International Journal of Greenhouse Gas Control 116 (2022) 103613

Contents lists available at ScienceDirect

International Journal of Greenhouse Gas Control

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

Dynamic estimates of extreme-case CO_2 storage capacity for basin-scale heterogeneous systems under geological uncertainty

Per Pettersson, Svenn Tveit, Sarah E. Gasda *

NORCE Norwegian Research Centre, Nygårdsgaten 112, Bergen 5008, Norway

ARTICLE INFO

Keywords:
Dynamic storage capacity
Pressure-limited capacity
VE methods
Uncertainty quantification
Basin-scale assessment
Heterogeneous field
Rare-event analysis
Subset simulation
Monte Carlo methods
Markov Chain Monte Carlo

ABSTRACT

Geological CO_2 storage is expected to grow dramatically in the coming decades to meet global climate targets. Assessment of worldwide storage resources using static methods indicates significant theoretical potential for large-scale deployment. Dynamic capacity estimates are needed at the basin-scale that fully capture the impact of geological uncertainty and account for regional limits on pressure buildup. Accurate quantification of the risk of low or critically low capacity under extreme occurrences of heterogeneity will be increasingly important. There are significant challenges associated with efficient computation of low probability capacity within Monte Carlo frameworks at these scales. In this paper, we propose a workflow for uncertainty quantification that is able to efficiently estimate increasingly outer percentiles of dynamic capacity such as P1, P0.1, or even lower probability events. Our approach is based on the rare-event methodology that uses a subset simulation approach to concentrate sampling of the parameter space in the tail regions of the capacity distributions. This approach greatly speeds up uncertainty quantification for very small probabilities compared to standard Monte Carlo. We demonstrate the method by introducing a correlated heterogeneity field to a highly prospective basin-scale system that can support regional injection rates of 100 million tons annually. We find that the outer quantiles are more sensitive to the underlying geostatistical model compared to the median P50 capacity. This implies that for large-scale systems, well characterized heterogeneity is essential to identify the likelihood of very rare yet still relevant dynamic estimates of storage capacity.

1. Introduction

Millions of tons of CO_2 emissions have been stored successfully at various sites around the world over the past several decades (Page et al., 2020). However, a dramatic scale-up of carbon capture and storage (CCS) is needed to realize the recommended emissions pathways laid out by the Intergovernmental Panel on Climate Change (IPCC) (Masson--Delmotte et al., 2018). Scale-up analyses indicate that CO_2 storage deployment at current growth rates could achieve 350 gigatonnes (Gt) stored by 2100, implying global injection rates approaching 20 Gt/y by the end of the century (Zahasky and Krevor, 2020). This exponential growth of CO_2 storage will rely on economies of scale in order to manage deployment costs, and therefore future storage development will likely become concentrated within a few large sedimentary basins. Development of several strategic continental margins at rates consistent with historic petroleum development could feasibly achieve >100 Gt CO_2 stored by 2050 (Ringrose and Meckel, 2019).

Basin-scale storage development involves maturing static capacity assessments for sedimentary basins (e.g. Bentham et al. (2014); Halland et al. (2014); Gray, 2015) through the use of dynamic simulation. The CO_2 Storage Resource Management System (SRMS) (Frailey et al., 2017) prescribes a process for maturing estimates that involves quantification of P10, P50 and P90 statistical quantiles to assess the impact of geologic uncertainty. However, the computational challenges associated with basin-scale dynamic capacity estimation are significant. Simulation grids should cover extensive areas on the order of 10^4 km^2 and consider multiple simultaneous storage sites. Multiphase flow processes associated with CO_2 injection and trapping within heterogeneous and geological complex systems must be considered. And finally, the uncertainty associated with a large number of geological parameters is significant, requiring 1000s of simulations within a Monte Carlo type framework to produce reliable estimates of expected capacity. Estimating very rare but important occurrences of critically low, extreme-case capacities would require significantly larger number of simulations, as the number of simulations required is inversely proportional to the probability of the occurrence of the event of interest.

* Corresponding author.
E-mail address: sarah.gasda@norceresearch.no (S.E. Gasda).

https://doi.org/10.1016/j.ijggc.2022.103613
Received 14 May 2021; Received in revised form 28 December 2021; Accepted 14 February 2022
Available online 8 March 2022

At the basin-scale, pressure is an important factor determining capacity that must be considered in addition to CO_2 containment (Birkholzer et al., 2015). Basin-scale capacity can be limited by pressure due to a number of factors, including multi-site pressure interference (De Simone and Krevor, 2021), geologic boundaries (Zhou and Birkholzer, 2011), groundwater protection (Birkholzer et al., 2011) and fault reactivation (Williams et al., 2014). While modeling and process understanding for these pressure-driven risks have advanced significantly in recent years, the added impact of permeability heterogeneity and associated uncertainty is less well known. Assessing the impact of heterogeneity involves a combination of two computationally intensive steps: (1) static geological modeling and (2) large-scale multiphase simulation. Despite the added challenges of incorporating uncertainty in heterogeneity, the use of uncertainty quantification for capacity estimation has received increasing attention. At the local scale of a single storage site, the impact of uncertainty on pressure-constrained capacity has been performed for synthetic shallow-marine depositional environments (Ashraf, 2014) as well as for realistic storage sites (Deng et al., 2012). A study of Rock Springs Uplift in Wyoming (Pawar et al., 2017) demonstrated that uncertainty in permeability heterogeneity significantly impacts the extent of the pressurized area, or Area of Review (AoR). The authors point to the difficulty in constraining permeability uncertainty leads to significant uncertainty in the estimated AoR.

While site-scale studies may test 10s to 100s of stochastic realizations of the permeability field, dynamic estimates of capacity at heterogeneous basin-scale systems often select only a few representative permeability cases to reduce the overall computational burden of the simulation study. Nevertheless, some useful insights have been obtained regarding the relation between pressure and heterogeneity for some prospective basins. For instance, capacity estimates for the Mount Simon basal sandstone increases when high-permeability zones are considered (Anderson and Jahediesfanjani, 2019; Birkholzer and Zhou, 2009). The capacity of the Utsira/Skade aquifer can be limited either by local injection pressure or by pressure limits in shallower zones depending on three chosen cases of heterogeneous aquifer properties (Elenius et al., 2018). A study of injection into dome structures of the Bunter Sandstone (Williams et al., 2013) also indicates that large-scale heterogeneity can cause local pressure buildup that constrains injection rates. In contrasting basins in the US and China, regional capacity estimates of saline aquifers were shown to be influenced by varying heterogeneity (high, mid, and low) in addition to a number of other factors such as number of wells, boundary conditions and water production (Gorecki et al., 2015). While these studies are valuable sensitivity studies, the limited sampling of the parameter space is likely insufficient to quantify the full effect of geological uncertainty on capacity estimates. Therefore, it would be difficult to rely on these estimates for the purposes of commercial deployment.

Quantifying the impact of geological uncertainty on capacity estimates involves consistent and statistically meaningful approaches to sample the underpinning parameter distributions. Approaches such as standard Monte Carlo (MC) can be readily applied, but the random sampling is often inefficient, resulting in many thousands of realizations to converge on a reliable estimate. There exists many available methods for accelerating uncertainty quantification (UQ) beyond standard MC methods. Sampling-based methods include variance reduction techniques, e.g., multilevel MC (Giles, 2015), latin hypercube sampling (Huntington and Lyrintzis, 1998), and importance sampling (Smith et al., 1997). Deterministic MC, e.g., space-filling methods such as quasi-MC, can be used to achieve higher convergence rates in the number of model evaluations (Caflisch, 1998). Another option is generalized spectral expansions, e.g., Karhunen-Loeve expansions and generalized polynomial chaos (Le Maître and Knio, 2010). These classes of methods have all been applied successfully in many different fields involving simulations of fluids, e.g., computational fluid dynamics (Barth et al., 2016; Najm, 2009), groundwater flow (Cliffe et al., 2011; Müller et al., 2011), flooding (Mondal and Mandal, 2020; Shaw et al., 2020), and weather forecasting (Gregory and Cotter, 2017; Sochala et al., 2020). Application of accelerated UQ methods within CO_2 storage, however, is more limited. Generalized polynomial chaos was used to quantify leakage of CO_2 from an aquifer in Oladyshkin et al. (2011), simulate residual trapping with a vertical equilibrium model in Pettersson (2016), and compared to other accelerated UQ methods for CO_2 storage in Köppel et al. (2019). Markov chain Monte Carlo (MCMC), a strategy for more efficient sampling, has been applied at the site-scale to estimate capacity under uncertainty (Deng et al., 2012). In Wriedt et al. (2014), the authors employed a surface response approach within a classical MC framework to efficiently quantify risk factors for CO_2 storage. The authors examined the impact of heterogeneity uncertainty on pressure development for random permeability fields in a simple box model. In Cao et al. (2020), a quasi-MC approach was used to train a surrogate model that was used to further predict impact of uncertainty on formation response. The above studies have been carried out for idealized or small-scale systems, and no study at this writing has applied UQ methods to basin-scale systems under uncertainty. We also note that while the surrogate approach in the latter two papers is a widely accepted substitute for full-physics models in applications such as classical fluid dynamics problems, their acceptance within subsurface applications such as reservoir engineering or CO_2 storage is limited.

Our study addresses the need for reliable and practical approaches for quantifying the impact of extreme occurrences of geologic uncertainty on basin-scale capacity estimation in heterogeneous systems. As many standard approaches can give good estimates of the expected value, or the median value (P50), we instead focus on estimation of the outer percentiles (i.e. P1, P0.1, ...) that are substantially more difficult to estimate efficiently and accurately using MC methods. Our approach is based on the so-called "rare-event" methodology, which is an umbrella term for estimating very small but presumably important probabilities. The advantage of this approach is that the estimation of increasingly unlikely events builds upon the previous more likely ones, such that there is natural and efficient path from one percentile to the next, e.g., from P10 to P1, from P1 to P0.1 and so forth. As an analogy, rare-event analysis can be applied to flood prediction to easily quantify the likelihood of a 1000-yr flood based on the likelihood of 100-year flood. For CO_2 capacity, the approach can be flipped such that one can identify the probability of a pre-defined *critical capacity*, i.e. the capacity at which regional, centralized storage development is not feasible. In other words, the rare event simulation allows one to effectively rule out the risk of not achieving some minimum acceptable capacity.

Rare-event simulation requires efficient conditional sampling restricted to regions of random space corresponding to critical events. By using MCMC sampling instead of standard MC, it is possible to concentrate sampling close to the region of interest without wasting computational resources on less important events. The efficiency can be improved by using surrogate models for the physical model, cf. Dostert et al., 2009; Ma and Zabaras, 2009 for examples from porous media, where sparse-grid collocation is used to reduce the number of full-scale simulations. Alternatively, the efficiency can be increased by more efficient MCMC sampling by reducing correlation between Markov states. Adaptive adjustment for nearly optimal performance was investigated in Zuev et al. (2012). Surrogate models and adaptive MCMC sampling can also be combined as in Elsheikh et al. (2014a), where polynomial chaos response surfaces were used to adjust the step lengths of the Markov chains used in nested sampling for subsurface flow models. Rare-event simulation within the context of CO_2 storage was proposed in Elsheikh et al. (2014b), where CO_2 leakage through an abandoned well was investigated.

