

Article

# Managing Uncertainty in Geological CO<sub>2</sub> Storage Using Bayesian Evidential Learning

Amine Tadjer \*  and Reidar B. Bratvold

Department of Energy Resources, University of Stavanger, 4021 Stavanger, Norway; reidar.bratvold@uis.no

\* Correspondence: amine.tadjer@uis.no; Tel.: +33-758-468-202

**Abstract:** Carbon capture and storage (CCS) has been increasingly looking like a promising strategy to reduce CO<sub>2</sub> emissions and meet the Paris agreement's climate target. To ensure that CCS is safe and successful, an efficient monitoring program that will prevent storage reservoir leakage and drinking water contamination in groundwater aquifers must be implemented. However, geologic CO<sub>2</sub> sequestration (GCS) sites are not completely certain about the geological properties, which makes it difficult to predict the behavior of the injected gases, CO<sub>2</sub> brine leakage rates through wellbores, and CO<sub>2</sub> plume migration. Significant effort is required to observe how CO<sub>2</sub> behaves in reservoirs. A key question is: Will the CO<sub>2</sub> injection and storage behave as expected, and can we anticipate leakages? History matching of reservoir models can mitigate uncertainty towards a predictive strategy. It could prove challenging to develop a set of history matching models that preserve geological realism. A new Bayesian evidential learning (BEL) protocol for uncertainty quantification was released through literature, as an alternative to the model-space inversion in the history-matching approach. Consequently, an ensemble of previous geological models was developed using a prior distribution's Monte Carlo simulation, followed by direct forecasting (DF) for joint uncertainty quantification. The goal of this work is to use prior models to identify a statistical relationship between data prediction, ensemble models, and data variables, without any explicit model inversion. The paper also introduces a new DF implementation using an ensemble smoother and shows that the new implementation can make the computation more robust than the standard method. The Utsira saline aquifer west of Norway is used to exemplify BEL's ability to predict the CO<sub>2</sub> mass and leakages and improve decision support regarding CO<sub>2</sub> storage projects.



**Citation:** TADJER, A.; Bratvold, R.B. Managing Uncertainty in Geological CO<sub>2</sub> Storage Using Bayesian Evidential Learning. *Energies* **2021**, *14*, 1557. <https://doi.org/10.3390/en14061557>

Academic Editor: Muhammad Aziz and Adam Smoliński

Received: 8 February 2021

Accepted: 8 March 2021

Published: 11 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** uncertainty quantification; carbon storage; Bayesian evidential learning; data assimilation

## 1. Introduction

Carbon capture and sequestration, also known as carbon capture and storage (CCS), represents a unique potential strategy to minimize carbon dioxide (CO<sub>2</sub>) emissions into the atmosphere. It creates a pathway toward a neutral carbon balance, which cannot be achieved solely with a combination of energy efficiency and other forms of low carbon energy. However, it can be achieved if CCS is added as a routine technology to any process that uses fossil fuels. Thus far, geological reservoirs, such as depleted oil or gas fields, or deep saline aquifers, have been considered as appropriate geologic formations for storing CO<sub>2</sub> emissions at a depth of several thousand meters [1–3]. Saline aquifers provide large storage capacities, are broadly distributed geographically, and are more accessible to capture sites as they facilitate the entire CO<sub>2</sub> transport process [4]. Several projects from the pilot-to commercial-scale have been implemented worldwide [5,6]. Cumulative injection of CO<sub>2</sub> in some countries like the United States, Norway, and Canada, is as high as 220 million tons (Mt). The majority of this cumulative (about 75%) is associated with enhanced oil recovery operations [7], and estimates show that geological reservoirs can store between 8000 to 55,000 Gt of CO<sub>2</sub> [8], which is the capacity of over 200 years of current global CO<sub>2</sub> emissions.

However, uncertainties in geological models and rock properties affect the flow modeling and CO<sub>2</sub> storage capacities, mitigating the risk of CO<sub>2</sub> leakage and consequently the contamination of clean groundwater. Ensuring the CCS is safe and successful requires both the storage capacity and CO<sub>2</sub> plume migration estimation, because they are used to identifying the significant uncertainty present in geomodel parameters like porosity, permeability, and caprock elevation. Storage operations monitoring must ensure the CO<sub>2</sub> remains trapped within the reservoir after the injection has stopped.

