Geological storage of CO_2: global sensitivity analysis and risk assessment using the arbitrary polynomial chaos expansion

Meisam Ashraf[a,b], Sergey Oladyshkin[c], Wolfgang Nowak[c]

[a]*Department of Mathematics, University of Bergen, 5028 Bergen, Norway*
[b]*SINTEF ICT, Department of Applied Mathematics P.O. Box 124 Blindern, N-0314 Oslo, Norway*
[c]*SRC Simulation Technology, Institute for Modelling Hydraulic and Environmental Systems (LH2), University of Stuttgart, Pfaffenwaldring 61, 70569 Stuttgart, Germany.*

Abstract

Geological storage of CO_2 is a proposed interim solution for mitigating the climate change. The considerable costs and potential hazards of the technique require feasibility studies to assess all possible risks involved in the process. Modeling CO_2 storage requires working with large time and space scales, which in practice are accompanied by huge geological uncertainties.

The overall goal of this study is to demonstrate the application and feasibility of global sensitivity analysis with Sobol indices and probabilistic risk assessment via the very recent arbitrary polynomial chaos expansion (aPC) method. We model a typical CO_2 injection scenario implemented in realistic geological realizations. A number of uncertain parameters control the structural heterogeneities with assumed probability distributions. These parameters include the effect of barriers, the aggradation angle, fault transmissibility and regional groundwater effects. They represent realistic types of structural features that have not been analyzed before in this context and manner.

Within the aPC, a so-called response surface is representing how the model output changes with respect to the uncertain parameter values by multi-variate polynomials for all output quantities of interest. We choose the aPC, because the resulting response surface is significantly faster than the full original model, and so facilitates global sensitivity analysis and probabilistic risk assessment even for

Email addresses: `meisam@resman.no` (Meisam Ashraf),
`Sergey.Oladyshkin@iws.uni-stuttgart.de` (Sergey Oladyshkin),
`Wolfgang.Nowak@iws.uni-stuttgart.de` (Wolfgang Nowak)

large problems.

Quantification of input parameter uncertainty is not the primary concern of this work, as it can change from one place to another. Therefore, no generalized conclusions on the reality of geological uncertainties can be drawn from this study.

1. Introduction

In the context of climate change mitigation, geological storage of CO_2 has been proposed as interim solution. The idea has been challenged during the last decades for its costs and potential hazards [1, 2]. A large number of studies has been performed in the industry and research communities to evaluate the safety and feasibility of CO_2 storage, addressing issues such as the status and barriers of CO_2 storage [3], screening and ranking of geological storage sites [4], large-scale impacts of CO_2 injection in deep saline aquifers [5], new solution methodologies for CO_2 leakage [6], the capture project [7], and leakage estimates [8]. Furthermore, many pilot projects have been installed, like In Salah [9], Ketzin [10], and Johansen [11]. A discussion on the experiences from the existing pilot projects is reported in [12].

Yet, there is a big demand for studies which demonstrate the appropriateness of the storage operation. Transparent scientific results are required to communicate the facts and evidences about feasibility and possible risks within public and industry. The large involved time and space scales, however, cause substantial computational issues in such studies (e.g., [13]), and the modeling procedure is accompanied by a huge extent of geological uncertainties (e.g., [14, 15, 16, 17]).

In an approach to quantify the impact of geological heterogeneity on model predictions of multiphase flow in geological formations, a large number of shallow marine depositional realizations has been generated and used in the sensitivity analysis of the impact of geological uncertainties on production forecasting (SAIGUP), see [18, 19, 20] . There, the impact of variable geological parameters has been quantified for oil recovery in different field development scenarios. The main general conclusion of that study is that realistic features of geological uncertainty in modeling (other than typical hydrological parameters) can lead to considerable uncertainties in prediction. Ashraf et al. [21, 22] used a number of SAIGUP realizations to study the impact of geological heterogeneity on the

injection and early migration of CO_2 in a shallow-marine aquifer with a complex, heterogeneous geological structure. That study transferred the significance of some of the geological structural features to the case of CO_2 injection.

In practice, modeling complicated physical phenomena in the subsurface requires stochastic approaches. Uncertainty can exist in different levels, from the formulation of dependency rules in the model to uncertainty about appropriate values for the model input parameters. Uncertainty coming from any source in the modeling procedure propagates through the model to the predicted responses. Ranking the important model parameters based on their influence on the model responses can support a better understanding of the system, and it can result in a better design of subsequent studies on the stochastic nature of the process. Hence, identifying and evaluating the sensitivities and uncertainties of model parameters and their impact on prediction uncertainties and projected risks is a significant task. Sensitivity analysis is known to be the right approach to identify the significance of uncertainty sources within the modeling process [23] and to improve the understanding of model behavior [24]. For example, the European Commission and the United States Environment Protection Agency recommend using sensitivity analysis in the context of extended compact assessment for policy making [25].

Uncertainty sources within the CO_2 storage problem can be classified in different types as geological, physical and operational uncertainties. This work is devoted to geological uncertainties. However the same procedure can be applied to extend the work for other types as well. Here, we use a set of SAIGUP realizations to perform a sensitivity analysis and to assess the risks caused by uncertainties in a choice of parameters that govern the geological structure of the featured shallow-marine deposit.

The goal of this study is to test and demonstrate the applicability of a recent set of methods to a realistic scenario. We choose a stochastic response surface method to project the model response to parameter changes onto high-dimensional polynomials via the arbitrary polynomial chaos expansion (aPC) [23, 26]. Highly similar ideas to the aPC have also been proposed in other scientific areas [27, 28, 29, 30]. As we review in Section 2, the involved orthogonal polynomial basis can be constructed for arbitrary probability distributions of the uncertain parameters. This data-driven approach provides fast convergence [23] in comparison to the classical polynomial chaos expansion (e.g., [31, 32, 33]). Moreover, it avoids the subjectivity of data treatment that would arise when being forced to fall back onto a limited number of theoretical distributions that can be tolerated with previous generalized versions of polynomial chaos expansions [34, 35]. The reduced

model represented by the response surface is significantly faster than the original complex one, and thus provides a promising starting point for global sensitivity analysis, uncertainty quantification, and probabilistic risk assessment.

In the current paper, we use global sensitivity analysis rather than a local one, because local analysis fails to cover the non- linear variation of model responses over the entire range of probability distributions of the input parameters. A practical approach in global sensitivity analysis is to work with the impact of uncertain parameters on prediction variances, because this shows a good success in nonlinear problems [36]. In the current study, we use Sobol indices [24] for sensitivity analysis, which are indeed working with variances. The fact that the aPC based response surface is based on orthonormal polynomials with exploitable known properties [37] substantially simplifies this analysis.

Finally, we perform risk analysis by applying a Monte-Carlo procedure to the response surface. The approximating polynomial is fast enough to be used for a large number of Monte-Carlo realizations. This makes it possible to cover the entire range of variations in the model input described by the assigned probability distributions, and thus provides accurate estimates for the risk in the system. We conclude with a discussion of the results.

The global sensitivity analysis and uncertainty quantification studies for CO_2 storage existing in the literature are concerned with classic hydrological uncertain parameters like porosity, pore volume and permeability as global constants (see for example [15, 26, 38]). To the best of our knowledge, the current study is the first one that implements the proposed mathematical analysis tools on realistic geological structural parameters at reservoir scale. The parameters we consider are the level of barriers presence, aggradation angle, fault transmissibility, and regional groundwater effects. The considered features are the structural and depositional features that dictate the distribution of hydrological parameters such as permeability and porosity, both in terms of value and spatial distribution. These are among the most uncertain geological parameters identified with the SAIGUP study (except the regional groundwater effect, which is specific to this study).

2. Response surface via arbitrary polynomial chaos expansion

Working with uncertain parameters in complex, non-linear and dynamic systems puts a high demand on stochastic tools to analyze the system and to propagate uncertainties through the system. Conceptually straightforward numerical Monte Carlo (MC) techniques are computationally demanding since the statistical accuracy of their predictions depends on the number of realizations used. The

Monte-Carlo estimation error (measured as standard deviation) for output statistics typically decreases only with the square root of the number of realizations used. Using a stochastic response surface is a promising approach in this respect.

Obviously, a response surface can be constructed in different ways, e.g. it can be constructed directly on a dense Cartesian grid of input parameters at extremely high computational efforts. In the current paper, we apply an alternative methodology which demands only a minimum number of model evaluations to construct the response surface. This approach is based on the theory of polynomial chaos expansion (PCE) introduced in [31]. Generally, all PCE techniques can be viewed as an efficient approximation to full-blown stochastic modeling (e.g., exhaustive MC). The basic idea is to represent the response of a model to changes in variables through a response surface that is defined with the help of an orthonormal polynomial basis in the parameter space. In simple words, the dependence of model output on all relevant input parameters is approximated by a high-dimensional polynomial. The resulting polynomials are functions of the model parameters. This projection can be interpreted as an advanced approach to statistical regression.

The PCE offers an efficient and accurate high-order way of including non-linear effects in stochastic analysis (e.g., [39, 40, 41]). One of the attractive features of PCE is the higher-order in uncertainty quantification, e.g., [32, 33, 42], as well as its computational speed when compared to other methods for uncertainty quantification performed on the full model, such as MC [43]. Due to its elegant reduction of models to polynomials, it allows performing many tasks analytically on the expansion coefficients. Alternatively, it allows performing excessive MC on the polynomials since they are vastly faster to evaluate than the original model.

Unfortunately, the original PCE concept [31] is optimal only for Gaussian distributed input parameters. To accommodate for a wide range of data distributions, a recent generalization of PCE is the arbitrary polynomial chaos (aPC [26]). Compared to earlier PCE techniques, the aPC adapts to arbitrary probability distribution shapes of input parameters and, in addition, can even work with unknown distribution shapes when only a few statistical moments can be inferred from limited data or from expert elicitation. The arbitrary distributions for the framework can be either discrete, continuous, or discretized continuous. They can be specified either analytically (as probability density/cumulative distribution functions), numerically as histogram or as raw data sets. This goes beyond the generalization of PCE in methods such as the generalized polynomial chaos (gPC) or the multi-element gPC (ME-gPC) [34, 35]. The aPC approach provides improved convergence in comparison to classical PCE techniques, when applied

to input distributions that fall outside the range of classical PCE. A more specific discussion and review of involved techniques will follow in Sections 2.1 to 2.3.