In this paper, we perform rare-event simulation by means of the subset simulation methodology (Au and Beck, 2001) combined with conditional Karhunen-Loeve expansion to sample increasingly extreme and rare occurrences of the random permeability field. This rare-event framework is efficient and can be implemented easily with any standard forward simulator. We apply the proposed framework to a realistic

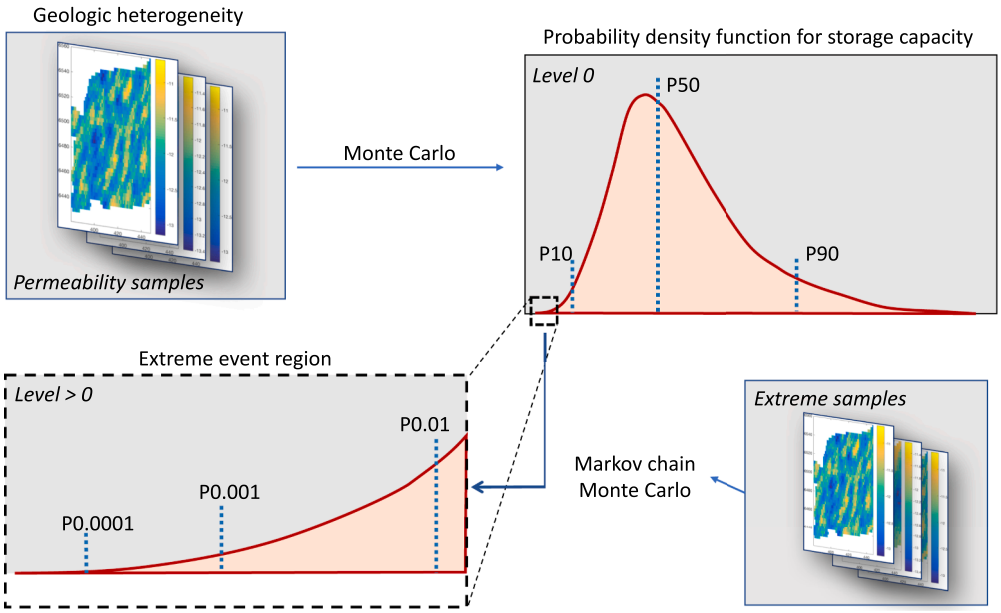


Fig. 1. Schematic figure of the overall approach for estimation of rare occurrences of storage capacity. A standard MC approach is used to generate the full PDF (level 0) from samples of geologic heterogeneity. MCMC is then used to sample the parameter space in the extreme event region. New levels of PDF (levels 1, 2, ...) are generated by subset simulation to progressively zoom in on smaller probability events (P1, P0.1, ...) with increased resolution.

aquifer based on the Utsira formation, a large aquifer located offshore on the Norwegian continental shelf. The formation is currently host to the Sleipner CCS project, and in addition has been subject of several large-scale injection studies, e.g. Gasda et al., 2017; Lindeberg et al., 2009; Møll Nilsen et al., 2015. In this study, we do not focus on further maturing capacity estimates for the Utsira. Instead, we use this formation as a test case for demonstrating the rare-event methodology. The Utsira is characterized by varying depth and topography, which makes it an interesting case for examining the impact of heterogeneity on the risk of triggering early pressure failure in shallower zones versus being constrained by local injection limits. However, we sample a larger range of permeability values than what is expected for the Utsira. We note that we focus solely on the constraints due to regional pressure buildup in order to characterize the link between heterogeneity, uncertainty and far-field pressure response that is thus far poorly understood. Overall, our motivation in choosing the Utsira is to understand the impact of heterogeneity on pressure-constrained saline aquifers that have optimal storage properties, i.e. thick, permeable sands with extensive well-connected pore volume.

2. Approach and numerical methods

The overall approach used to attain efficient and reliable estimates of extreme-case storage capacities is pictured schematically in Fig. 1. Our approach consists of three main elements: (1) a geostatistical model for heterogeneity to generate samples of the permeability field, (2) a method for targeted sampling of the parameter space, and (3) a forward simulator. These elements are connected together in a workflow that progressively zooms in and resolves the probability density function (PDF) in the extreme event region. The starting point is the full PDF, or level 0 distribution, for storage capacity using standard MC that samples the full geologic parameter space. Then, a MCMC method is applied to generate new levels of distributions that zoom in within intermediate critical regions. In this paper, we first concentrate on samples that fall within P10, when these are resolved we zoom in on P1, and so on, but other approaches are possible as long as the current focus region is a subset of the previous focus region. The main idea is that each new level progressively resolves a subset of the previous level PDF in a hierarchical manner. For instance, using samples from level 1 as seeds, level 2 resolves the subset below P1 to estimate P0.1, then level 3 resolves below P0.1 to estimate P0.01 where samples from level 2 have been used as seeds, and so forth. The end result is a set of highly resolved estimates of increasingly smaller probability events in the rare-event region. In this work it is convenient to let them be defined by percentiles, and then tenths of percentiles, and so forth, but any choice of increasingly small or increasingly large quantiles is possible. We note that the choice of how many levels to include in the estimation depends on the application, e.g., risk assessment. However, including very high number of levels is typically unnecessary since we only get estimates of highly unlikely events at an extraordinary computational cost.

The methods we use for each component are chosen to allow for high-resolution PDF in targeted regions to be carried out in an efficient and robust manner. Without efficient methods, it would be nearly impossible to resolve estimates of rare events within a reasonable computational time. We have applied methods to increase efficiency in all three components of the workflow:

- *Sampling method*: To maintain the number of model evaluations at a feasible level: A method to concentrate sampling of the parameter space to critical regions (cf. Section 2.1).
- *Geostatistical model*: An efficient parameterization of permeability (and other) uncertainty that honors data and allows for concentrated sampling by generating new samples that are in some sense similar to existing samples (cf. Section 2.2). This is in contrast to black-box methods where proximity of samples cannot be easily inferred.
- *Forward model*: An efficient numerical solver for the storage simulation to allow sufficient sampling at acceptable computational cost (cf. Section 2.3).

Next, we describe how these needs are satisfied in this work by means of subset simulation, geostatistical models based on conditional Karhunen-Loeve expansions of random fields, and vertical equilibrium models, respectively.

2.1. Monte Carlo subset simulation of extreme events

The standard MC method is a widely used approach to approximate integrals such as the expectation. Given a set of N samples $\eta^{(1)}, \ldots, \eta^{(N)}$ of

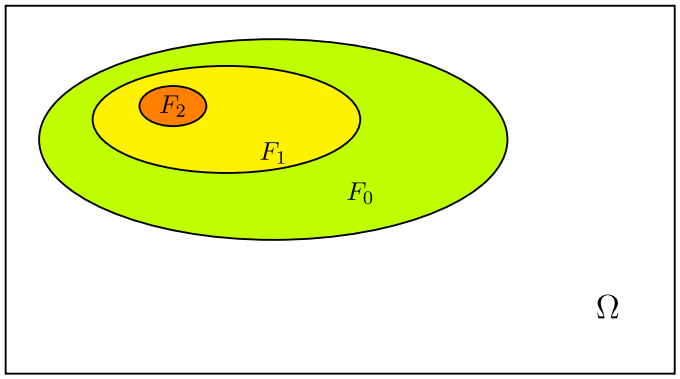


Fig. 2. Schematic figure of intermediate failure subsets. Sampling from the full domain Ω yields samples in F_0, which are used as seeds when sampling $F_1|F_0$, and so on for smaller subsets.

1. Generate candidate sample $\tilde{\eta}^{(n+1)}$ given $\eta^{(n)}$.
for j=1:M **do**
  Draw sample ξ_j from proposal PDF $p(\cdot|\eta_j^{(n)})$
  Compute ratio $r_j = \frac{\pi_j(\xi_j)}{\pi_j(\eta_j^{(n)})}$.
  Draw u from Uniform[0,1].
  if $u \leq \min(r, 1)$ **then**
    $\tilde{\eta}_j^{(n+1)} = \xi_j$.
  else
    $\tilde{\eta}_j^{(n+1)} = \eta_j^{(n)}$
  end
end
2. Accept or reject sample $\tilde{\eta}^{(n+1)}$.
% Check location of $\tilde{\eta}^{(n+1)}$.
if $Q(\tilde{\eta}^{(n+1)}) \in F_i$ **then**
  % Accept new sample
  $\eta^{(n+1)} = \tilde{\eta}^{(n+1)}$
  $Q^{(n+1)} = Q(\tilde{\eta}^{(n+1)})$
else
  % Reject new sample
  $\eta^{(n+1)} = \eta^{(n)}$
  $Q^{(n+1)} = Q^{(n)}$
end

Algorithm 1. Modified Metropolis algorithm (adapted from Au and Beck (2001)).

a vector of random variables $\eta = (\eta_1, \cdots, \eta_M) \in \mathbb{R}^M$ on the parameter space of interest, the MC method approximates the expectation of a quantity of interest (QoI), $Q(\eta)$, using the arithmetic mean,

$$\overline{Q}_{\mathrm{MC}} = \frac{1}{N}\sum_{n=1}^{N} Q(\eta^{(n)}). \tag{1}$$

The law of large numbers ensures that the arithmetic mean converges to the true expectation as $N \to \infty$. The main advantages with MC over deterministic numerical integration methods is that the convergence is not dependent of the dimension of the problem, and the QoI can be treated as a black box; that is only input-output interaction with $Q(\eta)$ is needed. The QoI in this paper is the total amount of stored CO_2, which we calculate from simulation results, and the vector of input parameters, η, is a parameterization of the permeability field. We assume that η has a product-type PDF $\pi(\eta) = \pi_1(\eta_1)\cdot\pi_2(\eta_2)\cdots\pi_M(\eta_M)$. We note that, in addition to approximation of the expectation, the MC samples can also be used to visualize the distribution of the QoI using, e.g., histograms or density estimations.

In this work, we are interested in computing the quantiles or probabilities of critical storage scenarios, i.e. capacities that are lower than some minimum threshold. In general terms, this can be described as computing the failure probability, denoted p_F, where "failure" should be interpreted in a wide sense, i.e. failure to meet expectations rather than system breakdown. An interesting feature of failure probability simulation with MC is that the accuracy of the resulting estimator, encoded as the coefficient of variation COV(·) (relative variability with respect to the mean value), has a simple expression that holds independent of physical model for the problem of interest:

$$\mathrm{COV}\left(p_F^{\mathrm{MC}}\right) = \sqrt{\frac{1-p_F}{Np_F}}.$$

This shows that for a desired accuracy, the number of MC samples N needed is proportional to $1/p_F$ which can be prohibitively expensive for realistic physical models since failure hopefully occurs only with a very small probability for robust systems. Subset simulation remedies this issue by successively zooming in to regions of parameter space of decreasing probability until a target region has been reached. In the context of critically low storage capacity, these regions or subsets of the space of all possible capacities, can be defined by certain quantiles, e.g. P10, P1, etc. More generally, let Q_{crit} be a critically low value of a QoI Q whose probability is very small but needs to be estimated to mitigate risk. In this work Q_{crit} will be understood as a storage capacity but the following description is general and the quantity of interest can be anything where unusual behavior needs to be quantified. Then $p_F = \mathbb{P}(Q \leq Q_{\mathrm{crit}})$, where $\mathbb{P}(\cdot)$ denotes the probability measure.

We now outline the subset simulation procedure introduced in the seminal work Au and Beck, 2001. Introduce a sequence of intermediate failure events $\{F_\ell\}_{\ell=0}^{L}$ where F_ℓ denotes the outcome that $Q \leq Q_\ell$, for some value Q_ℓ with the property $Q_{\mathrm{crit}} = Q_L < Q_{L-1} < \cdots < Q_0$. Then the probability of an intermediate failure is $\mathbb{P}(F_\ell) = \mathbb{P}(Q \leq Q_\ell)$. Note that these intermediate failure thresholds Q_ℓ do not need to correspond to any values that have some practical meaning. In fact, they will be estimated by the numerical method as a means to obtain the QoI Q_{crit}. The situation is illustrated schematically in Fig. 2 with $F_2 \subset F_1 \subset F_0$ subsets of the stochastic domain Ω. The boundaries between the subsets can be interpreted as the contour curves of the QoI Q at values Q_0, Q_1, Q_2.

By the definition of conditional probability, the probability of failure is given by

$$p_F = \mathbb{P}(F_L|F_{L-1})\mathbb{P}(F_{L-1}|F_{L-2})\cdots\mathbb{P}(F_1|F_0)\mathbb{P}(F_0), \tag{2}$$

where $\mathbb{P}(A|B)$ denotes the probability of outcome A given that B has occurred. Even if p_F is very small, if the intermediate failures are appropriately chosen, all factors of the right hand side of Eq. (2) can be large enough to be efficiently computed by a reasonable number of MC samples. For instance, if $p_F = 10^{-6}$ and all terms $\mathbb{P}(F_\ell|F_{\ell-1}) = 0.1$, then instead of using $\mathcal{O}(10^6)$ standard MC samples for accurate estimation, subset simulation requires $6 \times \mathcal{O}(10)$ samples.