The standard methods for quantifying uncertainty rely on the consideration of many plausible geological realizations (ensemble model) and quantification of the statistical measures of the ensemble parameters. Assessment of the static/volumetric capacity within a large ensemble model can be easily performed. However, creating highly resolved simulations for all members of a large model ensemble can quickly become computationally intractable, which can be solved either by reducing the number of members of the ensemble or accelerating the simulations required for acquiring each geomodel realization. The uncertainty of specific parameters has been discussed in several previous studies. For instance, Allen et al. [9] proposed a simplified method to investigate the causality and impact of uncertain parameters, including rock properties (permeability and porosity), fault transmissibility, top-surface elevation, and aquifer conditions in term of temperature and pressure in terms of both static trapping capacity and dynamic plume migration estimation.

Several studies have demonstrated the application of a data assimilation and optimization strategy for the minimization and mitigation of risks during CO<sub>2</sub> injections as well as the postinjection period at the storage sites. For instance, Dai et al. [10] introduced a method of analyzing data by employing the probabilistic collocation-based Kalman filter (PCKF) for the optimization of the surveillance operations within GCS projects. The method involves the development of surrogate models with the use of polynomial chaos expansions (PCE) that act as a replacement of the original flow model, followed by an assessment of the reduced variance of the field cumulative CO<sub>2</sub> leak by analyzing the data. Subsequently, a comparison of the data-worth values of each monitoring strategy is done in order to select an optimal monitoring operation scheme. Oladyshkin et al. [11] introduced a framework using polynomial chaos expansion (PCE) as well as bootstrap filters for the assimilation of the pressure data to reservoir models and quantifying the uncertainty reduction of the rate of CO<sub>2</sub> leak within the storage sites. Only three uncertainty parameters were considered: reservoir's permeability, reservoir's porosity, and wellbore's permeability. Sun and Nicot [12] and Sun et al. [13] utilized probability based collocations for the assessment of how detectable the CO<sub>2</sub> leaks were, with the use of the pressure data from the monitoring wells, for the heterogeneous aquifers that are not certain. Additionally, Chen et al. [14] introduced a method that focused on machine learning and filter-based data assimilations to create a CO<sub>2</sub> monitoring design, where one determines the optimal monitoring design by making a choice from the designs for the boosting of the model's ability to predict cumulative CO<sub>2</sub> leakage. Chen et al. [15] further introduced a framework that focuses on the ensemble smoother (ES) with multiple data assimilations (ES-MDA) and the geometric inflation factor (ES-MDA-GEO) to calibrate the reservoir model and monitor the data from storage sites to predict CO<sub>2</sub> migration or leakage detection. González-Nicolás et al. [16] made a comparison of the use of ES and the restarting of the ensemble Kalman filter (EnKF) algorithms to detect pathways of potential CO<sub>2</sub> leakage with the use of pressure data. In this vein, Cameron et al. [17] examined how pressure data works in a zone over the storage aquifer, identifying and quantifying potential leaks. It also performs CO<sub>2</sub> storage by using a particle swarm optimizing algorithm coupled with Karhunen–Loève representations porosities for model reductions, which detect the aquifer model that matches historical data. However, there are still conceptual and computational challenges associated with data assimilation and optimization procedure proposed in the previous listed methods, as generating a set of models properly conditioned to all historical data that preserve the geological realism is very challenging process. The limitations have been well-detailed in Olivier et al. [18], one issue is that of ensemble collapse, which may result in unrealistic un-

certainty and difficulty to coverage to the target distribution. Another practical limitation is to render these approaches relevant for a large variety of problems, such as different prior distributions, different forward models, etc. That also causes significant computational implementation challenges. This study attempts to introduce an alternative approach that can circumvent the different problems associated with model-based approaches.