With an introduction to response methods via the aPC, we describe here the theoretical background that we use in our modeling procedure. The related techniques for sensitivity and risk analysis used in this work are explained in Sections 4 and 5.

2.1. Definitions and polynomial chaos expansion

Suppose that we approximate a problem by a functional Υ, which represents the model responses Γ for the input variables Θ:

$$\Gamma \approx \Upsilon(\Theta). \tag{1}$$

Like all PCE methods, the aPC is a stochastic approach to approximate the response surface. Considering the uncertainty in the input variables, the aPC constructs a set of polynomial basis function and expands the solution in this basis. Thus, the response vector Γ in Eq. (1) can be approximated by [23]:

$$\Gamma \approx \sum_{i=1}^{n_c} c_i \Pi_i(\Theta). \tag{2}$$

Here, n_c is the number of expansion terms, c_i are the expansion coefficients, and Π_i are the multi-dimensional polynomials for the variables $\Theta = [\theta_1, ..., \theta_n]$, and n is the considered number of modeling parameters. If the model response $\Gamma(\Theta)$ depends on space and time, then so do the expansion coefficients c_i.

The number n_c of unknown coefficients c_i results from the number of possible polynomials with total degree equal to or less than d. This number depends on the degree d of the approximating polynomial, and the number of considered parameters n:

$$n_c = \frac{(d+n)!}{d!n!}. \tag{3}$$

2.2. Data-driven orthonormal basis

All polynomials Π_i in expansion (2) are orthogonal, i.e., they fulfill the following condition:

$$\int_{I\in\Omega} \Pi_l \Pi_m p(\Theta) d(\Theta) = \delta_{lm}, \tag{4}$$

where I is the support of Ω, δ is the Kronecker symbol, and $p(\Theta)$ is the probability density function for the input parameters. We obtain the orthonormal basis

6

with the moments-based method proposed in [23, 26]. Orthonormality has the advantage that many subsequent analysis steps are accessible to relatively simple analytical solutions.

Knowledge on variability never is so perfect such that we could express the probability of model parameter values in a unique distribution function. Available data are mostly scarce, and fitting a density function to observed frequencies is often biased by subjective choices of the modeler. Oladyshkin et al. [26] argued that, with aPC, it is possible to use available probabilistic information with no additional formal knowledge requirements for their probability distributions, only based on the statistical moments of the available data. They showed that, it is possible to calculate estimates for the mean, variance, and higher order moments of the model response $\Gamma(\Theta)$ even with incomplete information on the uncertainty of input data, provided in the form of only a few statistical moments up to some finite order.

2.3. Non-intrusive determination of the coefficients

The next task is to compute the coefficients c_i in Eq. 2. Generally, all PCE techniques can be sub-divided into intrusive [44, 45, 46] and non-intrusive [43, 47, 48, 49] approaches, i.e., methods that require or do not require modifications in the system of governing equations and corresponding changes in simulation codes. The challenge in choosing between the methods is to find a compromise between computational effort for model evaluations and a reasonable approximation of the physical processes by the interpolation.

For our study, we prefer the probabilistic collocation method (PCM: see [26, 43, 49]) from the group of non-intrusive approaches like sparse quadrature [50, 51, 52, 53]. In a simple sense, PCM can be interpreted as a smart (mathematically optimal) interpolation and extrapolation rule of model output between and beyond different input parameter sets. It is based on a minimal and optimally chosen set of model evaluations, each with a defined set of model parameters (called collocation points). For this reason, the collocation approach became more popular in the last years. Also, the collocation formulation does not require any knowledge of the initial model structure. It only requires knowledge on how to obtain the model output for a given set of input parameters, which allows treating the model like a "black-box". The distinctive feature of non-intrusive approaches is that any simulation model can be considered a "black-box", i.e. commercial software can be used without any modifications required.

According to [54], the optimal choice of collocation points corresponds to the roots of the polynomial of one degree higher $(d + 1)$ than the order used in the

chaos expansion (*d*). This choice adapts the position of collocation points to the involved distribution shape, and is based on the theory of Gaussian integration (e.g., [55]). For one-dimensional problems (i.e., when analyzing only one uncertain model parameter), it allows exact numerical integrations of order $2d$ given $d + 1$ values of the function to be integrated.

For multi-parameter analysis, the number of available points from the corresponding Gaussian integration rule is $(d + 1)^n$, which is larger than the necessary number *M* of collocation points. The minimum value of *M* is equal to the number of coefficients n_c in Expansion (2), according to Eq. (3). The full tensor grid can be used only for low-order (1st, 2nd) analysis of few parameters. For higher-order analysis of many parameters, the tensor grid suffers from the curse of dimensionality (a full tensor grid in *n* dimensions requires $(d + 1)^n$ points, which rises exponentially in *n*)[56]. In that case, a smart choice of a sparse subset from the tensor grid becomes necessary. Then, PCM chooses the minimum required number of collocation points, equal to the number of coefficients n_c, from the full tensor grid according to their probability weight, i.e. according to their importance as specified by the available probability distribution of Θ. This simply means to select the collocation points from the most probable regions of the input parameter distribution (see [43]).

The weighted-residual method in the random space is defined as [49]:

$$\int (\Gamma - \sum_{i=1}^{n_c} c_i \Pi_i(\Theta)) w(\Theta) p(\Theta) d\tau = 0, \tag{5}$$

where $w(\Theta)$ is the weighting function and $p(\Theta)$ is the joint probability density function of Θ. Please note that choosing $w_i = \Pi_i$ in Eq. 5 results in the method discussed by [32] and [33]. In PCM, the weighting function is chosen as the delta function:

$$w(\Theta) = \delta(\Theta - \Theta_c). \tag{6}$$

Θ_c is the set of collocation points. Substituting from Eq. (6) into Eq. (5) gives the following:

$$\Gamma_c - \sum_{i=1}^{n_c} c_i \Pi(\Theta_c) = 0, \tag{7}$$

where Γ_c are the response values corresponding to the collocation values Θ_c. We solve Eq. (7) to find the coefficients c_i.

Hence, in total, n_c detailed runs are required to determine the n_c unknown coefficients. The roots of the data-driven polynomial basis (see Section 2.2) define

8

the positions of the collocation points specific to the distribution of input parameters at hand and, thus, indicate the optimal parameter sets for model evaluation, using all available information about the input parameters. In our study, we have $n = 4$ uncertain parameters and we use a polynomial of degree $d = 2$. This means that only $n_c = 15$ detailed runs are necessary to obtain the expansion coefficients and approximate the response surface.

3. CO_2 storage problem

Here, we describe the injection scenario for which we analyze sensitivities, uncertainties, and risks in Sections 4 and 5. The same flow responses are studied here as in [21, 22]. These are aquifer pressure, CO_2 mobile and residual volumes and leakage risk as described below. Then, we describe the uncertain parameters considered in the study followed by a discussion on the uncertain structural aspects of the considered geological settings.

3.1. Modeling scenario

A typical scenario of CO_2 injection is defined in which a volume of 40×10^6 m^3 is injected via one well during an injection period of 30 years. This volume corresponds to 20% of the total aquifer pore volume. After stopping injection, simulation continues for 70 years to study the early migration of the CO_2 plume. For brevity, we omit the detailed model equations here and refer the interested reader to [23, 26].

In our scenario, we feature an aquifer system that is formed by shallow-marine deposits. There is one closed boundary on the top side of the model and the other sides are assumed to be open (Figure 1). All the open boundaries are modeled as Dirichlet boundaries, two of which with hydrostatic pressure distribution (the right and bottom boundaries in Figure 1). The remaining left boundary is also hydrostatic, but modified in order to account for the regional groundwater effect (see below).

The cells on the faces of the open boundaries are equipped with a very large pore volume multiplier, such that they numerically represent a much larger volume and effectively enlarge the domain. This helps to minimize the boundary effects of a computational domain that would otherwise be relatively small compared to the injected CO_2 volume (about 20% of the total pore volume, see above). The pore volume multiplier technique allows for a physically reasonable pressure build-up close to the boundary. Moreover, this allows the CO_2 that has left the domain to re-enter by gravity segregation after the injection has stopped.

A summary of the used parameter values is given in Table 1. The hydrological parameters like permeability and porosity vary within individual realizations due to the considered geological structure (see Figure 2 for the histograms of porosity and permeability in one selected realization). They also differ between the different realizations, as they are changed to represent different geological features. Although the geological realizations of this model vary in some geological features, but the same total pore volume, grid, and fault geometry is considered. The injection well is screened in the lower part of the model.

3.2. Analyzed model predictions

We seek to maximize the CO_2 storage volume and minimize the risk of leakage. These quantities are measured by various simulation outputs that are described in Table 2 and discussed in the following.

Aquifer pressure is considered as the spatial average of the pressure distribution in the domain, weighted by the pore volume of each rock type. Monitoring or predicting the pressure response is important to avoid over-pressurized injection operations.

Residual CO_2 volume is the volume of trapped CO_2 that is left in the small pores in an imbibition process. This volume is crucial for the long-term storage capacity of reservoirs.

Mobile CO_2 volume is the volume of CO_2 that can move in a continuous phase in the medium. It is considered as one of the important flow responses, because only mobile CO_2 volumes can lead to leakage through any failure in the sealing cap-rock or ill-plugged well.

Finally, we consider **leakage risk** through cap-rock failure. Cap-rock integrity is a major concern for the safety of CO_2 storage operations. An over-pressurized injection can lead to fractures that may extend up to the cap-rock, penetrate through the cap-rock, or activate pre-existing faults and fractures, and finally lead to CO_2 leakage. In addition, the capillary barrier effect of the cap-rock can be overcome by a local pressure build-up. Thus, the probability of cap-rock failure can depend on the geomechanical properties of the cap-rock and of the medium, on the topography of the cap-rock, and on the pressure build-up resulting from the CO_2 injection and migration. More details about failure mechanisms and failure criteria can be found in the literature (e.g., [57, 58, 59]). However, geomechanical modeling and knowledge about pre-existing features that can be activated during injection would be required to take these processes into account.