The last factor of Eq. (2) can be efficiently computed with standard MC:

$$\mathbb{P}(F_0) \approx \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{Q\leq Q_0}\left(Q^{(n)}\right),$$

where $Q^{(n)} = Q(\eta^{(n)})$, and the indicator function $\mathbb{1}_{Q\leq Q_0}$ is 1 if $Q \leq Q_0$ and 0 otherwise. Computing the remaining factors of Eq. (2) with standard MC is not efficient as it does not address the increasing rarity of the events simulated, i.e. the full random domain would be sampled to find increasingly rare events. However, the event $F_\ell|F_{\ell-1}$ is not very unlikely even though both F_ℓ and $F_{\ell-1}$ occur with very small probability. Mathematically, the estimator

User-defined parameters:
p (intermediate failure prob.), N_L (samples per level), $Q_{\text{crit.}}$ (critical value of QoI)
$N_{\text{M.chains}} = pN_{\text{L}}$
Level 0.
$L = 0$.
Run standard MC to produce QoI samples $\{Q_{L=0}^{(n,L=0)} : n = 1, \ldots, N_{\text{L}}\}$.
$N_{\text{fail}} = \#\{Q^{(n,L)} \leq Q_{\text{crit}} : n = 1, \ldots, N_{\text{L}}\}$ % *Compute number of failures*
Level 1,2,3,...
while $N_{\text{fail}} \leq pN_{\text{L}}$ **do** % *As long as only a small number of failures detected*
    Sort the samples: $Q^{(1)} \leq Q^{(2)} \leq \cdots \leq Q^{(N_L)}$.
    Set intermediate failure boundary $Q_L = (Q^{(N_{\text{M.ch}})} + Q^{(N_{\text{M.ch}+1})})/2$ to define region F_L.
    Save the smallest $N_{\text{M.ch}}$ samples as seeds for the next level.
    $L = L + 1$ % *Go to next subset level*
    for $m = 1 : N_{\text{M.ch}}$ **do** % *Total of* N_{L} *samples per level.*
        Produce $1/p$ samples of QoI within F_{L-1} using MMA [Alg. 1].
    end
    Collect new samples $Q^{(n,L)}$.
    $N_{\text{fail}} = \#\{Q^{(n,L)} \leq Q_{\text{crit}} : n = 1, \ldots, N_{\text{L}}\}$ % *Compute number of failures*
end
Outputs:
$p_F = p^L N_{\text{fail}}/N_L$
$Q_0, Q_1, \ldots, Q_{L-1}$ % *Quantiles* $p, p^2, \ldots, p^L$

Algorithm 2. Subset simulation (adapted from Au and Beck (2001) and Zuev (2013)).

$$\mathbb{P}(F_\ell | F_{\ell-1}) \approx \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{Q \leq Q_\ell}(Q_\ell^{(n)}),$$

with $Q_\ell^{(n)}$ being samples from the subset $\{Q \leq Q_\ell\}$, does not suffer from the needle-in-a-haystack phenomenon. The challenge is to make sure that Q is sampled within this set to avoid unnecessary and expensive full model evaluations. To sample the distribution restricted to a small subset rather than the full domain, a MCMC approach is used. The Modified Metropolis algorithm (MMA) generate samples that converge to the target distribution by means of Markov chain random walks starting from seeds within the probability domain of interest. Hence, all Markov chains start from their stationary states so there is no initial burn-in period of samples to be discarded. All samples generated will be used in the rare-event estimators, which is an example of so-called perfect simulation (Robert et al., 2004). Algorithm 1 outlines the MMA in pseudocode for the generation of a single new sample $\eta^{(n+1)}$ from an existing sample $\eta^{(n)}$. Note that the new sample is either a copy of the existing sample, or generated by taking a step determined by a user-defined proposal distribution $p(\cdot|\eta^{(n)})$. Algorithm 2 summarizes the full subset simulation method which calls Algorithm 1 to generate samples from the subsets.

The performance of subset simulation is dependent on the correlations between successive states of the output samples. Theoretically optimal acceptance rates are available for some distributions, with the frequently cited and perhaps surprisingly low number 0.234 for high-dimensional problems (Roberts et al., 1997). In practice, the exact distribution is usually not known, and one can rely on various rules-of-thumbs or empirical estimates to scale the proposal distribution to obtain acceptance rates within some interval, e.g., $[0.3,\ 0.5]$ as suggested in Zuev et al. (2012).

In this work, we perform adaptive scaling of the proposal distribution in Algorithm 1 by grouping the seeds of the Markov chains into smaller batches, run a batch and then evaluate its local empirical acceptance rate. If the rate is above 0.5, the scaling of the proposal distribution is increased by 50%. If the acceptance rate is below 0.3, the scaling is decreased by 50%. For the first batch on each level, the scaling is equal to the standard deviation of the sample seeds, where any correlation between conditional random variables has been ignored. This correlation could however be accounted for by scaling the proposal distribution with a Cholesky factorization of the covariance matrix of the seeds, instead of the main diagonal of the same covariance matrix, as done here.

2.2. Geostatistical modeling

The samples of the permeability field, which are used in the MCMC method (Section 2.1), are generated using a geostatistical modeling approach. In reservoir characterization, geostatistical methods are widely used since they capture the heterogeneity seen in many geological formations, and realizations of uncertainty can be generated with relatively simple procedures. Perhaps the simplest procedures are based on the underlying assumption of Gaussian random fields, leading to realizations generated using only a mean field and covariance matrix. While the mean is easily made as the most likely permeability field from prior knowledge, much effort can be made in the generation of the covariance matrix.

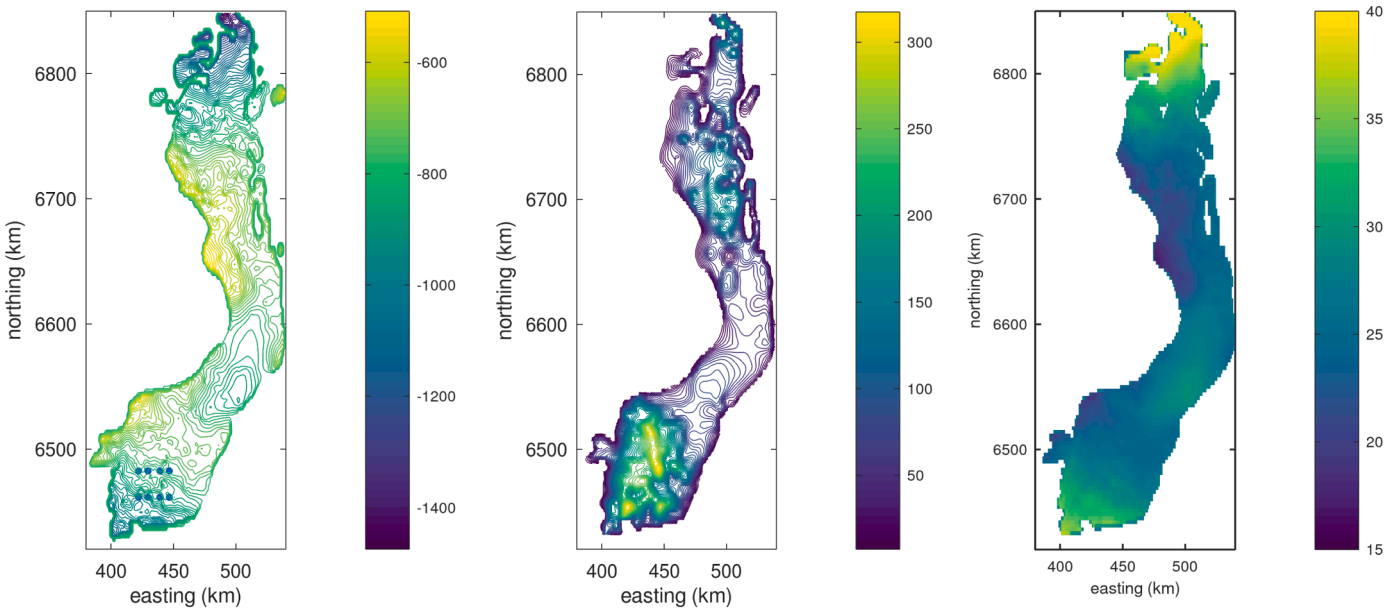
In this paper, we use a covariance function conditioned on data corresponding to fixed permeability values in the well cells. Furthermore, we assume that the permeability field, k, follows a log-normal distribution. The resulting Gaussian process, descibed in more detail in Rasmussen (2004), is $\mathcal{GP}(\log\bar{k}_{\text{cond}}, C_{\text{cond}})$ where

$$\log\bar{k}_{\text{cond}}(x) = \log\bar{k}(x) + C(x, x_{\text{data}})C^{-1}(x_{\text{data}}, x_{\text{data}})(\log k_{\text{data}} - \log\bar{k}(x_{\text{data}})), \quad (3)$$

$$C_{\text{cond}}(x, x') = C(x, x') - C(x, x_{\text{data}})C^{-1}(x_{\text{data}}, x_{\text{data}})C(x_{\text{data}}, x'), \quad (4)$$

where $\bar{k}$ is the mean permeability field, $C(\cdot,\cdot)$ is the assumed covariance function, and k_{data} is a vector of permeability data, located at the well positions x_{data}.

A parameterization of the uncertainty in permeability is both desirable for analysis purposes, and necessary for the subset simulation in Algorithm 2. It can be achieved by means of a truncated Karhunen-Loeve expansion of the conditional Gaussian process,



(a) formation top (mbs) (b) formation thickness (m) (c) maximum overpressure (bar)

Fig. 3. Utsira formation geometry. Data provided by Kirby et al. (2001). In (a) the injection locations are indicated by black dots.

Table 1
Formation and fluid properties for storage simulations.

Property	Description	Value
Porosity	homogeneous	0.38
Pore compressibility	constant	9 GPa^{-1}
Initial pressure	hydrostatic	100 bar @ 1000 m
Temperature	isothermal, 35 C/km	39 C @ 1000 m
Brine density	constant	1020 kg/m^3
Brine viscosity	constant	0.69 mPa·s
CO_2 density	variable	table (Span and Wagner, 1996)
CO_2 viscosity	variable	function (Vesovic et al., 1990)
Relative permeability under drainage conditions		
Brine	end-point value	1.0
	residual sat.	0.11
CO_2	end-point value	0.75
	residual sat.	0

$$\log k_{\text{cond}}(x,\eta) \approx \log \bar{k}_{\text{cond}} + \sum_{j=1}^{M} \sqrt{\lambda_j} v_j(x) \eta_j, \tag{5}$$

where $\{\eta_j\}$ are assumed standard Gaussian random variables, and $\{\lambda_j, v_j\}$ are the pairs of eigenvalues and eigenfunctions of the generalized eigenvalue problem

$$\int_D C_{\text{cond}}(x,x')v(x')\mathrm{d}x' = \lambda v(x), \tag{6}$$

where D denotes the physical domain. The generalized eigenvalue problem, Eq. (6), has analytical solutions for specific covariance functions (Ghanem and Spanos, 2002), but in most cases it must be solved using a numerical approach. In this paper, we approximate the integral in Eq. (6) with the midpoint rule on an equidistant grid, resulting in the following eigenvalue equation

$$\mathbf{CV} = \mathbf{\Lambda V}, \tag{7}$$

where

$$\mathbf{C}_{ij} = C_{\text{cond}}(x_i, x_j)\Delta.$$

Here, Δ is the grid size, and x_i is the cell center of grid cell i. Furthermore, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M]$, where $\mathbf{v}_j = [v_j(x_1), v_j(x_2), \ldots, v_j(x_n)]^T$ with n being the number of grid cells, and $\mathbf{\Lambda}$ is the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$. Lastly, we note that $M \leq n$, and that the eigenvalue Eq. (7) is solved using a standard QR method.

2.3. Storage simulation

A vertical equilibrium (VE) model (Gasda et al., 2011) was used to perform the storage simulations. The VE model assumes an instantaneous gravity segregation of CO_2 and brine at the time scale of the simulation (Nordbotten and Celia, 2012). We employ a sharp interface assumption, meaning that capillary pressure between CO_2 and brine is neglected. By neglecting capillarity, the CO_2 will migrate due to pressure and buoyancy faster than if capillarity is considered, and thus reflects the "worst-case scenario" with respect to storage security (Gasda et al., 2012). Variable CO_2 properties are modeled. Residual and structural trapping processes are considered, while solubility and mineral trapping are neglected. The VE model simulates the pressure and saturation change in the storage aquifer in space and time. Pressure is simulated at the formation top, and the pressure profile at depth is known due the VE assumption (fluid-static in the vertical dimension) and gravity segregation (CO_2 is lighter and overlies brine). Saturation is a depth-averaged quantity, which can be used to reconstruct the thickness of the CO_2 plume under the assumption of complete gravity segregation.

