig. 4. Dawid scores indicating prediction accuracy (estimated using LOO-CV) vs. number of K-L components retained (d_{train}).

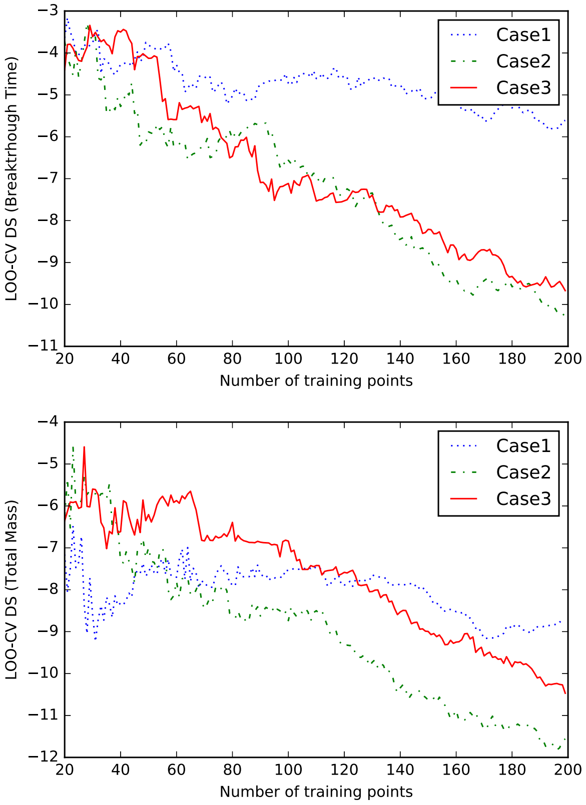

Fig. 5. LOO-CV Scores vs. é the size of the training set (n_{train}).

4.2. Cross-validation and optimization

We would like to use the smallest number of the training inputs possible to create an emulator that meets our accuracy requirements. To investigate this, we use the method of cross-validation (CV). The idea is to split the training set into two disjoint sets, one of which is used for the training and the other is used for the validation of the emulator. Notice that such splits can be done repeatedly in multiple ways (k-fold CV), one extreme case is when k=n, also known as leave-one-out cross-validation (LOO-CV). We can use CV scores to choose the optimum input dimensionality (the number of K-L coefficients, d_{train}) and the number of design training points (n_{train}) to be used in the GP. The evaluation is done by looking at the variance of the predicted value in LOO-CV as well as the Dawid score for the overall prediction error.

In our calculations, the size of the training ensemble is 800 for Case

1 but 400 for Case 2 and Case 3. The reason for using more training sets in Case 1 is that the correlation length for the log-Gaussian permeability field model is smaller in Case 1 than in Cases 2 and 3. Thus, the permeability is autocorrelated over shorter distances, and so we need more K-L components to describe the variation well, and consequently we need a larger training ensemble to build an adequate emulator. For predicting the BT ECDF (Fig. 4), using 15 K-L components provides good results, whereas for predicting the TM ECDF, around 20 K-L components is preferred. The indication is that the calculations of breakthrough time and total mass for the injection simulation of CO_2 are two very different processes.

A priori, it is difficult to provide a precise value for an adequate or appropriate number of training points required for a GP, as, to the best of our knowledge, a priori estimation of the error is not possible for GPs. Optimization of the design would mean changing the space filling design, which would mean drawing new samples ξ_i from $\mathbb{R}^{d=2000}$. To understand whether this design improved the GP performance, the simulator (TOUGH2/ECO2N) would need to be rerun so as to generate the corresponding new training ensemble. In other words, one would need to build new GPs based on additional simulator runs in order to understand the potential gain from optimization. This would be extremely computationally costly, and so a different approach has been used here.

Considering Case 1, for example, where we have generated 800 training pairs (n_{train} = 800), we start by building an emulator, $GP_{0,j=20}$, using a random draw (whilst trying to retain some of the space filling properties of the design) of j=20 training points from initial set of 800. A first DS score can then be calculated for $GP_{0,j=20}$ using LOO-CV. By randomly adding one training point at a time from the remaining training pairs, we can iteratively create new emulators, $GP_{i,j=20+i}$. The resulting Dawid scores then reflect how the predictive performance improves as the sample size increases. It should be noted that Latin-hypercube sampling has been used to create the initial 800 points. The re-sample of the existing Latin-hypercube set should be path-independent. Fig. 5 shows the decreasing trend of DS score reflecting that more information is provided by the training set as the sample size increases. It can been seen that 100 training pairs would be needed for Case 1 when building a GP for BT ECDF using only 15 K-L components. Note that the pattern of TM LOO-CV result for Case 1 (Fig. 5, lower panel) is different from the other cases. We further extended the LOO-CV test for Case 1 and the decreasing trend in the DS score was confirmed (Fig. 6). This indicates that for heterogeneous domain with a smaller correlation length, a larger training set may be needed for constructing the GP so as to achieve a similar predictive performance.