Here, we demonstrate how cap-rock integrity can be considered in the workflow of sensitivity analysis and uncertainty assessment in a simplified manner.

To avoid detailed studies of multiphase flow coupled with geomechanical simulations and fracture mechanics, we follow a pragmatic approach. The idea is to assign a spatial probability distribution of cap-rock failure over the area of the cap-rock layer, such that each point of the cap-rock has its own failure probability. In principle, this probability could be assigned in correspondence with the current pressure distribution and with geological features such as varying cap-rock thickness, material properties, faults and fractures. For the means of demonstration, we simply assign a spatial Gaussian function as a scenario assumption to provide the cap-rock failure probability for each point of the cap-rock (see Figure 3). Leakage risk is defined as the probability of leakage (due to cap-rock failure) times the amount of escaping CO_2 in case of leakage. Thus, we spatially integrate the product between cap-rock failure probability and the volume of mobile CO_2 below each point of the cap-rock over the entire area of the cap-rock.

3.3. Uncertain parameters

The most apparent uncertainty in CO_2 storage is the lack of geological knowledge. Large geological scales and diversity of rock properties make it impossible to obtain the whole descriptive picture for a study. A geological study will therefore be accompanied by huge levels of uncertainty. Many studies have shown the significance of geological heterogeneity on underground flow performance (e.g., [60, 61]). To obtain a descriptive image of a feature, like faults and depositional structure, such that uncertainty can be reduced, we must provide adequate data. The process of data collection from underground layers is very costly, therefore it is important to know the ranking of influence each feature has on the flow in order to optimize the cost of data acquisition in modeling.

From the geological parameters that are relevant for shallow-marine deposits used in [21, 22], we pick three parameters: the degree to which barriers may block horizontal and vertical flow, aggradation angle, and fault transmissibility. In addition to these, we consider the regional groundwater effect as an uncertain parameter in our study. Here, we give a brief description on each one, followed by the probabilities assigned to these parameters.

Barriers: During the formation of shallow-marine deposits, periodic floods result in a sheet of sandstone that dips, thins, and fines in a seaward direction. In the lower front, thin sheets of sandstone are inter-bedded with the mud-stones deposited from suspension. These mud-draped surfaces are potential significant barriers to both horizontal and vertical flow. In the SAIGUP realizations, these barriers were modeled by transmissibility multipliers in specific layers of the for-

mation. The position of the barriers is generated by creating an elliptic cone-shaped surface that follows the plan-view shoreline shape of the facies, characterized from real world data [18]. We define the degree of barrier presence by the areal percentage of zero-valued transmissibility multipliers. Figure 5 shows a medium level of barriers.

Aggradation angle: in shallow-marine systems, two main factors control the shape of the transition zone between river and basin: the amount of deposition supplied by the river and the accommodation space that the sea provides for these depositional masses. Deposition happens in a spectrum from larger grains depositing earlier on the land side, to fine deposits happening in the deep basin. If the river flux or sea level fluctuates, equilibrium changes into a new bedding shape based on the balance of these factors. In the SAIGUP study, progradational cases are considered, in which river flux increases and shifts the whole depositional system into the sea. The angle at which transitional deposits are stacked on each other because of this shifting is called the aggradation angle. Three levels of aggradation are shown in Figure 4: low, medium and high. The study reported in [21, 22] showed that aggradation can have a dramatic influence on the injection and migration process.

Fault transmissibility: Huge uncertainties can be involved when modeling the presence of faults. Faults are discrete objects that are modeled by changing the geometry of the simulation grid. The transmissibility for flow across faults changes during the process of faulting. This causes a spectrum of transmissibilities, from a sealing fault with no flow across it, to a fault that has not produced any barriers to the flow within its opening space.

Within a simulation grid, the influence of faults on the local and global flow behavior depends on a number of parameters including fault length, orientation, intensity and transmissibility. The well location with respect to the faults can change the overall behavior of injected CO_2 plume significantly. In the SAIGUP models, different levels of fault orientations, transmissibility, areal intensity, and well patterns are considered. For this study, we consider all fault modeling parameters at their medium level and consider to vary only the fault transmissibility. These variations, however, do not affect the definition of the no-flow boundary, which is motivated by the presence of an impermeable fault.

The used geology realizations contain compartmentalized fault systems comprising approximately equal densities of strike-parallel and strike-perpendicular faults based on a portion of the Gullfaks field [18, 62]. Figure 6 shows the fault pattern and location of the injector considered for the study.

It is shown in [63] that the transmissibility multiplier provides a numerically

more robust representation of faults within reservoir simulation than conventional permeability multipliers. We consider the fault transmissibility multipliers to range between zero and one. A multiplier value of one corresponds to a fault permeability equal to the harmonic average of cell permeabilities across the fault, i.e., to a fault without any influence on flow [63].

Regional groundwater effect: Geological modeling always comes with the uncertainty of how large the aquifer is and how it is connected to other underground aquifers. This is a direct consequence of the need to define boundary conditions to limit the computational domain, which cannot always coincide with meaningful physical boundaries in large-scale systems. However, connections to active external aquifers can be accounted for by adapting the values for the boundary conditions accordingly. Some connections might even change throughout the year, depending on rainfall. The flux across model boundaries might influence the CO_2 plume dynamics during and after injection. To simulate such effects, we changed the left boundary pressure by adding an uncertain additional pressure value Δp that varies between 0 and 100 bars.

As a scenario assumption, this pressure value is added at the start of injection, i.e., the pressure distribution is not at a steady state when the simulation starts, and this triggers a corresponding transient brine flow. We do so in order to analyze the effect of transient groundwater effects on the system. This may seem an arbitrary choice, but assuming a steady-state would also be arbitrary to some extent.

The overall process for sensitivity analysis, uncertainty propagation, and risk assessment starts by specifying probability information for the uncertain parameters. Next, one has to design and choose the simulation cases required to obtain the expansion coefficients in the approximating polynomial. However, in our study, we had access to the set of SAIGUP geological realizations and simulation results that had been designed without the considerations possible with the aPC. The computing time for each SAIGUP realization was about 2 hours on a 2.4GHz Intel Xeon CPU, and we decided to recycle these highly expensive simulations in our study. The large computing times are a key motivation to build a cheaper surrogate model for further analysis. Hence, we assume the histograms of uncertain parameters such that they result in collocation points that coincide with the SAIGUP designed values. Therefore, the histograms used in this study are almost uniform, as shown in Figure 7. In fact, these input distributions could also be handled with the gPC method already mentioned in the introduction, and would correspond to the use of Legendre polynomials. In our case, we use the aPC to avoid the step of modeling the input distributions as exactly uniform. Consequently, the polynomials resulting from the aPC approach are very close to Legendre polynomials.

The aPC, however, could be used for any type of histograms and so provides the freedom in other studies to adapt to arbitrary input statistics.

The main concern here is not a unique probability description of the input geological parameters, but rather we perform an uncertainty analysis practice, relying on a scenario assumption of probability distributions. Thus, no general geological conclusion is expected from this study, and results might change by feeding the work-flow with a different probability description.

4. Sensitivity analysis

In this section, we tackle global sensitivity analysis with Sobol indices based on the aPC technique, following the line of work on aPC by [23, 26, 37]. The big advantage of global aPC-based sensitivity analysis is that one can obtain global sensitivity information at computational costs that are hardly larger than those for local analysis. The reason is the following: local methods use infinitesimally small spacing between parameter sets for model evaluation to get numerical derivatives evaluated at a single point. The aPC-based method places the parameter sets for model evaluation at an optimized spacing in parameter space. This can be interpreted as fitting secants (or polynomials for non-linear analysis) to the model response. These secants (polynomials) approximate the model over the entire parameter space in a weighted least-square sense (compare with the best unbiased ensemble linearization approach described by [64]). This is more beneficial compared to computing a tangent or local second derivatives (compare FORM, SORM methods, e.g., [65]) that approximate the model well just around one point in the parameter space.

The system featured here is non-linear due to two reasons: First, the involved multi-phase flow equations [26] form a coupled system of non-linear partial differential equations, and second, these equations are non-linear in their coefficients. The latter is even more significant if parameters are spatially heterogeneous.

In the following, we briefly summarize the Sobol sensitivity indices technique for quantifying the relative importance of each individual input parameter in the final prediction. Then, we implement the method for our geological CO_2 storage problem, based on the aPC response surface.

The model responses featured here for global sensitivity analysis (this section) and for the probabilistic risk analysis (see Section 5) are listed in Table 2 and have been discussed in Section 3.1. In the sense of global sensitivity analysis [66], not only should the analysis technique be global, but also should the analyzed quantities be global. In the latter, global refers to the fact that they are relevant for the

14

engineer, are crucial in decision processes, etc. For example, an overall leakage risk is more informative in final decisions than the leakage rate at a specific point, and a total stored volume of CO_2 is more informative for volumetric efficiency considerations of the reservoir than the CO_2 saturation at individual points.

4.1. Sobol sensitivity indices

The method is well described in the literature [24, 36, 66, 67]. More recent works are concerned about expediting calculation pace by computing Sobol indices analytically from polynomial chaos expansions [26, 33, 37, 68, 69]. The idea behind the combination of PCE techniques with Sobol indices is to replace the analyzed system with an approximating function which leads to mathematical and numerical benefits in the sensitivity analysis.