We choose the VE method as it is an efficient forward model, which is a desirable property for speeding up the computational time of this study. VE methods have been tested in the literature and are suitable at large scales (e.g. Class et al. (2009)). We emphasize that at the basin scale, 99% of the total pore volume remains saturated with formation water even after 50 years of CO_2 injection. In addition, our focus is on pressure build-up, where full resolution of plume dynamics is a secondary effect. In any case, the subset simulation methodology described earlier can be combined with any forward simulator.

3. Test case parameters

3.1. Utsira formation description

The Utsira formation has been well described in the literature. In this

Table 2
Geostatistical parameters to generate log$k$ realizations for the test cases in Section 4.

Case	$\log\bar{k}$ [-]	β [-]	γ [km]	α [-]	M
Base case	−27.63	0.18	40.125	0.28	1861
Low-var	−27.63	0.018	40.125	0.28	1861
High-perm	−26.53	0.018	40.125	0.28	1861
Short-corr	−27.63	0.18	20.63	0.28	3701
Low-perm	−28.32	0.018	40.125	0.28	1861
Equal-corr	−27.63	0.18	40.125	1.00	853
Vert-corr	−27.63	0.18	9.63	9.35	1494

study, data on top surface (Fig. 3(a)) and thickness (Fig. 3(b)) are provided by the British Geological Survey (Kirby et al., 2001). The spatial domain is divided into 8110 active cells of equal size: 1070 m x 3000 m. The rock, petrophysical and fluid properties are taken from literature, see Table 1. The domain is initially fully saturated with brine. As migration and trapping of CO_2 is not the primary focus, a simplified model for capillarity is used and dissolution of CO_2 in brine is neglected. As CO_2 is injected continuously during all simulations, the system is exclusively under drainage conditions. Therefore, residual brine saturation is set to 0.11, while CO_2 residual saturation under drainage conditions is set to zero to be consistent with a fully brine-saturated initial condition.

The boundary conditions for the full Utsira model assume an impermeable top seal and bottom boundary. All outer lateral boundaries are impermeable to flow. The injected volume of CO_2 is constrained by the maximum overpressure (Fig. 3(c)), which is defined as the pressure that exceeds the least compressive stress assuming negligible tensile strength of the rock. Following Bohloli et al. (2015), the horizontal stress is the minimum stress and can be estimated by

$$\sigma_h = 0.134z, \tag{8}$$

where the units of σ_h is bars and for z is meters below sea level (mbs). If the initial pore pressure (Pa) is estimated by hydrostatic pressure given the sea depth h_{sea} (m), sea water density ρ_w, and brine density ρ_b,

$$p_0 = \rho_w g h_{\text{sea}} + \rho_b g z, \tag{9}$$

where g is the gravitational constant, then, the maximum allowable overpressure at any given depth is

$$\Delta p_{\max} = \sigma_h - p_0. \tag{10}$$

3.2. Geostatistical properties

To generate realizations of the permeability field for the test cases, we use the conditional Karhunen-Loeve method presented in Section 2.2. A common choice for the covariance function in geostatistical modelling is the spherical covariance function,

$$C(x_i, x_j) = C(h_{ij}) = \beta \begin{cases} 1 - \dfrac{3h_{ij}}{2\gamma} + \dfrac{h_{ij}^3}{2\gamma^3}, & \text{for } 0 \le h_{ij} \le \gamma, \\ 0, & \text{for } h_{ij} > \gamma, \end{cases}$$

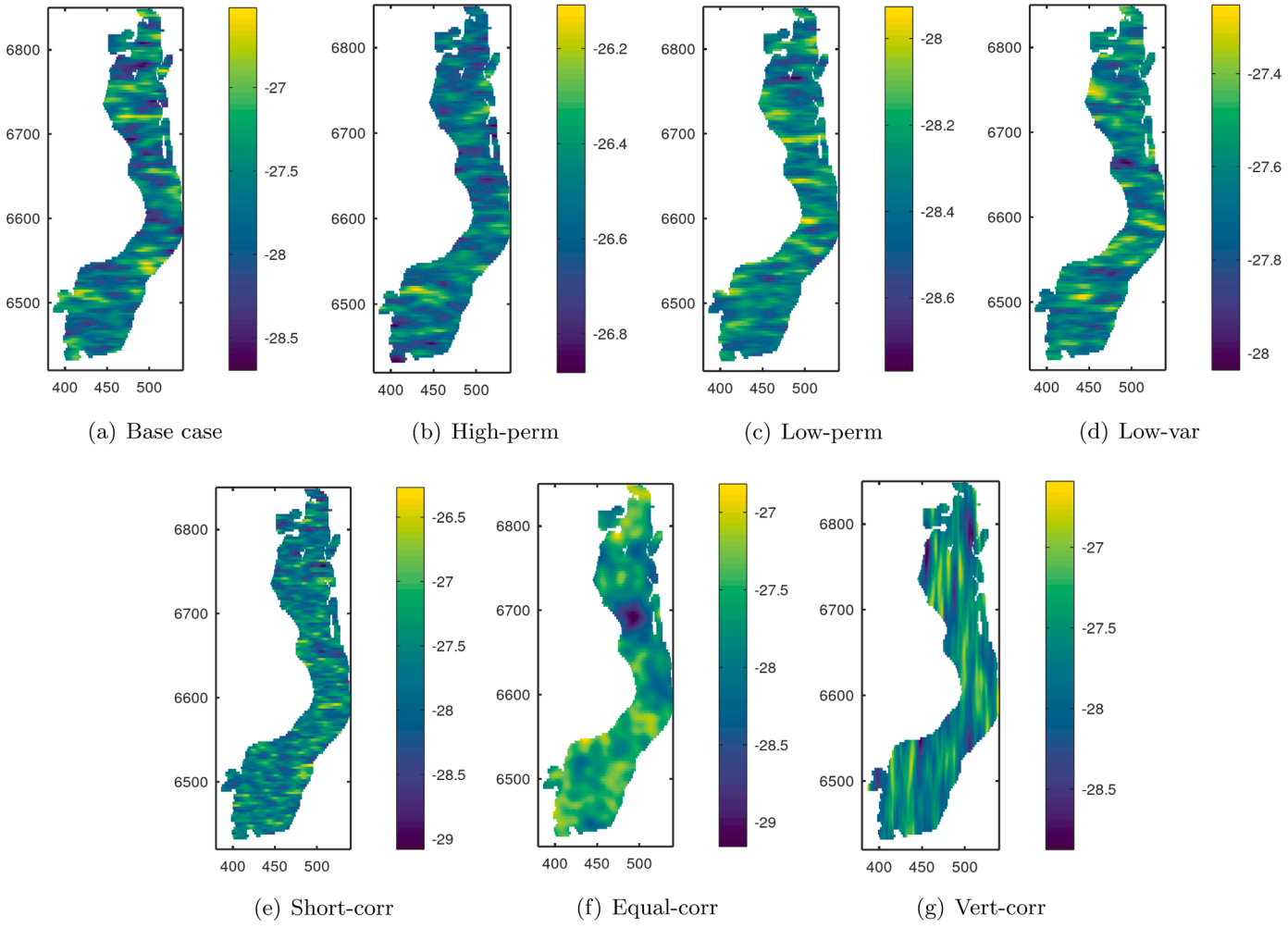
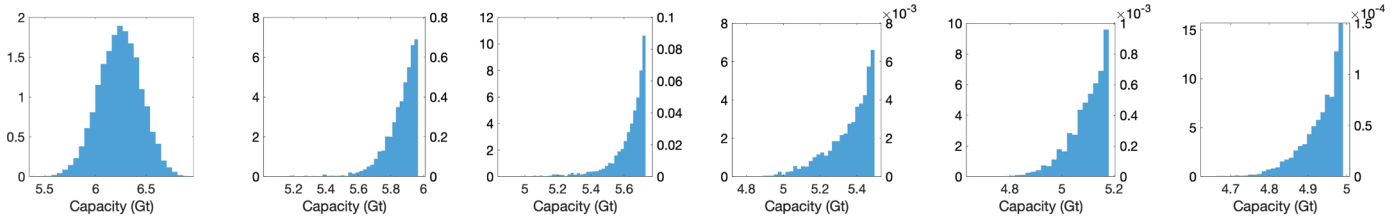


Fig. 4. Permeability maps (log$k$) for each test case. Plotted are representative samples close to the median.



(a) Level 0. (b) Level 1. (c) Level 2. (d) Level 3. (e) Level 4. (f) Level 5.

Fig. 5. Base case. Mean permeability 1 D. Histograms of samples of storage capacity on all levels.

Table 3
Percentiles for the different test cases. All values in Gt.

Case	P50	P10	P1	P0.1	P0.01	P0.001
Base case	6.24	5.97	5.73	5.50	5.18	4.99
Base case$_{5000}$	6.24	5.97	5.74	5.52	5.23	5.02
Base case$_{1000}$	6.25	5.97	5.70	5.47	5.28	5.09
Low-var	6.38	6.29	6.21	6.15	6.10	6.06
High-perm	6.88	6.79	6.72	6.67	6.63	6.60
Short-corr	6.25	6.06	5.62	5.21	5.04	4.69
Low-perm	4.98	4.88	4.79	4.73	4.68	4.63
Equal-corr	6.30	6.04	5.83	5.67	5.54	5.43
Vert-corr	6.34	6.14	5.97	5.77	4.82	4.74

where $h_{ij} = \| x_i - x_j \|_2$, β is the variance, and γ is the correlation length. To allow for correlations that are predominately in one direction (anisotropy), we scale the y-coordinate with a factor α.

In all test cases, the mean permeability field, $\bar{k}$, will be homogeneous, such that the structure of the generated permeability fields derives from the shape of the covariance function. Thus, we can systematically test different combinations of geostatistical parameters and investigate the impact they will have on the outcome of the subset simulations. The combinations investigated in Section 4 are given in Table 2. Representative realizations from each case are shown in Fig. 4. Lastly, we note that we only use 90% of the largest eigenvalues when generating realizations in Eq. (5) to lower the computational cost while still keeping the dominating features of the permeability fields. The number of eigenvalues, M, preserved in each case is also given in Table 2.

3.3. Injection locations and schedule

A total of 8 injection locations are spaced equally in the deepest and thickest zone of the southern Utsira (Fig. 3a). These wells are located 7.5 km apart in the easting direction and 21 km apart in the northing direction. In order to maintain a simplified computational approach, injection wells are not explicitly represented. Instead, the injection cells are average representations of a well at the subscale.

Large-scale deployment of CO_2 storage will likely start with smaller rates that gradually increase over time. The injection rates used in this study begin at 1 Mt/y and increase at a rate 1 Mt/y for 12 years. The maximum injection rate is 12 Mt/y for all cases except in the low-var case where it is 11 Mt/y. If the average pressure in the injection cell reaches the pre-defined limit, a well control function decreases the injection rate by a factor of 5. As wells are not explicitly defined, this control is not equivalent to BHP control, but is a control on the pressure at some effective radius from the well.

Injection continues until a pressure limit is reached anywhere in the domain outside of the injection cells, at which time the simulation is stopped. The subset simulation algorithm records the total amount of CO_2 in the domain, the simulated time, and the cell number that has reached the pressure limit.

4. Results

The subset simulation algorithm described in Section 2.1 is run on a total of six levels (0–5) with 10,000 samples per level, and intermediate failure threshold $p = 0.1$. With 1000 samples as seeds from the previous level, this results in $10,000 + 5{\cdot}9000 = 55,000$ VE model evaluations for each of the seven test cases. The choice $p = 0.1$ is common in the literature and has the advantage that the percentiles P10, P1, etc., are directly given by the output intermediate failure thresholds in Algorithm 2. Note that in the current problem setup there is no natural definition of the failure event Q_{crit} with corresponding probability p_F, so these values are not reported. If such QoIs are wanted, they can be computed directly from the simulation results used to estimate the quantiles and histograms to be reported next.