Recently, several approaches have demonstrated that it is possible to provide the outcomes of subsurface models without the need for model updating and solving the inverse problem [19]. In relation to this context, Scheidt et al. [20] with Satija et al. [21] introduced a new protocol for making decisions under uncertainty called Bayesian evidential learning (BEL). Based on the description provided in [19,22]. BEL relies mainly on data, model, prediction, and decision under Bayesianism scientific methodologies. BEL is usually divided into six main stages: (1) Formulation and definition of the decision problem; (2) prior model definition and specification; (3) Monte Carlo simulation and falsification of the prior uncertainty models; (4) Global sensitivity; (5) Uncertainty reduction using data; (6) Posterior falsification and decision making. In step 5, one may opt for classical inversion or direct forecasting (DF) [23]. DF utilizes a combination of statistical learning techniques and the Monte-Carlo sampling method to ensure direct relationships between the data and the prediction variables. It should be noted that this method requires no completed explicit model inversions. This results in it being less expensive by a computational amount when compared to the standard inversion methods. Despite the applications being still lower in number, DF was successful in applying case studies related to oil reservoir management, groundwater resources, and geothermal energy problems [21–25]. For instance, Satija et al. [23] used DF to forecast the future reservoir performance by mapping prior predictions into low-dimensional canonical space and estimating the joint distributions of historical and forecast data through linear Gaussian regression; they conclude by stating that this method displayed uncertainty estimates for production forecasts that reasonably agreed with rejection sampling. Yin et al. [22] proposed an extended approach based on direct forecasting, called direct forecasting of sequential model decompositions, in which both geological model parameters and borehole data are used simultaneously. The posterior results displayed large reductions of uncertainty both spatially, through a geological model and using gas volume predictions. In the context of CO<sub>2</sub> storage, Sun and Durlofsky [26] introduced a DF method named data-space inversion (DSI) that quantifies the uncertainties of CO<sub>2</sub> plume locations throughout GCS, where the generation of posterior forecasts of CO<sub>2</sub> saturation distributions were through the simulation results of prior model realizations along with observable data. Notably, the generation of posterior geological models were not in the DSI method, unlike the traditional methods of assimilating data, which involved ensemble-based data assimilations.

In this work, our intended contribution is to demonstrate how BEL protocol can be used in designing an uncertainty reduction strategy in predictions and minimizing the risk of CO<sub>2</sub> leakages, facing various sources in uncertainty in terms of permeability, porosity, temperature, pressure, and caprock depth. Here, we will use a case study problem based on the Utsira sand reservoir, which is a saline aquifer located in the Norwegian continental shelf (NCS). This paper also makes a key contribution in extending the DF procedure through implementing ES-MDA [26] and demonstrating that the DF with ES-MDA [27] is more robust than the standard procedure proposed in Satija et al. [21]. It also provides appropriate posterior uncertainty quantification with results that can be compared to those of the methods proposed in Yin et al. [22]. The paper is structured in multiple sections. In the following section, we present BEL framework and the associated statistical methods used to quantify uncertainty. Then, the proposed methodology can be tested by implementing it in Utsira CO<sub>2</sub> storage site involving many uncertainties. Finally, we provide some concluding remarks and recommend possible future research directions.

## 2. Methods: Uncertainty Quantification Framework

In this section, we will introduce the BEL procedure for the data assimilation. The BEL procedure is based on a Bayesian formulation in the data space, aiming to sample the conditional/posterior distribution of the interest quantities (in our case, the distribution of CO<sub>2</sub> mass and leakage in the top layer at a future time). BEL can usually be divided into six main stages [22]: (1) Formulation and definition of the decision problem; (2) prior model definition and specification; (3) Monte Carlo simulation and falsification of the prior uncertainty models; (4) Global sensitivity; (5) Uncertainty reduction using data; (6) Posterior falsification and decision making. Since this paper presents a hypothetical (but realistic) case study problem, we will focus on steps 2, 3, and 5.

### 2.1. Prior Model Definition

The prior sampling aims to identify the possible range of model parameterization and probability distribution for each geological parameter. Let  $m$  refer to the vector of uncertain parameters of a reservoir model using a historical data variable (CO<sub>2</sub> saturation near wellbore region, etc.) as vector  $d$ . The forecast (quantity of CO<sub>2</sub> mass and CO<sub>2</sub> leakage) is represented by  $h$ . The nonlinear function of  $m$  through both observed and forecast data forward model is defined as:

$$d = G_d(m) \quad \text{and} \quad h = G_h(m) \quad (1)$$

The functions,  $G_d$  and  $G_h$ , are generated through the use of a reservoir simulator and by forwarding them to prior geological model realizations, we obtain a set of  $N$  samples of both data and forecast variables.

$$d = (d^1, d^2, d^3, \dots, d^N), \quad \text{and} \quad h = (h^1, h^2, h^3, \dots, h^N) \quad (2)$$

Note that we refer  $d_{obs}$  as the vector of observation and acquired data.