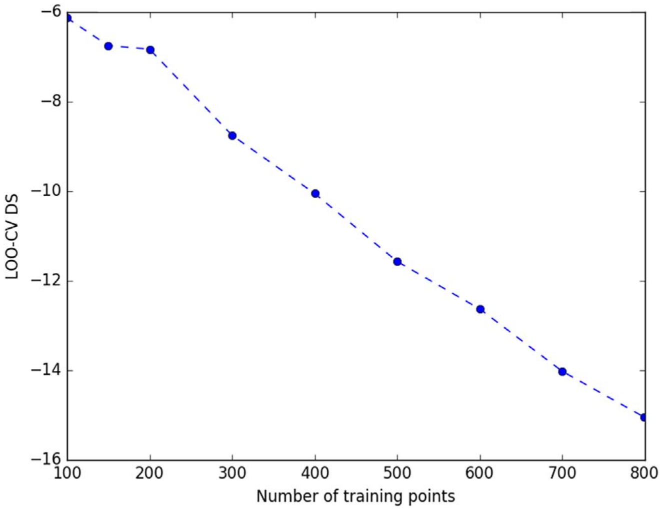

Fig. 6. LOO-CV Scores vs. the size of the training set (n_{train}), Case 1.

4.3. Using GP for uncertainty analysis

The output from each GP constructed in Section 3 is a collection of random variables indexed by $\boldsymbol{\xi}$. An assumption has been made that the spatial distribution of the heterogeneous field can be adequately described by $\boldsymbol{\xi}$. In a geostatistics perspective, the conventional perception of correlation length (λ), standard deviation (σ) and the descriptive covariance function (see SI) of the permeability field can all be interpreted as possible projections of $\boldsymbol{\xi}$.

We use standalone GPs in predicting the ECDF for each uncertain output of interest. It is worth noting that the two outputs, the breakthrough time and the total mass, are fundamentally different processes. Fig. 3 shows that the breakthrough time is log-normally distributed, while the total mass follows a normal distribution. The GP emulator prediction is noticeably better for $\log_{10}(BT)$ than for TM. This difference in reproducing the MC results may indicate that the dependence of TM on the underlying permeability field is more complex than that of BT. Additional metrics apart from the K-L expansion parameter (or alternative methods) describing the permeability fields may be needed to improve the uncertainty analysis of the total CO_2 mass.

We have shown that the use of GP for UA, in our case exploring the ECDFs of BT and TM, results in considerably lower computational cost compared to classical MC analyses. By improving the experimental design, it is possible to further improve the model performance.

5. Concluding remarks

We have carried out uncertainty analysis of the simulations of CO_2 injection and migration into a deep heterogeneous saline aquifer using both MC simulation and GP emulation. We have shown how GPEs can successfully be used to predict ECDFs of the breakthrough time and total CO_2 mass, replicating the ECDF estimates obtained using Monte Carlo simulation, at only a small fraction of the computational cost. The GPs automatically provide confidence intervals for the estimates of the CDF, which compare well to those calculated from classical MC. Our work demonstrates that GP emulators with truncated Karhunen-Loève expansion can be effectively applied to uncertainty analysis associated with modelling of multiphase flow and transport processes in heterogeneous media.

We have also examined the issues surrounding experimental design, including the possibilities to further optimize the GP. An optimum design may need to re-sample the input space, and therefore need additional simulator runs. To address this, an alternative approach has been taken by down-sampling the training set. The results from the cross-validation exercise indicate significant performance gain from potential optimization. This information provides a good starting point for further applications.

We have treated the two outputs, namely the CO_2 breakthrough time and the total CO_2 mass as two independent processes, and built standalone GPs for each one. It is possible to construct a single GP with multiple outputs (Alvarez et al., 2011), and this may provide one future perspective for exploring the internal physical mechanism for a complex system. Another future aspect would be to use simulations of varying fidelity and then to use multilevel emulation to further increase the accuracy of the GPE (cf. multi-level Monte Carlo in Giles et al. (2015)).

We have also explored the indication from modelling of heterogeneous media and identified that the conventional perception on correlation length is, from a geostatistic perspective, a matter of parameter bounds and dimensions. Finally, we note that future work is needed to address the limitation associated with the use of truncated Karhunen-Loève expansion, which is a smooth representation of the random field, for application to real reservoirs which often exhibit multi-scale permeability heterogeneity.

Acknowledgment

The authors would like to acknowledge Prof. Andrew Cliffe at the School of Mathematical Sciences, University of Nottingham for his invaluable contributions to this work. Unfortunately Prof. Cliffe passed away during the preparation of this manuscript. He will be remembered for his great qualities as a researcher and a friend.