We use adaptive scaling of the proposal distribution so that all acceptance rates fall in the interval $[0.3,\ 0.5]$, as described in Section 2.1. Although the literature suggests Gaussian or uniform proposal distributions with no clear preference (Au and Beck, 2001), we have found that uniform proposal distributions perform better for the test cases investigated here, and that has been used in all reported results. The reason for this is that in the high-dimensional spaces we sample, acceptance often occurs when the Markov chain moves in only one or very few out of up to more than 3000 dimensions. With a Gaussian proposal distribution with step size σ_{prop}, obtaining at least one proposed random step of size on the order of $1/\sigma_{\text{prop}}$ is quite likely even for big values of σ_{prop}. As a result, one may have to increase the scaling by orders of magnitude with a Gaussian proposal distribution before observing a significant decrease in the acceptance rates. This is not the case with uniform proposal distributions, where relatively modest changes in the proposal distributions yields observable effects in acceptance rates.

We provide histograms of the storage distributions (in Gt) for all levels. These are estimates of the conditional PDFs, e.g. the PDFs given QoIs below some threshold value. For all cases, we provide histograms for the full distribution with MC (level 0), and levels 1 to 5. For the latter the left vertical axis show the estimated PDF density, i.e., it is normalized to integrate to unity. The right vertical axis has the same scale as the level 0 histogram to aid in assessing how small these event regions are compared to the full level 0 PDF. Histogram appearance is sensitive to the number of bins, which can be chosen according to various rule-of-thumbs. An inappropriate choice of bins can make the histogram appear to display peaks not present in the underlying data, or gaps that are also not expected for a continuous distribution. To avoid such artifacts, the number of bins in the following histograms have been set manually. We emphasize that this design parameter does not affect the numerical results presented, i.e., the percentiles in Table 3.

It is a delicate matter to make sure that acceptance rates remain within the interval [0.3 0.5]. Even if the prescribed rates are maintained on average, batches generated before the scaling of the proposal distribution has been suitably adapted may show unwanted behavior such as excessive smearing and multimodality (locally too high acceptance rates), or sharp peaks (locally too low acceptance rates). The histograms of the base case, low-var and high-perm are qualitatively similar, as

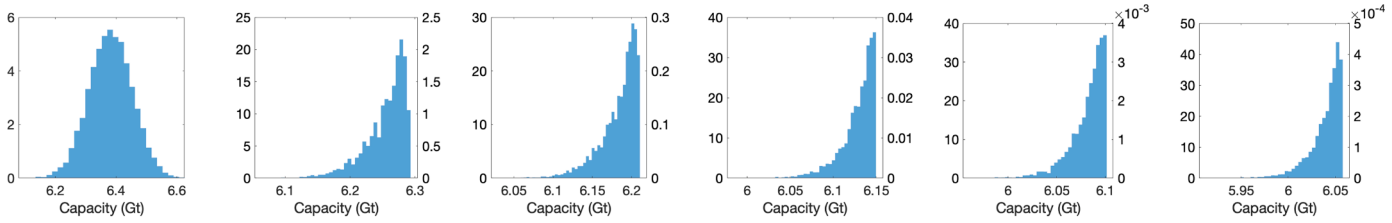


Fig. 6. Low-var. Permeability field variance reduced by a factor 10. Histograms of samples of storage capacity on all levels..

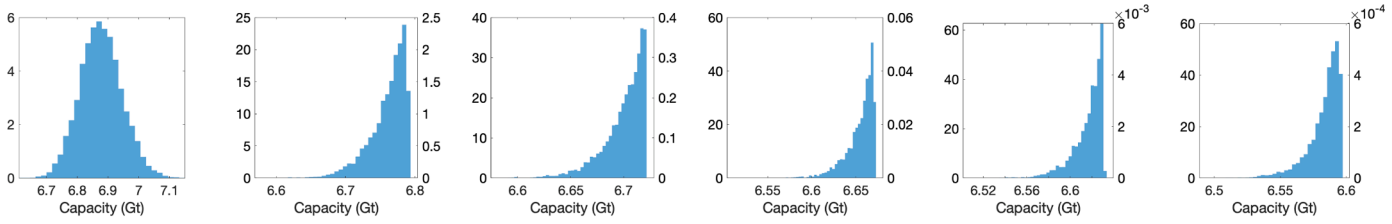


Fig. 7. High-perm. Mean permeability of 3 D. Histograms of samples of storage capacity on all levels..

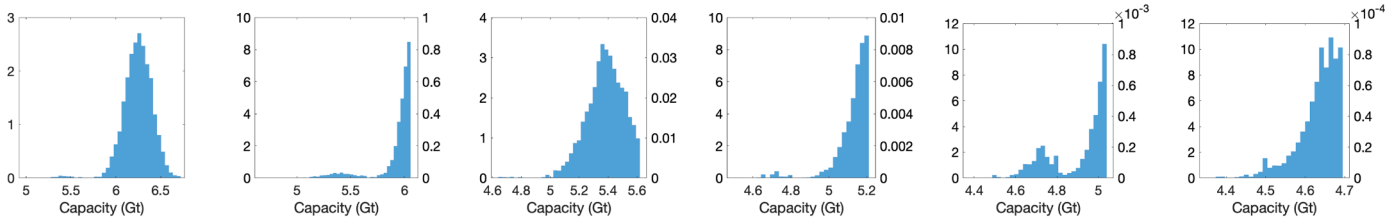


Fig. 8. Short-corr. Correlation length reduced by a factor 2 compared to base case. Histograms of samples of storage capacity on all levels..

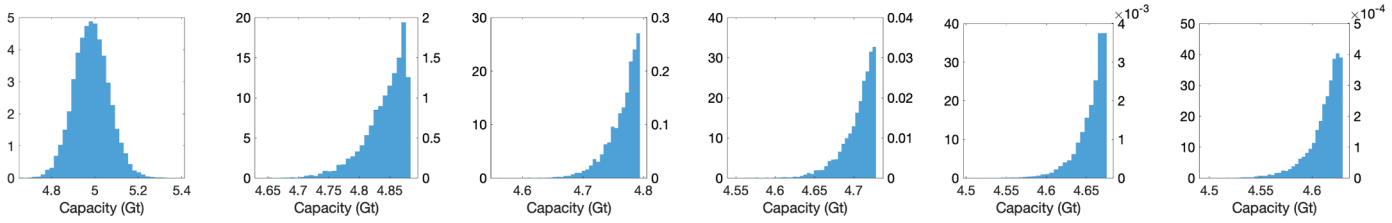


Fig. 9. Low-perm. Mean permeability of 0.5 D, and correspondingly decreased injection rate. Histograms of samples of storage capacity on all levels..

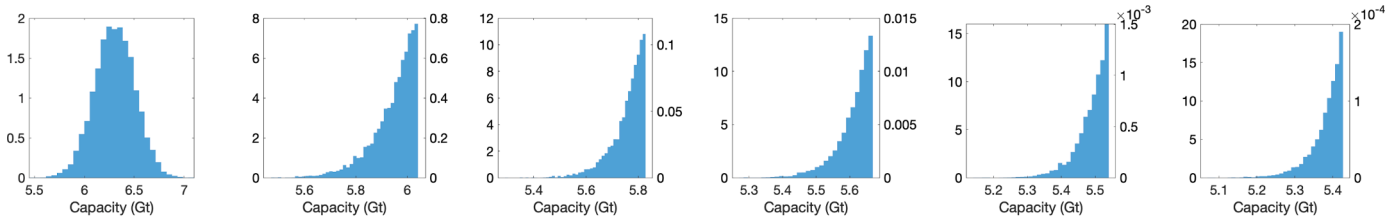
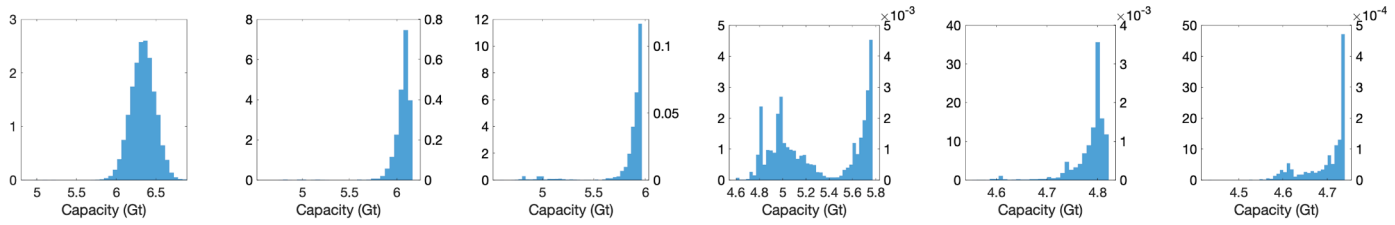


Fig. 10. Equal-corr. Same correlation length in horizontal and vertical direction. Histograms of samples of storage capacity on all levels..



(a) Level 0. (b) Level 1. (c) Level 2. (d) Level 3. (e) Level 4. (f) Level 5.

Fig. 11. Vert-corr. Correlation length in vertical direction is increased by a factor 9.35 compared to horizontal direction. Histograms of samples of storage capacity on all levels..

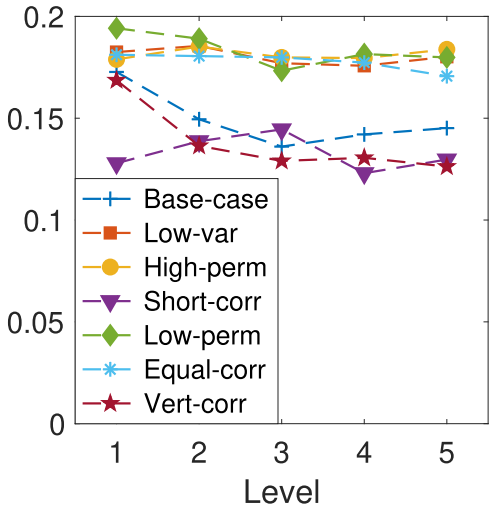
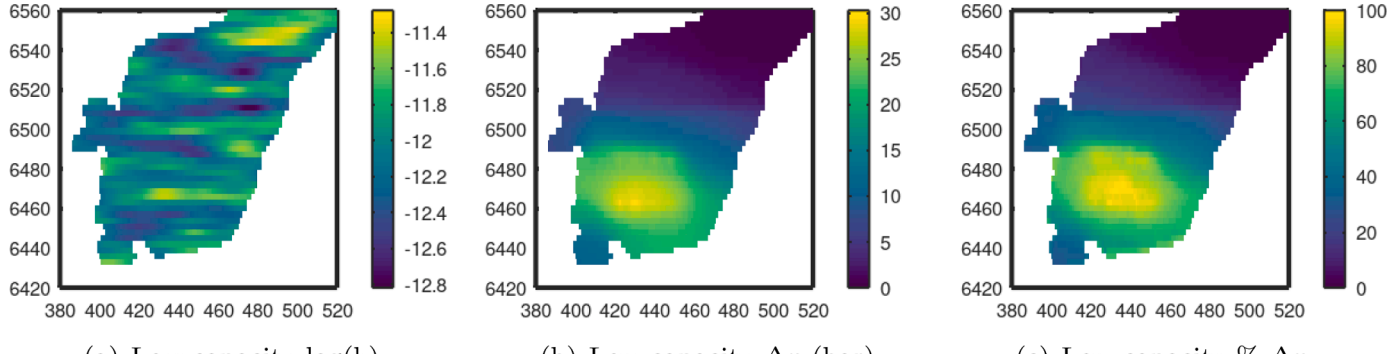


Fig. 12. Efficiency factors for all test cases as a function of level.