### 2.2. Prior Model Falsification

Once the prior samples (both historical and forecast data) are extracted, it is important to check whether the observed (reference) prior data can predict posterior distribution that appertains to the prior range. Otherwise, there is a risk that the prediction may be erroneous. If the prior model is false, suggesting data inconsistency, we must revise the prior data distribution herein to assess the prior model's quality and ability to predict the posterior data [19]. A statistical procedure based on Mahalanobis distance ( $MD$ ) [28] is used that handles high dimensional and different types of data measurements with the primary objective of detecting outliers and determining whether the prior model is false or not. The  $MD$  for each data variable realization  $d$  or  $d_{obs}$  can be computed as follows:

$$MD(d) = \sqrt{(d - \rho)\beta^{-1}((d - \rho))}, \quad \text{for } n = 1, 2, 3, \dots, N \quad (3)$$

Here  $\rho$ , and  $\beta$  are the mean and covariance of the data  $d$ . Assuming that the distribution of the data is multivariate Gaussian, the distribution of  $[MD(dn)]^2$  would be chi-squared  $\chi_d^2$ . We set the 95th percentile of the  $\chi_d^2$  as the tolerance threshold for the multivariate dimensional point  $d^n$ . If  $MD(d_{obs})$  falls outside the tolerance threshold ( $MD(d_{obs}) > MD(d^n)$ ), the  $d_{obs}$  would be regarded as outliers, and the prior model would be determined as false, as it would mean that it has a very small probability. It should also be noted that this method requires data distribution to be Gaussian; if it is not, other outlier detection techniques such as local outliers detection [29], isolation forest [30], and One-Class Support Vector Machines [31] are highly recommended.

Next, a machine learning dimension reduction method is applied (e.g., functional principal component analysis (FPCA) [32] and canonical functional component analysis (CFCA) [33]) are applied to generate reduced dimension vectors in canonical,  $d^c$  and  $h^c$ , where  $\text{dimension}(d^c) \ll \text{dimension}(d)$ ; and  $\text{dimension}(h^c) \ll \text{dimension}(h)$ .

### 2.3. Direct Forecasting

Direct forecasting (DF) is a prediction-focused analysis [21,23,25], the main objective is to build a statistical relationship between the observed data and prediction without resolving any geological model inversion problem. More precisely, the main idea behind DF is to make an estimation of the conditional distribution  $f(h|d)$  from the prior Monte-Carlo sampling. This conditional distribution can be used to generate posterior samples  $h$ . In this paper, we introduce the method proposed in Satija and Caers (2017) [21], which has been successfully used in a variety of previous studies [19,24,25]. This learning strategy depends on mapping the problem into a lower-dimensional space through bijective transformations using machine learning reduction techniques- principal component analysis (PCA) and canonical correlation analysis (CCA) to maximize linearity between both data and prediction variables and then fitting the data through multivariate Gaussian distribution. In the multivariate Gaussian, all the conditional distributions can be identified analytically and described as follows:

The linear relationship between data variables and forecast implies that:

$$d^c = G_h^c \quad (4)$$

$G$  is the linear coefficient that maps  $h^c$  to  $d^c$ . Then, the Gaussian likelihood model is formulated as:

$$L(h^c) = \exp\left(-\frac{1}{2}(Gh^c - d_{obs}^c)^T C_{d^c}^{-1}(Gh^c - d_{obs}^c)\right) \quad (5)$$

Here,  $C_{d^c}$  is the data covariance matrix of the canonical space. Since the prior and likelihood data are multivariate Gaussian, the posterior is as well Gaussian, and the posterior mean and covariance are easily computed using the standard methods [34]. With the likelihood and prior data and a linear model being multivariate Gaussian, the posterior distribution  $f(h^c|d_{obs}^c)$  is also multivariate Gaussian with mean  $\tilde{h}^c$  and covariance model  $\tilde{C}^H$  that has an analytics solution:

$$\tilde{h}^c = \bar{h}_{prior}^c + C_h G^T (G C_h G^T + C_{d^c} + C_T)^{-1} (d_{obs}^c + G \bar{h}_{prior}^c) \quad (6)$$

$$\tilde{C}_h = C_h - C_h G^T (G C_h G^T + C_{d^c} + C_T)^{-1} G C_h \quad (7)$$

where  $C_T$  is the error covariance that occurs as a result of the linear fitting. Thus, we can generate the posterior data by simply sampling from this multivariate Gaussian.