Using polynomials for this approximation is convenient, because it is easy to analytically obtain the output variances from the statistics of the input variables of the polynomials. In our case, the solution is approximated by orthogonal polynomials with ascending polynomial degree. We expand the variance of model output into individual components originating from all possible combinations of input parameters. Assume that we break the system output into components as follows:

$$\Gamma = \Gamma_0 + \sum_i \Gamma_i + \sum_i \sum_{j>i} \Gamma_{ij} + ... \tag{8}$$

A single index (here: i) shows dependency to a specific input variable. More than one index (e.g.: i and j) shows interaction of two or more input variables. If we consider the input vector Θ to have n components θ_i for $i = 1,..,n$, then $\Gamma_i = f_i(\theta_i)$ and $\Gamma_{ij} = f_{ij}(\theta_i, \theta_j)$. In practice, we stop at a finite number of terms in Eq. (8). The first order sensitivity index, the so called Sobol index, is defined statistically as follows [66]:

$$S_i = \frac{V[E(\Gamma \mid \theta_i)]}{V(\Gamma)}, \tag{9}$$

where $E(\Gamma \mid \theta_i)$ is the conditional expectation of output Γ for a given value of θ_i and V is the variance operator. In plain words, S_i is the fraction of total variance $V(\Gamma)$ that can be explained by the parameter θ_i. Since θ_i can be fixed at any value in its uncertainty interval, each of those values produces a distinct expectation. In Eq. (9), the variance of those expectations is divided by the unconditional variance

of output (i.e., with no input variable fixed). For more than one index, a higher-order Sobol index can be defined as:

$$S_{ij} = \frac{\mathrm{V}[\mathrm{E}(\Gamma \mid \theta_i, \theta_j)] - \mathrm{V}[\mathrm{E}(\Gamma \mid \theta_i)] - \mathrm{V}[\mathrm{E}(\Gamma \mid \theta_j)]}{\mathrm{V}(\Gamma)}. \tag{10}$$

Here, $\mathrm{V}[\mathrm{E}(\Gamma \mid \theta_i, \theta_j)]$ is the variance of output expectations after fixing θ_i and θ_j. This index represents the significance of variation in output generated from the joint uncertainty in several input variables, i.e., from the interaction of uncertain parameters. If we add all indices that contain a given variable θ_i, the sum is called the total Sobol index:

$$S_{Ti} = S_i + \sum_{j \neq i} S_{ij} + \sum_{j \neq i} \sum_{k \neq i} S_{ijk} + \dots \tag{11}$$

The total Sobol index is a sensitivity measure to rank parameters according to their influence on the model results. When this index is close to zero, the corresponding parameter has a negligible role in the variation of the system response. In that case, the uncertainty in that parameter does not introduce a considerable uncertainty in the response, and the parameter could be omitted from further analyses.

In practice, we evaluate the Sobol indices analytically from the expansion coefficients of the aPC as described by [37].

4.2. Sensitivity analysis

We calculate the total Sobol indices for the geological CO_2 storage problem that is described earlier. The results are based on an aPC expansion of order two that is obtained by fifteen detailed simulations. The choice of order two is supported by the results of [43], where the authors found in a similar CO_2 storage problem that second order may be the cheapest non-linear expansion, but still sufficiently accurate for this type of purpose. Recently, [37] provided the results of a numerical convergence analysis for aPC-based Sobol analysis. They report that increasing the expansion order beyond 2 introduces only small changes to the sensitivity values for their considered system, and does not change the ranking of the analyzed parameters anymore. A study similar to the current study without aPC needed one hundred and sixty runs to perform a sensitivity analysis with a different method [22]. The pattern of sensitivity reported here is similar to what is produced in that study, but at dramatically reduced costs.

4.3. Results

The flow behavior in the domain is influenced by the type and intensity of different heterogeneities. This influence can be traced in the CO_2 pressure and saturation distributions over time. During injection, viscous forces imposed by the injector dominate the force balance. Viscous forces act in the form of spatial pressure gradients in all directions. After 30 years, the injection stops, and gravity starts playing the major role in the flow dynamics, acting in the vertical direction [21, 22].

Barriers and aggradation angle have different impacts on the flow during each flow regime, i.e., injection or after injection. Low fault transmissibility hinders the flow and keeps the pressure in compartments. Geometry distortion in the geological layers because of the faulting processes plays a considerable role in the splitting of CO_2 plumes within the domain. Water flux from lateral boundaries due to the regional groundwater effect enhances the spread of CO_2 and leads the mass of CO_2 toward the other open boundaries.

Figure 8 shows the sensitivity of different responses to the uncertain parameters. Total Sobol indices are plotted at specific times. End of simulation refers to the year 100, i.e., 70 years after injection stops. This time duration is long enough for the flow to stabilize at a stationary condition for the majority of the model runs.

As already observed in [21, 22], the aggradation angle plays a significant role in the flow behavior. In cases with low aggradation angle, the stratigraphy of rock types is a pattern of parallel layering. For higher aggradation angles, rock-types are distributed between more modeling layers. The effective vertical permeability changes from the harmonic average (in Figure 9a) toward the arithmetic average (in Figure 9c), as the aggradation angle increases from 0 to 90 degrees. The harmonic average might be much smaller than the arithmetic average, in particular when there are vertically impermeable rock-types in the medium. The shallow marine depositional system contains some rock-types with almost zero transmissibility in the vertical direction. Therefore, a low aggradation angle can hinder the flow from traveling upward across layers in the domain and force it to stay trapped in some lower layers, as seen for many of the low aggradation angle realizations in our study. The relatively large sensitivities to the level of barrier presence is based on the same effects.

Our results show a relatively weak sensitivity of responses with respect to the water influx from one side of the model. This sensitivity is in particular low during injection, when the high pressure imposed from the well dominates the dynamics of flow in the medium (Figure 8a). The sensitivity patterns for the mobile and

17

residual CO_2 volume are similar in Figures 8b and 8c, because the mobile and residual CO_2 volume add up to the total injected CO_2 volume, with the exception of the CO_2 volume that has left the domain. Hence, they are highly dependent on each other.

More detailed results are shown in Figures 10a to 10d. Total Sobol indices are plotted for each response during the entire time interval. When the flow regime switches from injection to a gravity-dominated system, we observe a jump or sharp drop in some of the sensitivity plots (Figures 10a and 10c at 30 years).

The sensitivity of the aquifer pressure with respect to the presence of barriers jumps up, right after stopping the injection. This happens because barriers slow down the pressure release through open boundaries, resulting in local pressure build-ups.

The sensitivity of the residual CO_2 volume with respect to barriers presence drops soon after injection. This is reasonable since the residual trapped volume is more a function of lateral flow in the medium, compared to the vertical flow in the relatively small thickness of the aquifer.

5. Risk analysis

The risk R of a process is quantitatively defined as the extent of consequence C caused by the process, multiplied by the probability P of that consequence to happen:

$$R = P \times C. \tag{12}$$

The consequence can be defined by direct measures in the simulation responses, or it can be related to consequences caused in the environment outside the considered system. For example, in the case of CO_2 injection into deep aquifers, the amount of CO_2 which stays mobile and undissolved in the medium for a time after injection can be considered as a consequence, bearing the potential of leakage up to the surface if exposed to a geological leakage point. The consequence could also be defined by a criterion for external consequences, like the rate of climate change (either locally or globally) due to CO_2 leakage, the costs of pumping CO_2 that does not remain in the subsurface, or via the related costs for CO_2 emission certificates.

The other part is the probability of these consequences to happen. This depends on the stochastic behavior of the process which results in the respective outcomes.

We use the polynomial-based reduced model for risk analysis, because it is fast enough to perform a Monte-Carlo analysis with a large number (here: 10000)

of realizations on the polynomials. Thanks to the higher-order approximation via the aPC, the principal non-linear physical behavior of CO_2 storage is included in the analysis, and detailed probabilistic risk assessment becomes feasible. We analyze here the same quantities as in Section 4, i.e., average pressure, the volume of mobile or immobile CO_2, and leakage risk. For definitions, see Section 3.1.

5.1. Quantification of expected values in CO$_2$ storage

Average response values can be calculated analytically from the polynomial (e.g., [26]) or via the Monte-Carlo post-process as mentioned above. Figures 11a to 11d show some of the calculated expectations as functions of time. In Figure 11a, the mobile CO_2 volume increases linearly in the medium because of the constant injection rate during the first thirty years. After injection, the mobile volume of CO_2 is reduced due to the trapped volume in residual form and the migration of CO_2 across open boundaries.

Figure 11b shows the expected values for the volume of residually trapped CO_2 as a function of time. The plot shows the significance of imbibition during the plume migration period, when water replaces CO_2 that is moving upward because of gravity segregation. During injection, CO_2 invades the aquifer and drainage is dominant. Therefore, the expected residual CO_2 plot shows a smaller slope during injection than what it shows later in time.

When injection starts, a pressure pulse travels through the medium at a finite velocity because of the slight compressibility of brine. The initially built-up pressure releases through open boundaries over time and the average pressure drops in the aquifer (Figure 11c). Under realistic injection settings, an average (spatial and statistical) pressure rise of up to 400 bars (from 270 to 670 bars in the first simulation time step, not visible in Figure 11c) would be very unrealistic and would not be allowed to occur. This high pressure rise occurs because large pressure values have to be exceeded in the injection cell before CO_2 becomes mobile at saturations above the residual value.

Also, during early injection time, the pressure is larger than at the end of injection. There are a few realizations where the contributions from the external aquifer support, a dense barrier system close to 100% areal coverage, an adverse aggradation angle of the formation and extremely low fault transmissibilities interact to effectively block the CO_2 flow close to the well. This has strong effects on pressure when the rock at the injector position happens to be poorly permeable, leading to a very poor injectivity. An adapted CO_2 injection strategy would react by lowering the injection rate, by choosing a different injection position, or by even abandoning the entire site.

Based on the results of the current study, it is possible to identify such adverse combinations and guide site investigation strategies to pay attention to such situations. In a follow-up study (ready for submission), we are currently investigating an active injection strategy controlled by an upper allowable pressure limit.

However, the initial sharp pressure increase is released very quickly. This happens, when first parts of the CO_2 plume have found flow pathways into regions with better rock properties, providing the possibility to relax the pressure build-up, and also to let the CO_2 escape towards the boundaries.

The expected leakage risk is plotted in Figure 11d, and increases in value as the injected CO_2 travels upward and accumulates beneath the sealing cap-rock.

5.2. Results of CO_2 storage risk assessment

In this section, the probability distributions (rather than expected values) of system responses during and after injection are studied. Results from the MC analysis of the response surface are given as histograms of output values and also as cumulative distribution functions (CDF) for probabilities (Figures 12 and 13).