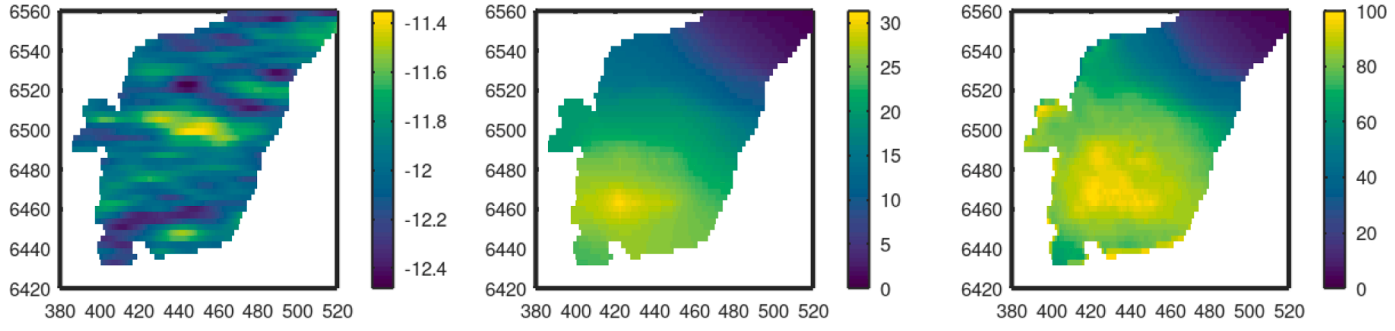
shown in Figs. 6 and 7. This is not surprising as the permeabilities of the low-var and high-perm cases are simple multiplicative and additive transformations of the base case permeability random field.

Decreasing the correlation lengths of the geostatistical model leads to more pronounced local effects, see Fig. 8. This test case is likely the most challenging for the subset simulation method as the problem has 3701 dimensions - almost twice as many as the base case. Repeating the simulations led to similar shapes of the histograms, and the adaptive scaling of the proposal distributions vary only moderately (between 0.3 and 1.5 times the initial scaling with the empirical standard deviation of the seeds). The low-perm test case is a relatively simple variation of the base case, and the results are qualitatively similar as can be seen by comparing Fig. 9 to Figs. 6 and 7. Finally, it is interesting to compare the different outcomes when the orientation of the correlation structure is varied. Assuming equal correlation leads to an almost symmetric Gaussian-like distribution, see Fig. 10. A vertical correlation structure leads to a long tail of small-valued capacities of varying probability, as can be observed in Fig. 11. At the finest levels, the distributions appear poorly resolved with spurious peaks.

The percentiles estimated from the subset simulations are displayed in Table 3. To quantify the reliability of these numbers, confidence

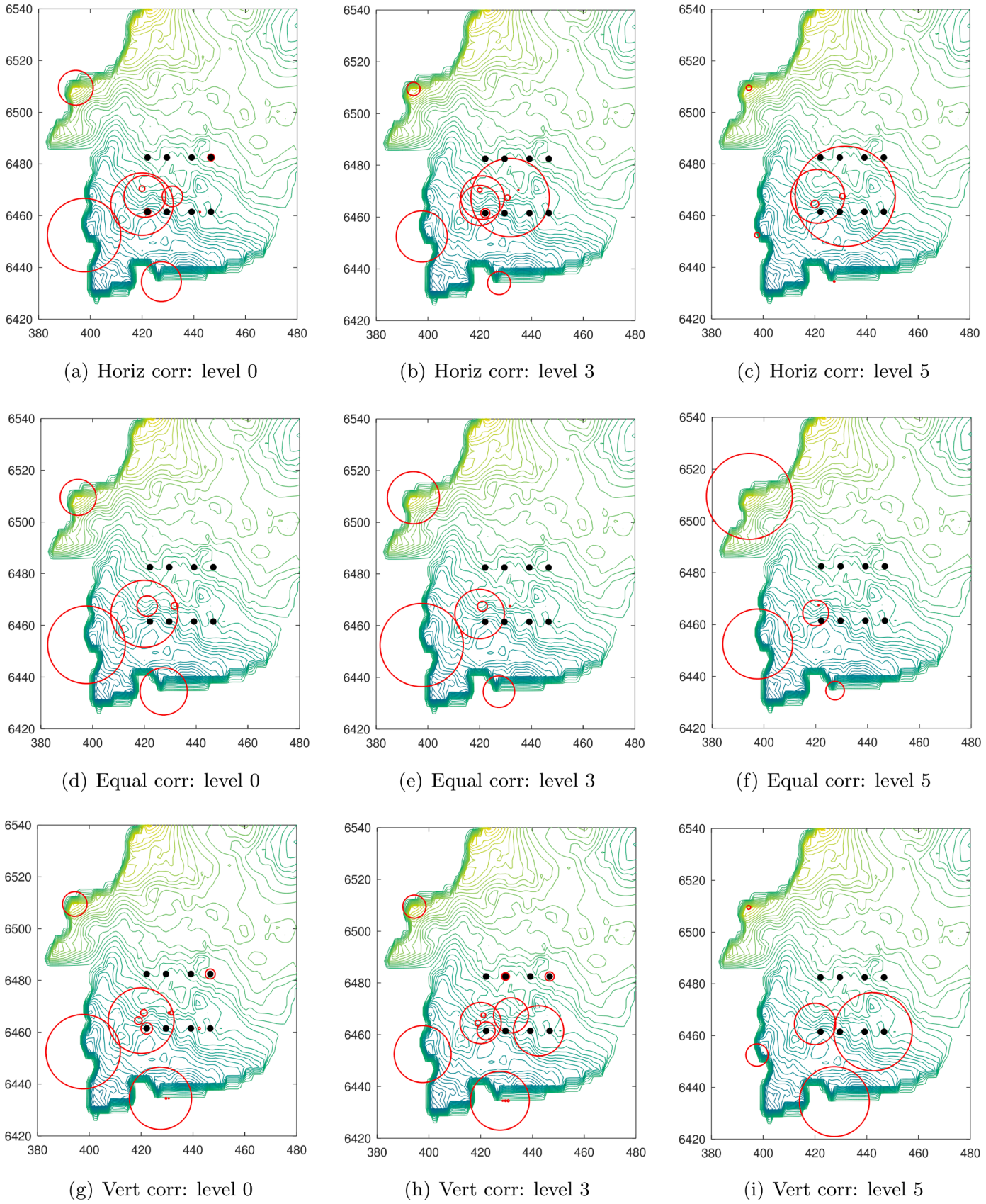


(a) Low capacity log(k) (b) Low capacity Δp (bar) (c) Low capacity % $\Delta p_{\max}$



(d) High capacity log(k) (e) High capacity Δp (bar) (f) High capacity % $\Delta p_{\max}$

Fig. 13. Comparison of permeability, overpressure, and percentage of maximum pressure for representative cases of high and low capacity.



(a) Horiz corr: level 0 (b) Horiz corr: level 3 (c) Horiz corr: level 5

(d) Equal corr: level 0 (e) Equal corr: level 3 (f) Equal corr: level 5

(g) Vert corr: level 0 (h) Vert corr: level 3 (i) Vert corr: level 5

Fig. 14. Bubble plot showing the distribution of pressure-limiting locations (red circles) for the base case as a function of subset simulation level (left to right) and correlation (top to bottom). Circle radius are scaled by the frequency of occurrence. The injection sites are given in dark filled circles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

intervals can be estimated for the quantiles. Let $F_Q(Q)$ be the unknown cumulative distribution function of the QoI Q at some level $\ell = 0, \cdots, L$. Then each sample $Q^{(n)}$ takes a value lower than $F_Q^{-1}(q)$ with probability q, i.e., it follows a binomial distribution $B(N_\ell, q)$. Hence, if the samples are ordered so that $Q_{(1)} \leq Q_{(2)} \leq \ldots \leq Q_{(N)}$, then $[Q_{(l_q)}, \ Q_{(u_q)}]$ is an $100(1 - \alpha)\%$ two-sided confidence interval for the q-quantile, where $l_q = B^{-1}(\alpha/2, N_\ell, q) - 1$ and $u_q = B^{-1}(1 - \alpha/2, N_\ell, q) + 1$. For the base case, a two-sided 95% confidence interval for the P10, P1, and P0.1 percentiles are respectively [5.959, 5.978], [5.726, 5.734], and [5.486, 5.501]. These are considered sufficiently narrow that the numbers in Table 3 are good approximations up to one digit. However, these numbers rely on the assumption that the samples are uncorrelated, which is not the case for subset simulation samples. Two alternative measures of accuracy will therefore be presented, both of which take empirical correlations into account.

An empirical measure of the accuracy of the computed percentiles is given by repeating the experiments with varying numbers of samples per level. The percentiles using 1000 and 5000 samples per level for the base case are also shown in Table 3, and the correpsonding histograms are displayed in Appendix A. These results also indicate that the results are good up to about a single decimal point.

Rather than repeating the experiments for all cases and percentiles, one may instead estimate so called efficiency factors based on the existing results using the framework presented in Au and Beck (2001); Papaioannou et al. (2015). The efficiency factor is the relative performance compared to the ideal case of completely uncorrelated sampling, and defined as

$$\text{eff}_\ell = \frac{1}{1 + \gamma_\ell}$$

where

$$\gamma_\ell = 2 \sum_{k=1}^{1/p-1} (1 - kp)\text{AC}_\ell(k), \qquad \text{for } \ell = 1, \cdots, L,$$

with $\text{AC}_\ell(k)$ being the $\text{lag} - k$ auto-correlation of the set $\{\mathbb{1}_{Q \leq Q_\ell}(Q^{(n,\ell)})\}_{n=1:N_\ell}$ averaged over all Markov chains of level ℓ. The efficiency factors are shown in Fig. 12. They are all comparable to those for the base case, and we conclude that the accuracy of all percentiles are of the same order.

4.1. Comparison of median to rare-event pressure maps

It is useful to make a comparison of the physical behavior of two cases that represent low and high capacity, respectively. For the low capacity case, we select one sample from the level 5 extreme region (P0.001), with estimated capacity of 5 Gt. For the high capacity case, we take a sample from the upper tail of the level 0 histogram, choosing a case where the estimated capacity is 7.5 Gt. Both representative cases are selected from the base case (1 D mean permeability and horizontal correlation).

We recall that the underlying geostatistics for each capacity case are identical, however the permeability map (Figs. 13(a) and 13(d)) shows a stark contrast in distribution. Both cases have correlated regions of lower and higher permeability that extend across the width of the domain, but the difference lies in the highs and lows. The low capacity permeability field varies between several moderately high and low permeability regions, while the high capacity permeability field is marked by a few focused regions of very high and low permeability.

The differences in the permeability field are evident in the maps of overpressure (difference between dynamic pressure and initial pressure) at the end of the simulation time, which is 57 years (low capacity) and 96 years (high capacity). In the low case (Fig. 13(b)), the more frequent but moderate low permeability regions inhibit pressure dissipation into the northern region of the domain. In the high peremability case (Fig. 13(e)), pressure can dissipate easily around the more severe but isolated low permeability regions.

The result of the differences in pressure dissipation is that pressure build-up becomes concentrated in the lower portion of the domain in the low capacity case (Fig. (c)), reaching the maximum allowable pressure in the region in between the injection sites. For the high permeability case (Fig. 13(f)), the pressure approaches the maximum pressure more evenly across the domain.

4.2. Location of pressure limitation

The capacity for each simulation is defined by the time at which the maximum pressure limit is reached. For the Utsira-based test case, the maximum allowable pressure is heterogeneous across the domain as a function of the formation top (see Fig. 3(c)). Thus, the capacity can be limited by locations in shallower regions rather than the location where the overpressure is highest, i.e. at the injection sites. Due to uncertainty in permeability, the location of the pressure limit can vary significantly from one sample to the next. For the base case, the simulated samples in the level 0 distribution (P50 quantile) indicate that many different locations across the lower Utsira can be controlling the capacity (Fig. 14(a)), ranging from near the injection sites to along the far boundaries. However, for the progressively zoomed in subsets of the rare-event region, we see that the location becomes increasingly concentrated in a few central locations with increasing levels. For level 5 (P0.001 quantile) shown in Fig. 14(c), there is essentially only one location that controls capacity, which is located just north of the lower row of injection sites. These results related to the base case correspond with the observations of pressure build-up made earlier regarding Fig. 13.

The observed trend for the predominantly horizontal correlation (base case) is also evident for other correlations. For the equally correlated case (Figs. 14(d)-14(f)), the shift from wide to narrow distribution is not as stark, but we observe an increasing occurrence of pressure limit towards the north-west and south-west corners. For the predominantly vertically correlated case, there is shift in level 3 towards the east (Fig. 14(h)) that reverses slightly in level 5 (Fig. 14(h)). This fluctuation is likely due to the orientation of the correlation perpendicular to the dip of formation and the two rows of injection. In any case, it is interesting to observe that although the three correlation cases appear similar at level 0 (P50), the level 5 results are quite different. This shows that the rare-event capacity is clearly more controlled by the type of heterogeneity than the median expected capacity.