The work has been supported by the European Community's Seventh Framework Programme (FP7) MUSTANG (No. 227286), PANACEA (No. 282900) and TRUST (No. 309067) projects. The computations were performed on resources at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) and at Chalmers Centre for Computational Science and Engineering (C3SE), both provided by the Swedish National Infrastructure for Computing (SNIC).

The authors thank Alberto Guadagnini and two anonymous reviewers for their helpful comments.

References

Alvarez, M.A., Rosasco, L., Lawrence, N.D., 2011. Kernels for Vector-Valued Functions: A Review. arXiv preprint arXiv:1106.6251, 4(3), 1–37.

Ambrose, W.A., Lakshminarasimhan, S., Holtz, M.H., Núñez-López, V., Hovorka, S.D., Duncan, I., 2007. Geologic factors controlling CO_2 storage capacity and permanence: case studies based on experience with heterogeneity in oil and gas reservoirs applied to CO_2 storage. Environ. Geol. 54 (8), 1619–1633.

Ampomah, W., Balch, R., Cather, M., Rose-Coss, D., Dai, Z., Heath, J., Dewers, T., Mozley, P., 2016. Evaluation of co2 storage mechanisms in co2 enhanced oil recovery sites: application to morrow sandstone reservoir. Energy Fuels 30 (10), 8545–8555.

Bacon, D.H., Qafoku, N.P., Dai, Z., Keating, E.H., Brown, C.F., 2016. Modeling the impact of carbon dioxide leakage into an unconfined, oxidizing carbonate aquifer. Int. J. Greenh. Gas Control 44, 290–299.

Claprood, M., Gloaguen, E., Sauvageau, M., Giroux, B., Malo, M., 2014. Adapted sequential Gaussian simulations with Bayesian approach to evaluate the CO_2 storage potential in low porosity environment. Greenh. Gases: Sci. Technol. 4 (6), 761–776.

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

Crevillén-García, D., 2016. Uncertainty Quantification for Flow and Transport in Porous Media (Ph.D. thesis). University of Nottingham.

Dai, Z., Stauffer, P.H., Carey, J.W., Middleton, R.S., Lu, Z., Jacobs, J.F., Hnottavange-Telleen, K., Spangler, L.H., 2014. Pre-site characterization risk analysis for commercial-scale carbon sequestration. Environ. Sci. Technol. 48 (7), 3908–3915.

Dai, Z., Viswanathan, H., Middleton, R., Pan, F., Ampomah, W., Yang, C., Jia, W., Xiao, T., Lee, S.-Y., McPherson, B., Balch, R., Grigg, R., White, M., 2016. CO_2 accounting and risk analysis for CO_2 sequestration at enhanced oil recovery sites. Environ. Sci. Technol. 50 (14), (6b01744 acs.est.).

Deng, H., Stauffer, P.H., Dai, Z., Jiao, Z., Surdam, R.C., 2012. Simulation of industrial-scale CO_2 storage: multi-scale heterogeneity and its impacts on storage capacity, injectivity and leakage. Int. J. Greenh. Gas Control 10 (0), 397–418.

Doughty, C., 2007. Modeling geologic storage of carbon dioxide: comparison of non-hysteretic and hysteretic characteristic curves. Energy Convers. Manag. 48 (6), 1768–1781.

Doughty, C., Pruess, K., 2004. Modeling supercritical carbon dioxide injection in heterogeneous porous media. Vadose Zone J. 3 (3), 837–847.

Flett, M., Gurton, R., Weir, G., 2007. Heterogeneous saline formations for carbon dioxide disposal: impact of varying heterogeneity on containment and trapping. J. Pet. Sci. Eng. 57 (1–2), 106–118.

Gershenzon, N.I., Ritzi, R.W., Dominic, D.F., Soltanian, M., Mehnert, E., Okwen, R.T., 2015. Influence of small-scale fluvial architecture on CO_2 trapping processes in deep brine reservoirs. Water Resour. Res. 51 (10), 8240–8256.

Giles, M.B., Nagapetyan, T., Ritter, K., 2015. Multilevel monte carlo approximation of distribution functions and densities. SIAM J. Uncertain. Quantif. 3, 267–295.

Graham, I., Kuo, F., Nuyens, D., Scheichl, R., Sloan, I., 2011. Quasi-monte carlo methods for elliptic pdes with random coefficients and applications. J. Comput. Phys. 230 (10), 3668–3694.

Han, W.S., Lee, S.-Y., Lu, C., McPherson, B.J., 2010. Effects of permeability on CO_2 trapping mechanisms and buoyancy-driven CO_2 migration in saline formations. Water Resour. Res. 46 (7), W07510.

Jahangiri, H.R., Zhang, D., 2011. Effect of spatial heterogeneity on plume distribution and dilution during CO_2 sequestration. Int. J. Greenh. Gas Control 5 (2), 281–293.