Figures 12a to 12c show the histograms of responses obtained from the Monte-Carlo process at the end of injection. A long tail is observed for lower residual and mobile CO_2 values in Figures 12b and 12c. The long tail means a large range of possible low values. Pressure shows a long tail for higher values. This means that even high critical values still have substantial probabilities to be exceeded, indicating that the possibility of geomechanical damage to sealing layers will have to receive a large attention. We observe an issue of mass conservation in Figure 12b, where a few realizations show more mobile CO_2 in the domain than the total injected volume (which is about 40×10^6 m^2). This is a typical issue for a large class of statistical methods that interpolate or extrapolate simulation results in the parameter space, because their setup is not based on the mass conservation equation. In this specific case, the mass conservation issue is caused by approximating the response surface via polynomials, with vanishing residuals only at the collocation points. The polynomials are evaluated at many randomly chosen parameter sets drawn from the histograms shown in Figure 7, which do not coincide with the collocation points.

Finally, we report how the corresponding probabilities change over time in Figures 13a to 13c. High pressure buildup is considerable during the early injection time, and it is negligible after injection during plume migration (Figure 13a). An over-pressurized injection can induce fracturing in the medium, extending to the sealing layers. Any fractures caused in the structural traps can expose the mobile CO_2 to leakage paths. Therefore, higher pressure values can be interpreted

20

as high risk in early time. The injection scenario in this study is set to a fixed injection rate. To avoid risky pressures during the start of injection, the injector can be set on pressure control instead, by changing the injection rate in order to keep the injection pressure limited.

6. Conclusion

In this paper, we used the arbitrary polynomial chaos expansion (aPC) method in a sensitivity analysis and risk assessment process. The goal was to demonstrate the application and feasibility of aPC-based methods in the context of realistic CO_2 injection scenarios. We implemented this method for a typical CO_2 storage problem. Four uncertain parameters with assumed uncertainty distributions are considered. Injection and early migration of CO_2 is studied. The flow sensitivity to geological heterogeneity is evaluated and quantified using Sobol indices. Risk analysis is performed on the defined problem. Flow dynamics are discussed and corresponding interpretations and explanations of the sensitivity and risk results are provided.

The performance of the aPC method has been satisfactory. It is very fast, compared to other stochastic methods for low-parametric systems, and this speed-up allows us to perform an extensive Monte-Carlo process on the aPC-based response surface to calculate the probability of response values throughout simulation time. This study was a first-time application of the aPC to study a realistically complex type of geological structural uncertainty. Based on our assessment of aPC feasibility, we can strongly encourage the use of aPC for sensitivity and risk analysis in complex situations.

The results have shown that the most influential parameter for most of the responses is the aggradation angle of deposition layers of the considered shallow-marine aquifer. The least relevant parameter is the regional groundwater effect, especially during injection time. We re-iterate that the aim of this study was to demonstrate a practice of using arbitrary polynomial chaos expansion for the sensitivity and risk analysis of a typical CO_2 storage problem. Since, in general, the levels of involved input uncertainty are not unique, the physical and geological conclusions of this study are restricted to the probability assumptions taken here and should not be generalized to systems that are very different.

7. Acknowledgements

References

[1] M. Lenzen, Global warming effect of leakage from CO_2 storage, Critical Reviews in Environmental Science and Technology 41 (24) (2011) 2169–2185.

[2] P. Viebahn, J. Nitsch, M. Fischedick, A. Esken, D. Schüwer, N. Supersberger, U. Zuberbühler, O. Edenhofer, Comparison of carbon capture and storage with renewable energy technologies regarding structural, economic, and ecological aspects in Germany, International Journal of Greenhouse Gas Control 1 (1) (2007) 121–133.

[3] S. Bachu, CO_2 storage in geological media: Role, means, status and barriers to deployment, Progress in Energy and Combustion Science 34 (2) (2008) 254–273.

[4] S. Bachu, Screening and ranking of sedimentary basins for sequestration of CO_2 in geological media in response to climate change, Environmental Geology 44 (3) (2003) 277–289.

[5] J. Birkholzer, Q. Zhou, C. Tsang, Large-scale impact of CO_2 storage in deep saline aquifers: A sensitivity study on pressure response in stratified systems, International Journal of Greenhouse Gas Control 3 (2) (2009) 181–194.

[6] J. Nordbotten, M. Celia, S. Bachu, H. Dahle, Semianalytical solution for CO_2 leakage through an abandoned well, Environmental Science & Technology 39 (2) (2005) 602–611.

[7] D. Thomas, Carbon Dioxide Capture for Storage in Deep Geologic Formations-Results from the CO_2 Capture Project, Vol. 1, Elsevier Science Ltd, 2005.

[8] M. Celia, S. Bachu, J. Nordbotten, S. Gasda, H. Dahle, Quantitative estimation of CO_2 leakage from geological storage: Analytical models, numerical models and data needs, in: Proceedings of 7th International Conference on Greenhouse Gas Control Technologies.(GHGT-7), 2004.

[9] F. Riddiford, I. Wright, C. Bishop, T. Espie, A. Tourqui, Monitoring geological storage the in salah gas CO_2 storage project, in: Proceedings of the 7th International Greenhouse Gas Technologies Conference, Vancouver, BC, 2004.

[10] A. Förster, B. Norden, K. Zinck-Jørgensen, P. Frykman, J. Kulenkampff, E. Spangenberg, J. Erzinger, M. Zimmer, J. Kopp, G. Borm, et al., Baseline characterization of the CO2SINK geological storage site at Ketzin, Germany, Environmental Geosciences 13 (3) (2006) 145–161.

[11] G. Eigestad, H. Dahle, B. Hellevang, F. Riis, W. Johansen, E. Øian, Geological modeling and simulation of CO_2 injection in the Johansen formation, Computational Geosciences 13 (4) (2009) 435–450.

[12] K. Michael, A. Golab, V. Shulakova, J. Ennis-King, G. Allinson, S. Sharma, T. Aiken, Geological storage of CO_2 in saline aquifers–a review of the experience from existing storage operations, International Journal of Greenhouse Gas Control 4 (4) (2010) 659–667.

[13] H. Class, A. Ebigbo, R. Helmig, H. Dahle, J. Nordbotten, M. Celia, P. Audigane, M. Darcis, J. Ennis-King, Y. Fan, et al., A benchmark study on problems related to CO_2 storage in geologic formations, Computational Geosciences 13 (4) (2009) 409–434.

[14] F. Walton, J. Tait, D. LeNeveu, M. Sheppard, Geological storage of CO_2: A statistical approach to assessing performance and risk, in: Proceedings of the 7th International Conference on Greenhouse Gas Control Technologies (GHGT-7), 2004.

[15] S. Brennan, R. Burruss, M. Merrill, P. Freeman, L. Ruppert, A probabilistic assessment methodology for the evaluation of geologic carbon dioxide storage, US Geological Survey Open-File Report 1127 (2010) 31.

[16] E. Wilson, T. Johnson, D. Keith, Regulating the ultimate sink: Managing the risks of geologic CO_2 storage, Environmental Science and Technology 37 (16) (2003) 3476–3483.

[17] A. Hansson, M. Bryngelsson, Expert opinions on carbon dioxide capture and storage–a framing of uncertainties and possibilities, Energy Policy 37 (6) (2009) 2273–2282.

[18] J. Howell, A. Skorstad, A. MacDonald, A. Fordham, S. Flint, B. Fjellvoll, T. Manzocchi, Sedimentological parameterization of shallow-marine reservoirs, Petroleum Geoscience 14 (1) (2008) 17–34.

[19] T. Manzocchi, J. Carter, A. Skorstad, B. Fjellvoll, K. Stephen, J. Howell, J. Matthews, J. Walsh, M. Nepveu, C. Bos, et al., Sensitivity of the impact of geological uncertainty on production from faulted and unfaulted shallow-marine oil reservoirs: objectives and methods, Petroleum Geoscience 14 (1) (2008) 3–11.

[20] J. Matthews, J. Carter, K. Stephen, R. Zimmerman, A. Skorstad, T. Manzocchi, J. Howell, Assessing the effect of geological uncertainty on recovery estimates in shallow-marine reservoirs: the application of reservoir engineering to the SAIGUP project, Petroleum Geoscience 14 (1) (2008) 35–44.

[21] M. Ashraf, K. Lie, H. Nilsen, J. Nordbotten, A. Skorstad, Impact of geological heterogeneity on early-stage CO_2 plume migration, in: CMWR, 2010.

[22] M. Ashraf, K.-A. Lie, H. M. Nilsen, A. Skorstad, Impact of geological heterogeneity on early-stage CO_2 plume migration: sensitivity study., in: Proceedings of ECMOR XII, 2010.

[23] S. Oladyshkin, W. Nowak, Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion, Reliability Engineering & System Safety 106 (2012) 179–190.

[24] I. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation 55 (1-3) (2001) 271–280.

[25] E. Commission, European Commission's Communication on Extended Impact Assessment, Brussels, Office for Official Publications of the European Communities, 2002.

[26] S. Oladyshkin, H. Class, R. Helmig, W. Nowak, A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations, Advances in Water Resources 34 (2011) 1508–1518. doi:10.1016/j.advwatres.2011.08.005.

[27] J. Witteveen, S. Sarkar, H. Bijl, Modeling physical uncertainties in dynamic stall induced fluid–structure interaction of turbine blades using arbitrary polynomial chaos, Computers & Structures 85 (11) (2007) 866–878.

[28] J. Witteveen, H. Bijl, Modeling arbitrary uncertainties using Gram-Schmidt polynomial chaos, in: Proceedings of the 44th AIAA Aerospace Sciences Meeting and Exhibit, number AIAA-2006-0896, Reno, NV, Vol. 117, 2006.

[29] R. Ghanem, A. Doostan, On the construction and analysis of stochastic models: characterization and propagation of the errors associated with limited data, Journal of Computational Physics 217 (1) (2006) 63–81.