5. Discussion

In this study, we applied a rare-event uncertainty quantification workflow to estimate critically low storage capacity with low probability of occurrence for a heterogeneous storage aquifer. A few interesting points of discussion can be made given the results of this study.

5.1. Extension to other constraints or quantities of interest

Our study quantifies storage capacity as constrained by pressure, where the upper limit on pressure is heterogeneously distributed across the study area. A pressure constraint on capacity is appropriate for basin-scale resource assessment for which the emphasis is placed on estimating maximum injected volumes over several injection sites. In this instance, pressure plays an important role due to pressure

interference between sites and/or with regional boundaries. Local site-scale constraints such as trapping or other leakage risk features such as faults would need to be handled separately.

The workflow applied in this study is not limited to estimating pressure-limited basin-scale capacity, and other performance metrics of CO_2 storage that depend on uncertainty in heterogeneity can be estimated with the same approach. For instance, it is straightforward to change the forward simulator shut-off criterion to other constraints of interest, including those related to CO_2 migration and trapping. For example, if capacity is constrained by CO_2 migration beyond a spill point or other defined boundary, then these constraints can be implemented in the forward simulator.

With regard to the uncertainty quantification approach, the QoI could be adjusted to estimate quantiles of trapped CO_2, fault reactivation or leaked CO_2, or any number of quantities whose importance will depend on the question being posed. However, it remains to be seen if the subset simulation method will perform as well for another QoI as we observed in this study for estimate of storage capacity. We observe that storage capacity is a relatively smooth and well behaved distribution with respect to geologic heterogeneity. A heavily skewed, bi-modal or other complicated distribution may require additional research for the subset simulation method to be reliable and efficient approach for quantification of rare events.

5.2. Implications for basin-scale storage assessment

This study has demonstrated a fast and reliable approach to quantifying critically low capacity that can be very valuable for application to real systems. Basin-scale systems are large and the uncertainty in geologic parameters is often significant. This implies that estimating very low quantiles such as P1, P0.1 or lower using standard Monte Carlo approaches may lead to extraordinary computational times. The targeted sampling approach presented here can be used to give comparable confidence in capacity estimates under geologic uncertainty with fewer forward simulations.

The concept of critical storage capacity adds increased understanding of the behavior of a system under geologic uncertainty. For the high-quality sand formation we studied with extensive thickness and mean permeability of 1 D, we found a high expected median capacity of several Gt and some degree of variability around the P50 value. However, while the impact of heterogeneity uncertainty is significant, one can still expect to achieve 75% of the P50 capacity for even the most rare capacity occurrences (P0.001). For a system with lower mean permeability of 0.5 D, the P50 capacity is lower due to the lower overall injectivity, but the P0.001 capacity is only 15% lower than the P50 value in this case. These results imply that the risk of capacity being so low that large-scale CO_2 storage would be no longer feasible is essentially zero.

This study gives an important contribution to improving understanding of how geological complexity affects pressure dissipation and thus capacity. We have found that the underlying geostatistics of the permeability field do not significantly affect the median P50 value for a given mean permeability. However, pinpointing parameters such as variance and correlations become increasingly important for the outer quantiles. Additional geological features not investigated here include regional faults, vertical pressure leak-off, geological boundaries, and stacked storage systems. Uncertainty in these features will have an important effect on quantifying pressure dissipation and capacity estimates. The subset simulation methodology can be applied to investigate additional geology, but will require a different approach to geostatistics modeling than applied herein. This topic is the subject of ongoing work.

Further work is needed to examine other types of storage systems not represented by the model system studied here. But our results indicate that the ultimate lower bound on capacity can still be an acceptably high value that would enable further appraisal of basin-scale storage resources. Our work also indicates that while pressure evolution is sensitive to heterogeneity, the very rare occurrences in heterogeneity do not severely impact the overall response of the system to multi-site injection. That said, quantification of the extremely low (or high) percentiles is still a necessary exercise to determine the outer bounds on capacity in a way that P10 and P90 percentiles do not fully capture.

The methods we present here can also be adapted to aid in designing effective pressure management strategies such as injection schedules and brine production. Optimizing strategies under uncertainty is a challenging task, and the targeted subset simulation approach can be combined with optimization algorithms to increase the efficiency of those workflows.

5.3. Geostatistical model

In this paper, we chose a relatively simple approach to geostatistical modeling by employing the assumption of Gaussian random fields. The advantage to this approach is the ability to parameterize the uncertainty in permeability by the Karhunen-Loeve expansion, which is necessary to achieve targeted sampling in the subset simulation step. However, there are many geological systems that require more complex approaches to heterogeneity modeling, which is often achieved by employing commercial earth modeling software. These packages are closed source and thereby will not provide the same opportunity for combining with the subset simulation approach described here. More work is needed to examine the underlying modeling in other types of geostatistical software, e.g. facies modeling, and to explore the potential to include these different modeling approaches with the rare-event methodology we have presented herein.

5.4. Numerical methods

Subset simulation is a versatile tool that scales well with the number of random dimensions, and can be applied to a very wide class of problems, including highly nonlinear and nonsmooth problems without any significant changes to the simulator setup. The problem investigated in this paper is a good example of that, featuring between 853 and 3701 random dimensions, and nonlinear input-to-output dependence. Compared to standard Monte Carlo sampling, the computational budget is greatly reduced: a total of 55,000 samples with 10,000 samples per levels (of which 1000 are the seeds taken from the previous level for all but the zeroth level). One would need approximately 10^9 standard Monte Carlo samples to expect as many as 10,000 samples to fall within the sample space defined by the most extreme region given by level 5. Unlike standard Monte Carlo, the subset simulation samples are not independent, so one should not come to the conclusion that the proposed method is $(1{\cdot}10^9)/55,000 \approx 18,000$ times more efficient, but the numbers still give a feeling for the possibilities of exploring otherwise unknown regions of stochastic space.

The VE simulation approach is an efficient way to achieve basin-scale simulations of multi-site CO_2 storage systems. The reduction in dimensionality from 3D to 2D is critical for achieving the computational requirements of uncertainty quantification even with the acceleration achieved by subset simulation. We also greatly simplified the CO_2 migration and trapping to only the essential features of compressible two-phase flow under drainage conditions, which is reasonable given the focus on pressure dissipation over very large scales where over 95% of the system is fully saturated with brine. With this set-up, each forward simulation takes approximately 1 minute to run on a standard laptop computer. By deploying a small computing cluster, we are able to perform each test case in a few hours. That said, one might prefer to

increase the complexity of the VE simulation to include hysteresis, capillary trapping and dissolution, and these are readily done given existing literature but the computational times will increase slightly. However, there are well-known limitations to the use of VE for CO_2 storage simulation in general. For detailed site-scale appraisal, a full 3D simulation could be desired. The workflow for critical capacity estimation presented herein is independent of the forward simulation, and the only requirements are a readable output file of the quantity of interest. However, it will likely be necessary to explore simplifications to reduce the computational burden when combining this approach with 3D flow simulation.

6. Summary and conclusions

In this study, we have applied rare-event methodology in order to estimate very unlikely but still relevant values of CO_2 storage capacity for heterogeneous basin-scale systems. The methodology presented herein involves a subset simulation approach for sampling a random permeability field within a rare-event region by using Markov chain Monte Carlo combined with conditional Karhunen-Loeve expansion to successively zoom in on increasingly low probability occurrences. This approach is highly efficient for quantifying very low probability events compared to application of standard Monte Carlo alone. This accelerated approach to uncertainty quantification is complemented by an efficient numerical solver suitable for basin-scale storage analysis, which gives a distinct advantage in maintaining acceptable computational cost.

We demonstrate our approach on a realistic basin-scale storage resource based on the Utsira Sand formation in the northern North Sea. We choose a set of test cases to evaluate the performance of the rare-event method for quantifying the increasing outer percentiles – P10, P1, P0.1, P0.01 and P0.001 – given a specified set of geostatistical parameters. The applied methodology was successful in resolving percentiles up to P0.001. In certain instances, non-monotonic behavior in the outer region is resolved well by the subset simulation method. Thus, we can conclude that the subset simulation method functions as expected for a realistic system. The method is also exceedingly efficient, performing accurate estimates of rare probability events for a given system in a couple of hours on a small computing cluster.

In all, seven sets of geostatistical parameters, including correlation length, orientation, variance and mean permeability, were tested in order to understand their impact on the estimated quantiles. We found that for a given expected permeability (basin-wide mean), the median behavior exhibits little variation for a variety of different choices of geostatistical parameters. The P50 percentile varies by only small margins, particularly compared with the variation for different mean permeability. However, the impact of the same underlying statistics increases for estimates of outer percentiles in the tail region. The controlling impact of geostatistics in the extreme region is supported by differences in pressure behavior at the large scale.

And finally, we have identified and quantified a set of increasingly rare occurrences of storage capacity for a realistic basin-scale system. The extreme instances capacity only occur for very rare realizations of the random permeability field, but their estimation is relevant for understanding the ultimate lower bound on capacity, however unlikely it may be. The actual values we estimated are highly dependent on many factors, and thus should not be taken as truth. However, we gained important insight that the quantified value of extreme-case capacity is greater than anticipated even for very rare occurrences of the permeability field. Our study indicates that rare occurrences of heterogeneity, that themselves cannot be ruled out, may not lead to a very severe reduction in basin-scale capacity. This result is promising for reducing risks of large-scale storage deployment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Financial support for this study has been provided by Gassnova to project no. 620073 through the CLIMIT-Demo funding program of the Research Council of Norway.

Appendix A. Base case with varying number of samples per level

Figs. A.15 and A.16 show the storage capacity samples from each level 0–5 for the base case using 1000 and 5000 samples per level, respectively.

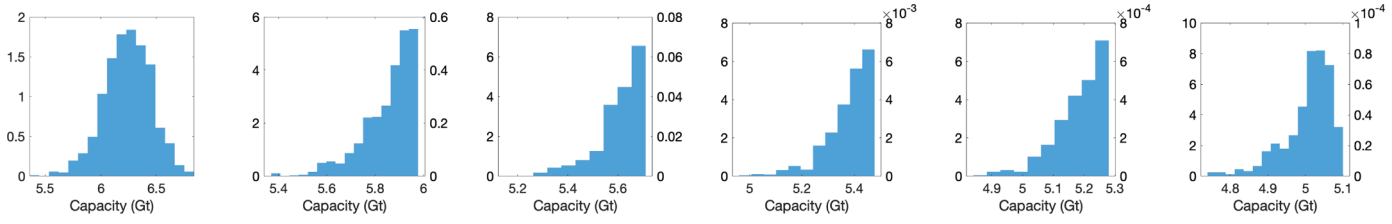


Fig. A1. Base case. Mean permeability 1 D. Histograms of samples of storage capacity on all levels using 1000 samples per level..

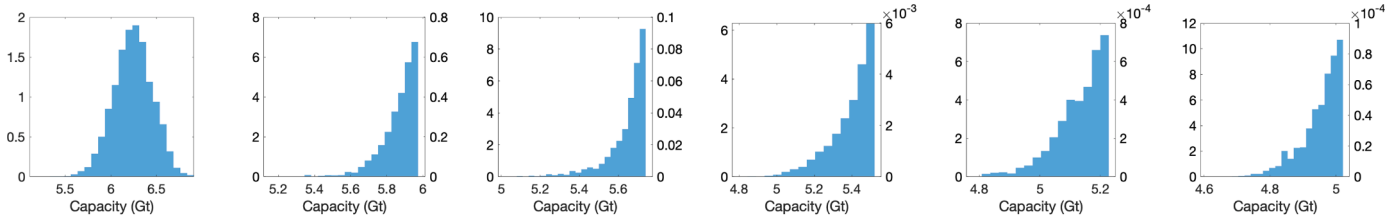


Fig. A2. Base case. Mean permeability 1 D. Histograms of samples of storage capacity on all levels using 5000 samples per level..

References

Anderson, S.T., Jahediesfanjani, H., 2019. Estimating the pressure-limited dynamic capacity and costs of basin-scale CO_2 storage in a saline formation. Int. J. Greenhouse Gas Control 88, 156–167.

Ashraf, M., 2014. Geological storage of CO_2: heterogeneity impact on the behavior of pressure. Int. J. Greenhouse Gas Control 28, 356–368.