James, F., 1980. Monte-Carlo theory and practice. Rep. Progress. Phys. 43 (9), 1145–1189.

Juanes, R., Spiteri, E.J., Orr, F.M., Blunt, M.J., 2006. Impact of relative permeability hysteresis on geological CO_2 storage. Water Resour. Res. 42 (12), 1–13.

Kennedy, M.C., O'Hagan, A., 2000. Predicting the output from a complex computer code when fast approximations are available. Biometrika 87 (1), 1–13.

Lengler, U., De Lucia, M., Kühn, M., 2010. The impact of heterogeneity on the distribution of CO_2: numerical simulation of CO_2 storage at Ketzin. Int. J. Greenh. Gas Control 4 (6), 1016–1025.

Liu, X., Zhou, Q., Birkholzer, J., Illman, W.A., 2013. Geostatistical reduced-order models in underdetermined inverse problems. Water Resour. Res. 49 (10), 6587–6600.

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

Morris, M.D., Mitchell, T.J., 1995. Exploratory designs for computational experiments. J. Stat. Plan. Inference 43 (3), 381–402.

Nordbotten, J.M., Flemisch, B., Gasda, S.E., Nilsen, H.M., Fan, Y., Pickup, G.E., Wiese, B., Celia, M.A., Dahle, H.K., Eigestad, G.T., Pruess, K., 2012. Uncertainties in practical simulation of CO_2 storage. Int. J. Greenh. Gas Control 9 (0), 234–242.

Oakley, J., O'Hagan, A., 2002. Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika 89 (4), 769–784.

Pruess, K., García, J., 2002. Multiphase flow dynamics during CO_2 disposal into saline aquifers. Environ. Geol. 42 (2–3), 282–295.

Pruess, K., Spycher, N., 2007. ECO_2N-A fluid property module for the TOUGH2 code for studies of CO_2 storage in saline aquifers. Energy Convers. Manag. 48 (6), 1761–1767.

Pruess, K., Oldenburg, C., Moridis, G., 1999. TOUGH2 User's Guide, Version 2.0. Report LBNL-43134, Lawrence Berkeley National Laboratory, Berkeley, California.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. University Press Group Limited.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 (7), W07401.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour. Res. 46 (5), W05521.

Ritzi, R.W., Freiburg, J.T., Webb, N.D., 2016. Understanding the (co)variance in petrophysical properties of CO_2 reservoirs comprising sedimentary architecture. Int. J. Greenh. Gas Control 51, 423–434.

Stein, M.L., 1999. Interpolation of Spatial Data Springer Series in Statistics. Springer, New York, New York, NY.

Tian, L., Yang, Z., Fagerlund, F., Niemi, A., 2016a. Effects of permeability heterogeneity on CO_2 injectivity and storage efficiency coefficient. Greenh. Gases: Sci. Technol. 6 (1), 112–124.

Tian, L., Yang, Z., Jung, B., Joodaki, S., Erlström, M., Zhou, Q., Niemi, A., 2016b. Integrated simulations of CO_2 spreading and pressure response in the multilayer saline aquifer of South Scania Site, Sweden. Greenh. Gases: Sci. Technol. 6 (4), 531–545.

Tsang, C.-F., Birkholzer, J., Rutqvist, J., 2008. A comparative review of hydrologic issues involved in geologic storage of CO_2 and injection disposal of liquid waste. Environ. Geol., 54 (8), 1723–1737.

Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen, V., Vehtari, A., 2012. Bayesian Modeling with Gaussian Processes using the GPstuff Tool-box. ArXiv e-prints, 14(1), 48.

Wilkinson, R.D., Vrettas, M., Cornford, D., Oakley, J.E., 2011. Quantifying simulator discrepancy in discrete-time dynamical simulators. J. Agric. Biol. Environ. Stat. 16 (4), 554–570.

Xiao, T., McPherson, B., Pan, F., Esser, R., Jia, W., Bordelon, A., Bacon, D., 2016. Potential chemical impacts of CO_2 leakage on underground source of drinking water assessed by quantitative risk analysis. Int. J. Greenh. Gas Control 50, 305–316.

Yang, Z., Tian, L., Niemi, A., Fagerlund, F., 2013. Upscaling of the constitutive relationships for CO_2 migration in multimodal heterogeneous formations. Int. J. Greenh. Gas Control 19 (0), 743–755.

Yang, Z., Tian, L., Jung, B., Joodaki, S., Fagerlund, F., Pasquali, R., Vernon, R., O'Neill, N., Niemi, A., 2015. Assessing CO_2 storage capacity in the Dalders Monocline of the Baltic Sea Basin using dynamic models of varying complexity. Int. J. Greenh. Gas Control 43, 149–160.