[30] C. Soize, R. Ghanem, Physical systems with random uncertainties: chaos representations with arbitrary probability measure, SIAM Journal on Scientific Computing 26 (2) (2004) 395–410.

[31] N. Wiener, The homogeneous chaos, American Journal of Mathematics 60 (1938) 897–936.

[32] R. G. Ghanem, P. D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag, New York, 1991.

[33] O. Le Maître, O. Knio, Spectral methods for uncertainty quantification: with applications to computational fluid dynamics, Springer, 2010.

[34] X. Wan, G. Karniadakis, Multi-element generalized polynomial chaos for arbitrary probability measures, SIAM Journal on Scientific Computing 28 (3) (2007) 901–928.

[35] D. Xiu, G. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2002) 619–644.

[36] U. Reuter, M. Liebscher, Global sensitivity analysis in view of nonlinear structural behavior, in: Proceedings of the 7th LS-Dyna Forum, Bamberg, 2008.

[37] S. Oladyshkin, F. P. J. de Barros, W. Nowak, Global sensitivity analysis: a flexible and efficient framework with an example from stochastic hydrogeology, Advances in Water Resources 37 (2012) 10–22.

[38] A. Kovscek, Y. Wang, Geologic storage of carbon dioxide and enhanced oil recovery. i. uncertainty quantification employing a streamline based proxy for reservoir flow simulation, Energy Conversion and Management 46 (11) (2005) 1920–1940.

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

[40] J. Foo, G. Karniadakis, Multi-element probabilistic collocation method in high dimensions, Journal of Computational Physics 229 (5) (2010) 1536–1557.

[41] N. Fajraoui, F. Ramasomanana, A. Younes, T. A. Mara, P. Ackerer, A. Guadagnini, Use of global sensitivity analysis and polynomial chaos expansion for interpretation of non-reactive transport experiments in laboratory-scale porous media, Water Resources Research (doi:10.1029/2010WR009639).

[42] R. Ghanem, P. D. Spanos, Polynomial chaos in stochastic finite elements, Journal of Applied Mechanics 57 (1990) 197–202.

[43] S. Oladyshkin, H. Class, R. Helmig, W. Nowak, An integrative approach to robust design and probabilistic risk assessment for CO_2 storage in geological formations, Computational Geosciences 15 (3) (2011) 565–577. doi:10.1007/s10596-011-9224-8.

[44] R. Ghanem, P. Spanos, A stochastic Galerkin expansion for nonlinear random vibration analysis, Probabilistic Engineering Mechanics 8 (1993) 255–264.

[45] G. Matthies, H., A. Keese., Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, Computer Methods in Applied Mechanics and Engineering 194 (2005) 1295–1331.

[46] D. Xiu, G. E. Karniadakis, Modeling uncertainty in flow simulations via generalized polynomial chaos, Journal of Computational Physics 187 (2003) 137–167.

[47] A. Keese, H. G. Matthies, Sparse quadrature as an alternative to MC for stochastic finite element techniques, Proc. Appl. Math. Mech. 3 (2003) 493Ű–494.

[48] S. Isukapalli, S., A. Roy, G. Georgopoulos, P., Stochastic response surface methods (srsms) for uncertainty propagation: Application to environmental and biological systems, Risk Analysis 18 (3) (1998) 351–363.

[49] H. Li, D. Zhang, Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods, Water Resources Research 43 (2007) 44–48.

[50] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM Journal on Numerical Analysis 45 (3) (2007) 1005–1034.

[51] D. Xiu, J. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM Journal on Scientific Computing 27 (3) (2005) 1118–1139.

[52] T. Gerstner, M. Griebel, Dimension–adaptive tensor–product quadrature, Computing 71 (1) (2003) 65–87.

[53] V. Barthelmann, E. Novak, K. Ritter, High dimensional polynomial interpolation on sparse grids, Advances in Computational Mathematics 12 (4) (2000) 273–288.

[54] J. Villadsen, M. L. Michelsen, Solution of differential equation models by polynomial approximation, Prentice-Hall, 1978.

[55] M. Abramowitz, A. Stegun, I., Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, New York: Dover, 1965.

[56] F. Nobile, R. Tempone, C. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, SIAM Journal on Numerical Analysis 46 (5) (2008) 2309–2345.

[57] P. Zweigel, L. Heill, Studies on the likelihood for caprock fracturing in the sleipner CO_2 injection case, Tech. rep., Sintef Petroleum Research Report (2003).

[58] E. Aker, E. Skurtveit, L. Grande, F. Cuisiat, Ø. Johnsen, M. Soldal, B. Bohloli, Experimental methods for characterization of cap rock properties for CO_2 storage, Multiphysical Testing of Soils and Shales 303–308.

[59] J. Rohmer, D. Seyedi, Coupled large scale hydromechanical modelling for caprock failure risk assessment of co_2 storage in deep saline aquifers, Oil & Gas Science and Technology–Revue de l'Institut Français du Pétrole 65 (3) (2010) 503–517.

[60] S. Dutton, W. Flanders, M. Barton, Reservoir characterization of a Permian deep-water sandstone, East Ford field, Delaware basin, Texas, AAPG bulletin 87 (4) (2003) 609–628.

[61] T. Eaton, On the importance of geological heterogeneity for flow simulation, Sedimentary Geology 184 (3-4) (2006) 187–201.

[62] T. Manzocchi, J. Matthews, J. Strand, J. Carter, A. Skorstad, J. Howell, K. Stephen, J. Walsh, A study of the structural controls on oil recovery from shallow-marine reservoirs, Petroleum Geoscience 14 (1) (2008) 55–70.

[63] T. Manzocchi, J. Walsh, P. Nell, G. Yielding, Fault transmissibility multipliers for flow simulation models, Petroleum Geoscience 5 (1) (1999) 53–63.

[64] W. Nowak, Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator, Water Resources Research 45 (W04431). doi:10.1029/2008WR007328.

[65] S. Jang, Y., N. Sitar, D. Kiureghian, A., Reliability analysis of contaminant transport in saturated porous media., Water Resources Research 30 (8) (1994) 2435–2448.

[66] A. Saltelli, Global sensitivity analysis: the primer, Wiley, 2007.

[67] A. Saltelli, Global sensitivity analysis: an introduction, in: Proc. 4th International Conference on Sensitivity Analysis of Model Output, pp. 27–43.

[68] T. Crestaux, O. Le MaıĹtre, J. Martinez, Polynomial chaos expansion for sensitivity analysis, Reliability Engineering & System Safety 94 (7) (2009) 1161–1172.

[69] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering & System Safety 93 (7) (2008) 964–979.

Table 1: Aquifer model information.

Parameter	Value	Unit
Number of active cells in the model	78720	-
Resolution X,Y,Z	$40 \times 120 \times 20$	-
Scale X,Y,Z	$3000 \times 9000 \times 80$	m
Injection rate	3650	m^3/day
Initial pressure	266.5	bar
Critical CO_2 and water saturations	0.2	-
CO_2 viscosity	0.04	cp
Water viscosity	0.4	cp
Rock compressibility	0.3e-6	1/bar

Table 2: Important model responses and their brief description. For more information, see [21, 22].

Response	Description
Aquifer pressure	Volume average of pressure, weighted by porosity
Mobile CO_2	Volume of CO_2 in places with saturation above critical value
Residual CO_2	Volume of CO_2 in places with saturation below critical value
Leakage risk	A risk value for the leakage through the cap-rock.
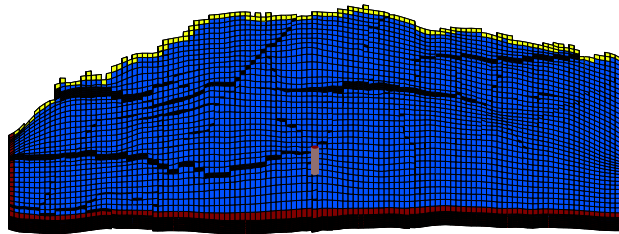
Figure 1: Boundary conditions and the well location in the designed injection scenario. Red color corresponds to the open boundaries and yellow color shows the closed side on the crest.

(a) Porosity, fraction.