Au, S.-K., Beck, J.L., 2001. Estimation of small failure probabilities in high dimensions by subset simulation. Probab. Eng. Mech. 16 (4), 263–277.

Barth, A., Schwab, C., Šukys, J., 2016. Multilevel Monte Carlo simulation of statistical solutions to the Navier–Stokes equations. In: Cools, R., Nuyens, D. (Eds.), Monte Carlo and Quasi-Monte Carlo Methods. Springer International Publishing, pp. 209–227.

Bentham, M., Mallows, T., Lowndes, J., Green, A., 2014. CO_2 storage evaluation database (CO_2 stored). the UK's online storage atlas. Energy Procedia 63, 5103–5113.12th International Conference on Greenhouse Gas Control Technologies, GHGT - 12

Birkholzer, J.T., Nicot, J.P., Oldenburg, C.M., Zhou, Q., Kraemer, S., Bandilla, K., 2011. Brine flow up a well caused by pressure perturbation from geologic carbon sequestration: static and dynamic evaluations. Int. J. Greenhouse Gas Control 5 (4), 850–861.

Birkholzer, J.T., Oldenburg, C.M., Zhou, Q., 2015. CO_2 Migration and pressure evolution in deep saline aquifers. Int. J. Greenhouse Gas Control 40 (SI), 203–220.

Birkholzer, J.T., Zhou, Q., 2009. Basin-scale hydrogeologic impacts of CO_2 storage: capacity and regulatory implications. Int. J. Greenhouse Gas Control 3 (6), 745–756.

Bohloli, B., Choi, J.C., Skurtveit, E., Grande, L., Park, J., Vannset, M., 2015. Criteria of fault geomechanical stability during a pressure build-up. NGI Report 2015/04.

Caflisch, R.E., 1998. Monte Carlo and quasi-Monte Carlo methods. Acta Numerica 7, 1–49.

Cao, C., Liao, J., Hou, Z., Wang, G., Feng, W., Fang, Y., 2020. Parametric uncertainty analysis for CO_2 sequestration based on distance correlation and support vector regression. J Nat Gas Sci Eng 77.

Class, H., Ebigbo, A., Helmig, R., Dahle, H.K., Nordbotten, J.M., Celia, M.A., Audigane, P., Darcis, M., Ennis-King, J., Fan, Y., Flemisch, B., Gasda, S.E., Jin, M., Krug, S., Labregere, D., Naderi Beni, A., Pawar, R.J., Sbai, A., Thomas, S.G., Trenty, L., Wei, L., 2009. A benchmark study on problems related to CO_2 storage in geologic formations. Comput. Geosci. 13 (4), 409.

Cliffe, K.A., Giles, M.B., Scheichl, R., Teckentrup, A.L., 2011. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. Comput. Visual Sci. 3 (14).

De Simone, S., Krevor, S., 2021. A tool for first order estimates and optimisation of dynamic storage resource capacity in saline aquifers. Int. J. Greenhouse Gas Control 106.

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. Greenhouse Gas Control 10, 397–418.

Dostert, P., Efendiev, Y., Mohanty, B., 2009. Efficient uncertainty quantification techniques in inverse problems for Richards equation using coarse-scale simulation models. Adv Water Resour 32 (3), 329–339.

Elenius, M., Skurtveit, E., Yarushina, V., Baig, I., Sundal, A., Wangen, M., Landschulze, K., Kaufmann, R., Choi, J.C., Hellevang, H., Podladchikov, Y., Aavatsmark, I., Gasda, S.E., 2018. Assessment of CO_2 storage capacity based on sparse data: Skade formation. Int. J. Greenhouse Gas Control 79, 252–271.

Elsheikh, A.H., Hoteit, I., Wheeler, M.F., 2014. Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. Comput Methods Appl Mech Eng 269, 515–537.

Elsheikh, A.H., Oladyshkin, S., Nowak, W., Christie, M., 2014. Estimating the probability of CO_2 leakage using rare event simulation. Conference Proceedings, ECMOR XIV - 14th European Conference on the Mathematics of Oil Recovery.

Frailey, S.M., Tucker, O., Koperna, G.J., 2017. The genesis of the CO_2 storage resources management system (SRMS). Energy Procedia 114, 4262–4269.13th International Conference on Greenhouse Gas Control Technologies, GHGT - 13, 14–18 November 2016, Lausanne, Switzerland

Gasda, S.E., Nordbotten, J.M., Celia, M.A., 2011. Vertically-averaged approaches for CO_2 migration with solubility trapping. Water Resour Res 47, W05528.

Gasda, S.E., Nordbotten, J.M., Celia, M.A., 2012. Application of simplified models to CO_2 migration and immobilization in large-scale geological systems. Int Journal of Greenhouse Gas Control 9, 72–84.

Gasda, S.E., Wangen, M., Bjørnarå, T.I., Elenius, M.T., 2017. Investigation of caprock integrity due to pressure build-up during high-volume injection into the Utsira formation. Energy Procedia 114, 3157–3166.

Ghanem, R., Spanos, P., 2002. Stochastic Finite Elements: A Spectral Approach. Dover New York.

Giles, M.B., 2015. Multilevel Monte Carlo methods. Acta Numerica 24, 259–328.

Gorecki, C.D., Ayash, S.C., Liu, G., Braunberger, J.R., Dotzenrod, N.W., 2015. A comparison of volumetric and dynamic CO_2 storage resource and efficiency in deep saline formations. Int. J. Greenhouse Gas Control 42, 213–225.

Gray, K., 2015. Carbon Storage Atlas, 5th. US Department of Energy, Office of Fossil Energy.

Gregory, A., Cotter, C.J., 2017. On the calibration of multilevel Monte Carlo ensemble forecasts. Q. J. R. Meteorolog. Soc. 143 (705), 1929–1935.

CO_2 Storage atlas: Norwegian continental shelf. In: Halland, E., Mujezinovi, J., Riis, F. (Eds.), 2014. Norwegian Petroleum Directorate.

Huntington, D.E., Lyrintzis, C.S., 1998. Improvements to and limitations of Latin hypercube sampling. Probab. Eng. Mech. 13 (4), 245–253.

Kirby, G.A., Chadwick, R.A., Holloway, S., 2001. Depth mapping and characterisation of the Utsira Sand Saline Aquifer, Northern North Sea. British Geological Survey Commissioned Report CR/01/218.

Köppel, M., Franzelin, F., Kröker, I., Oladyshkin, S., Santin, G., Wittwar, D., Barth, A., Haasdonk, B., Nowak, W., Pflüger, D., et al., 2019. Comparison of data-driven uncertainty quantification methods for a carbon dioxide storage benchmark scenario. Comput. Geosci. 23 (2), 339–354.

Le Maître, O.P., Knio, O.M., 2010. Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics. Springer Science & Business Media.

Lindeberg, E., Vuillaume, J.-F., Ghaderi, A., 2009. Determination of the CO_2 storage capacity of the Utsira formation. Energy Procedia 1 (1), 2777–2784.

Ma, X., Zabaras, N., 2009. An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method. Inverse Probl 25 (3), 035013.

Mondal, A., Mandal, A., 2020. Stratified random sampling for dependent inputs in Monte Carlo simulations from computer experiments. J Stat Plan Inference 205, 269–282.

Müller, F., Jenny, P., Meyer, D.W., 2011. Probabilistic collocation and Lagrangian sampling for advective tracer transport in randomly heterogeneous porous media. Adv Water Resour 34 (12), 1527–1538.

Masson-Delmotte, V., Zhai, P., Prtner, H.-O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Pan, C., Pidcock, R., Connors, S., Matthews, R., Chen, Y., Zhou, X., Gomis, M., Lonnoy, E., Maycock, T., Tignor, M., Tabatabaei, M., 2018. Global warming of 1.5C. an IPCC special report on the impacts of global warming of 1.5C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. IPCC Special report. IPCC.

Møll Nilsen, H., Lie, K.-A., Andersen, O., 2015. Analysis of CO_2 trapping capacities and long-term migration for geological formations in the Norwegian North Sea using MRST-co2lab. Computers & Geosciences 79, 15–26.

Najm, H.N., 2009. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. Annu. Rev. Fluid Mech. 41, 35–52.

Nordbotten, J.M., Celia, M.A., 2012. Geological Storage of CO_2. Wiley & Sons.

Oladyshkin, S., Class, H., Helmig, R., Nowak, W., 2011. A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. Adv Water Resour 34 (11), 1508–1518.

Page, B., Turan, G., Zapantis, A., Burrows, J., Consoli, C., Erikson, J., Havercroft, I., Kearns, D., Liu, H., Rassool, D., Tamme, E., 2020. The Global Status of CCS 2020: Vital to Achieve Net Zero. Hydrogen Knowledge Centre.

Papaioannou, I., Betz, W., Zwirglmaier, K., Straub, D., 2015. MCMC algorithms for subset simulation. Probab. Eng. Mech. 41, 89–103.

Pawar, R., Dempsey, D., Guthrie, G., 2017. Effect of permeability heterogeneity on area of review. Energy Procedia 114, 7459–7465.13th International Conference on Greenhouse Gas Control Technologies, GHGT - 13, 14–18 November 2016, Lausanne, Switzerland

Pettersson, P., 2016. Stochastic Galerkin formulations for CO_2 transport in aquifers: numerical solutions with uncertain material properties. Transp Porous Media 114 (2), 457–483.

Rasmussen, C.E., 2004. Gaussian Processes in Machine Learning. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 63–71.

Ringrose, P.S., Meckel, T.A., 2019. Maturing global CO_2 storage resources on offshore continental margins to achieve 2DS emissions reductions. Sci Rep 9.

Robert, C.P., Casella, G., Casella, G., 2004. Monte Carlo Statistical Methods, Vol. 2. Springer.

Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. Annals of Applied Probability 7 (1), 110–120.

Shaw, J., Kesserwani, G., Pettersson, P., 2020. Probabilistic Godunov-type hydrodynamic modelling under multiple uncertainties: robust wavelet-based formulations. Adv Water Resour 137, 103526.

Smith, P.J., Shafi, M., Gao, H., 1997. Quick simulation: a review of importance sampling techniques in communications systems. IEEE J. Sel. Areas Commun. 15 (4), 597–613.

Sochala, P., Chen, C., Dawson, C., Iskandarani, M., 2020. A polynomial chaos framework for probabilistic predictions of storm surge events. Comput. Geosci. 24, 109–128.

Span, R., Wagner, W., 1996. A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa. J. Phys. Chem. Ref. Data 25 (6), 1509–1596.

Vesovic, V., Wakeham, W.A., Olchowy, G.A., Sengers, J.V., Watson, J.T.R., Millat, J., 1990. The transport properties of carbon dioxide. J. Phys. Chem. Ref. Data 19 (3), 763–808.

Williams, J.D.O., Holloway, S., Williams, G.A., 2014. Pressure constraints on the CO_2 storage capacity of the saline water-bearing parts of the Bunter sandstone formation in the UK southern North Sea. Pet. Geosci. 20 (2), 155–167.

Williams, J.D.O., Jin, M., Bentham, M., Pickup, G.E., Hannis, S.D., Mackay, E.J., 2013. Modelling carbon dioxide storage within closed structures in the UK Bunter sandstone formation. Int. J. Greenhouse Gas Control 18, 38–50.

Wriedt, J., Deo, M., Han, W.S., Lepinski, J., 2014. A methodology for quantifying risk and likelihood of failure for carbon dioxide injection into deep saline reservoirs. Int. J. Greenhouse Gas Control 20, 196–211.
Zahasky, C., Krevor, S., 2020. Global geologic carbon storage requirements of climate change mitigation scenarios. Energy & Environmental Science 13 (6), 1561–1567.
Zhou, Q., Birkholzer, J.T., 2011. On scale and magnitude of pressure build-up induced by large-scale geologic storage of CO_2. Greenhouse Gases Sci. Technol. 1 (1), 11–20.
Zuev, K.M., 2013. Subset Simulation Method for Rare Event Estimation: An Introduction. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–25.
Zuev, K.M., Beck, J.L., Au, S.-K., Katafygiotis, L.S., 2012. Bayesian post-processor and other enhancements of subset simulation for estimating failure probabilities in high dimensions. Computers & Structures 92–93, 283–296.