(b) Horizontal Permeability, log(mD).
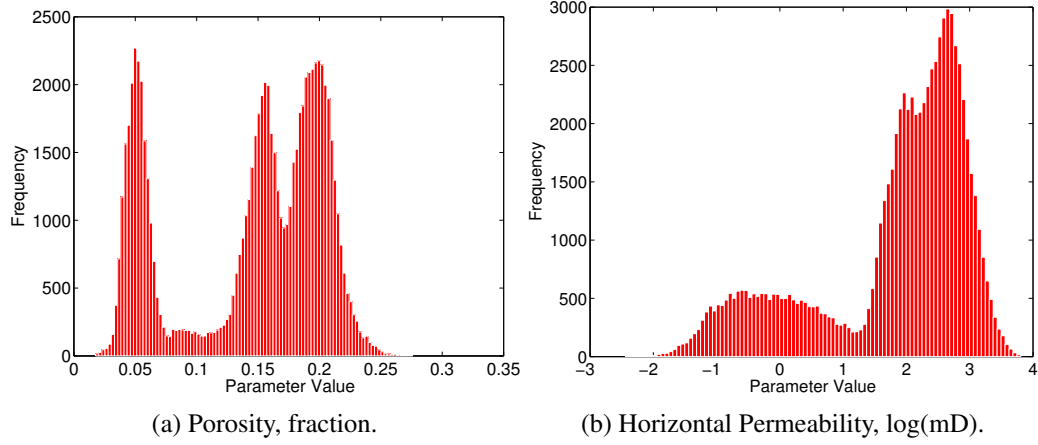
Figure 2: The histograms of hydrological parameters shown for a realization with low levels of heterogeneity. The vertical permeabilities are approximately one order of magnitude lower than the horizontal permeabilities.
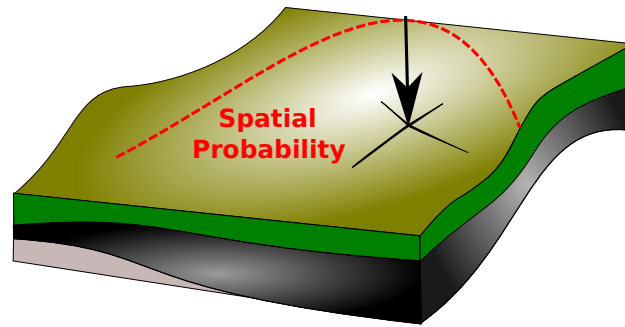
Figure 3: CO_2 leakage risk is computed as the product of a cap-rock failure probability and the amount of mobile CO_2 beneath the cap-rock, integrated over the entire surface area of the cap-rock. Here, we use a Gaussian function as simple scenario assumption for the cap-rock failure probability (indicated schematically by the color shading and the dashed red line with the black coordinate system).
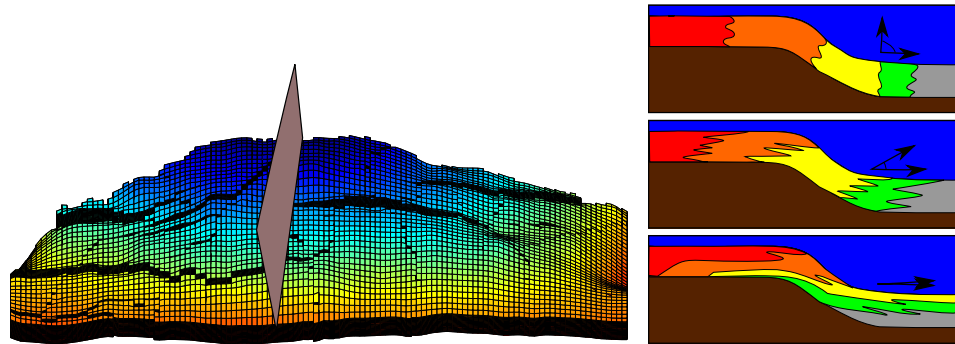
Figure 4: The river flows from left to right toward the sea on the model vertical section shown here (left figure). Aggradation angle is demonstrated in three levels (right figure); from top: low, medium and high aggradation angle. Between deposition and now, the entire system was rotated by tectonic effects such that the original river flow direction is oriented upward, not downward.
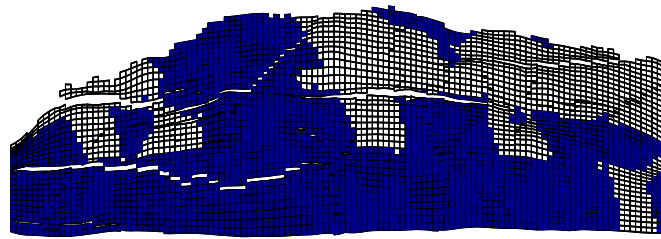
Figure 5: The figure shows 50% of zero transmissibility multipliers in a specific model layer representing a medium level of barriers. One layer of the model is shown in the figure.

| 2000 | 2100 | 2200 | 2300 | 2400 | 2500 |
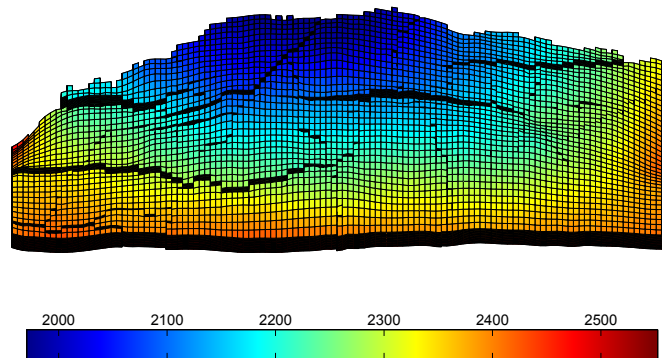
Figure 6: Fault orientation and intensity of the model used in the study. Depth in meter is shown by color on the grid.

31

(a) Barriers, percentage.

(b) Aggradation angle, degree.

(c) Fault transmissibility multiplier,-.

(d) Lateral pressure support, bar.
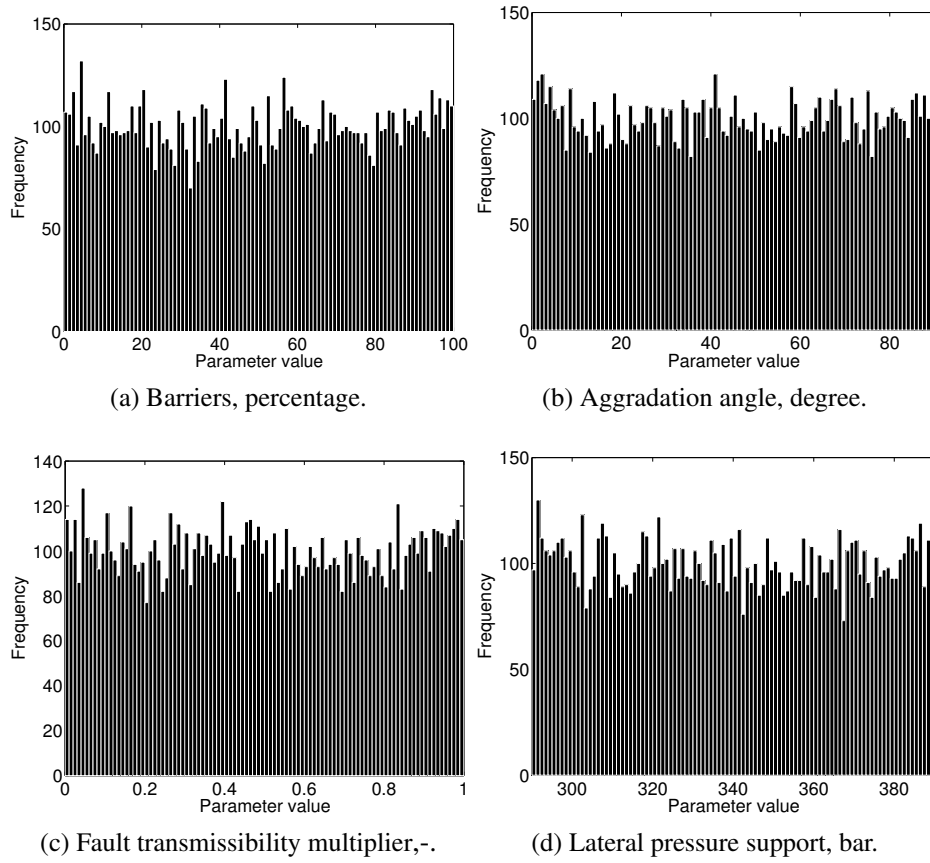
Figure 7: The histograms of geological variables used in this study are sampled from uniform distributions.

(a) Average aquifer pressure.

(b) Mobile CO_2.

(c) Residual CO_2.
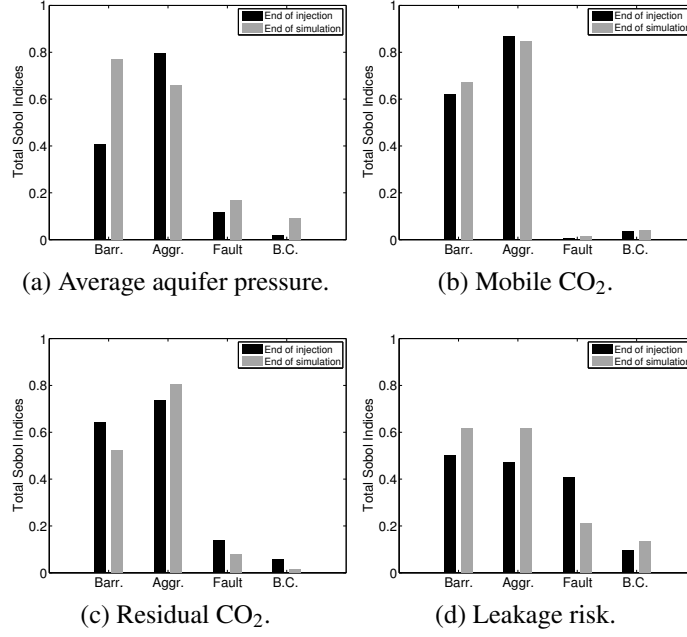
(d) Leakage risk.

Figure 8: Sensitivity analysis for different responses (a: average aquifer pressure, b: mobile CO_2, c: residual CO_2, and d: leakage risk) with respect to the uncertain parameters. In the figures above, Barr. is for barriers, Aggr. for aggradation angle, Fault for fault transmissibility, and B.C. for regional groundwater effect.

(a) Low aggradation angle.　(b) Medium aggradation angle.　(c) High aggradation angle.
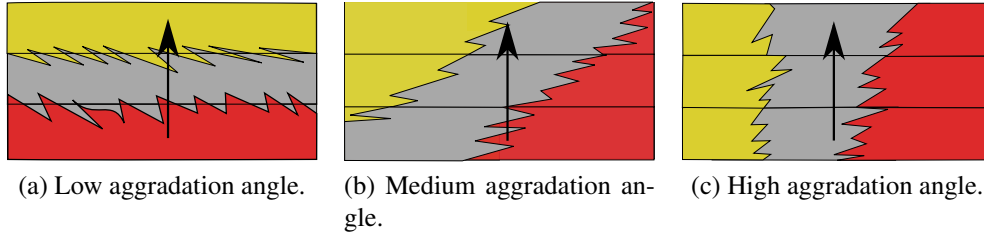
Figure 9: Illustration of how the aggradation angle affects the effective vertical conductivity.
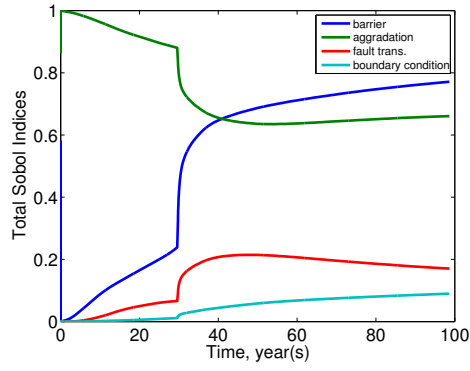
(a) Aquifer pressure.
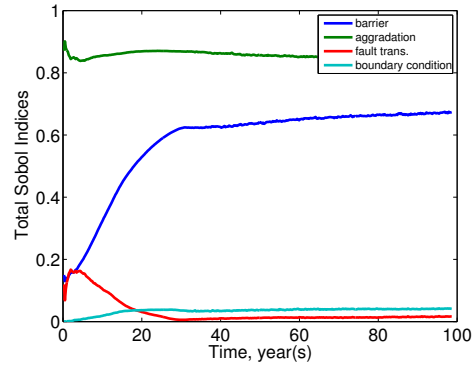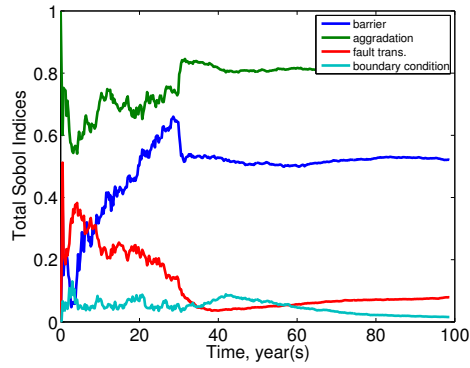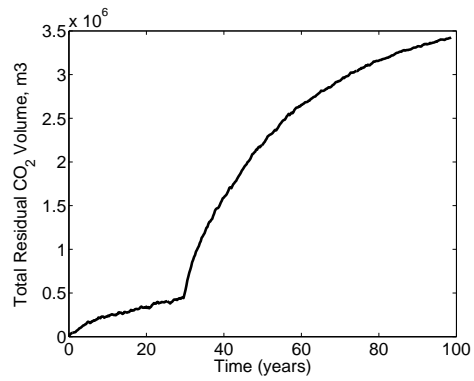
(b) Mobile CO_2 volume.

(c) Residual CO_2 volume.
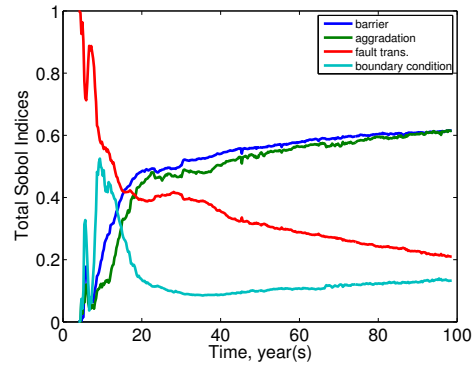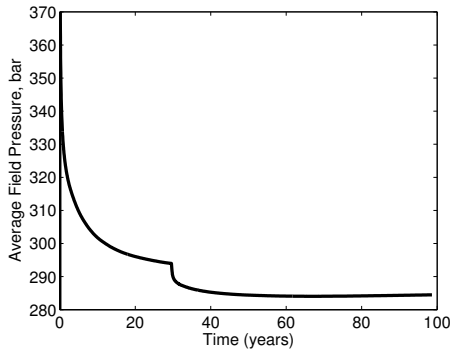
(d) Leakage risk.

Figure 10: Sensitivities (expressed by total Sobol indices) plotted versus time for different responses.
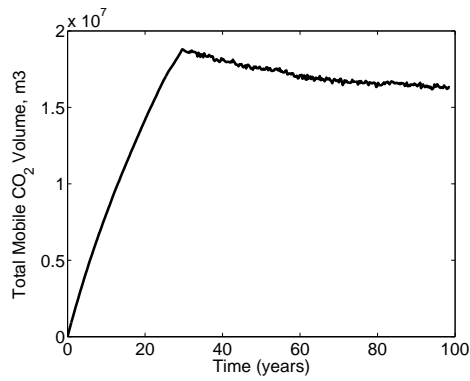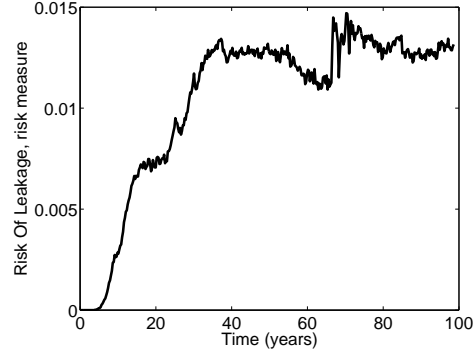
(a) Expected mobile CO$_2$ volume.

(b) Expected residual CO$_2$ volume.

(c) Expected pressure.

(d) Expected leakage risk.

Figure 11: Expectation for response values versus time. The presssure value for initial time step in Figure c goes up to 670 bars.
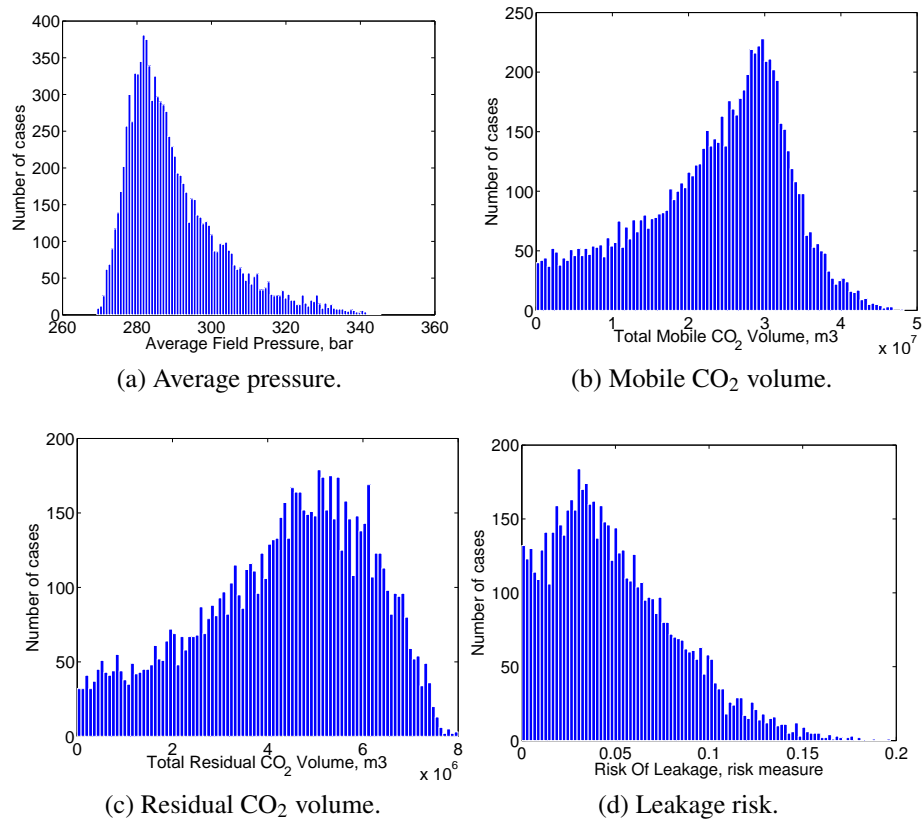
35

(a) Average pressure.

(b) Mobile CO_2 volume.

(c) Residual CO_2 volume.

(d) Leakage risk.

Figure 12: Histograms of selected response values.

(a) Average pressure.

(b) Mobile CO_2 volume.

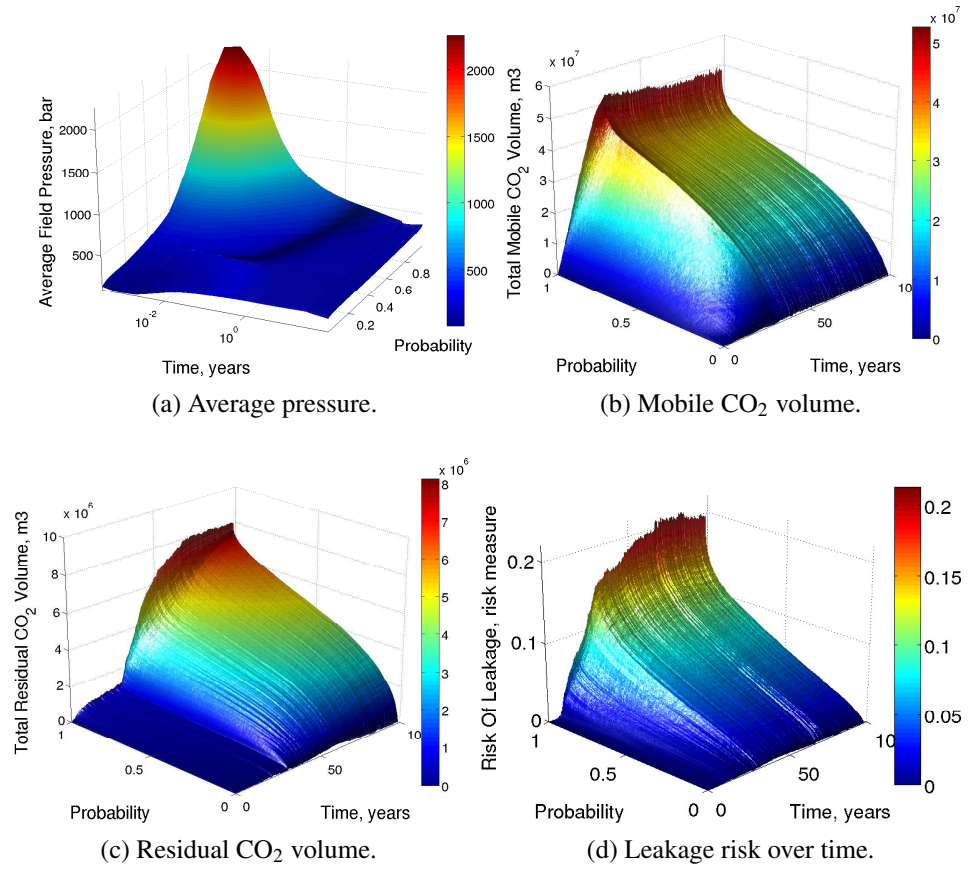(c) Residual CO_2 volume.

(d) Leakage risk over time.

Figure 13: Evolution of the cumulative distribution function of different response values over time.