One key element of DF is the way a sufficient Monte-Carlo samples of size  $N$  are determined. Following the results of previous studies on hydrogeophysics [25], and on oil reservoirs [23], the range of the realizations size  $N$  is generally between 100 and 1000. DF can also be modified. Instead of using linear Gaussian, we can integrate the ensemble smoother with multiple data assimilation (ES-MDA).

### 2.4. Direct Forecasting-ES-MDA (DF-ES-MDA)

ES-MDA is an ensemble-based method introduced by Emerick and Reynolds in 2013 [27], as an alternative to the sequential data assimilation scheme of EnKF. ES-MDA has successfully improved the performance of history matching, and it is simple to implement. In its simplest form, the method employs the standard smoother analysis equation a predefined number of times along with the covariance matrix of the measured data error that is multiplied by a coefficient  $a$ . The coefficients must be selected in a way that the following equation is satisfied.

$$\sum_{k=1}^{N_a} \frac{1}{a_k} = 1 \quad (8)$$

Here,  $N_a$  is the number of times the analysis is repeated. The standard ES-MDA analysis that is applied to a vector of model parameters,  $m$ , can be written as:

$$m_i = m_i^b + C_{md}(C_{dd} + a_p C_D)^{-1}(d_{obs} - d_{sim,i}), \text{ for } i = 1 \dots \dots N \quad (9)$$

Here, the subscript  $i$  refers to the  $i$ th ensemble member,  $C_{md}$  is the cross-covariance matrix between the vector of model parameters  $m$  and predicted data  $d$ ;  $C_{dd}$  is the auto-covariance matrix of the predicted data  $d$ ;  $a_p$  is the coefficient that inflates  $C_D$ , the covariance matrix of the observed data measurement error;  $d_{obs}$  is the observation data perturbed by the inflated observed data measurement error;  $d_{sim}$  the simulated data based on the previous simulation models; and  $N$  is the ensemble size (i.e., number of reservoir models in the ensemble). Conventionally, the ensemble-based history matching simultaneously updates  $N$  reservoir models.  $C_{md}(C_{dd} + a_p C_D)^{-1}$  refers to Kalman gain  $K$  which is computed with regularization by SVD using 99.9% of the total energy in singular values. Refer to [27,35] for more information about the method.

In this work, we modified ES-MDA to generate samples of both historical and predicted data variables  $d = [d^c, h^c]^T$ , given that the vector of observations  $d_{obs}^c$  are considered after applying PCA and CCA. Thus, we can write the DF-ESMDA update equation as:

$$d_i^{k+1} = d_i^k + K^k(d_{obs}^c - d^{c,k}) \quad (10)$$

$$K^k = C_{hd^c}(C_{d^c d^c} + a_p C_e)^{-1} \quad (11)$$

$$C_{d^c d^c} = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (d_i^c - \bar{d}^c)(d_i^c - \bar{d}^c)^T \quad (12)$$

$$C_{dd^c} = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (d_i - \bar{d}^c)(d - \bar{d}^c)^T \quad (13)$$

Here,  $N_e$  is the number of components (selected dimension);  $K$  refers to the Kalman gain;  $C_{hd^c}$  is the cross-covariance matrix between the vector  $d$  and historical data; and  $C_{d^c d^c}$  is the auto covariance matrix historical data.

### 2.5. Direct Forecasting on a Sequential Model Decomposition (DF-SMD)

DF can also be extended by replacing the prediction variable  $h$  with geological model variable  $m$  (porosity, permeability, etc.) to update the geological model variables and to obtain  $f(m|d_{obs})$  without traditional iterative model inversions. We employ the same method of [22] which has been successfully applied to update geological uncertainty with borehole data. In case of a reservoir, the geological model  $m$  consists of structural model  $s$ , rock types  $r$ , petrophysical model  $p$ , and subsurface fluid distribution  $f$ ,  $m = (s, r, p, f)$ .

The prior model uncertainty is defined as the sequential decomposition of specific model variables. In order to condition these model variables to wellbore data, we propose the following direct forecasting equation in a sequential scheme:

$$\begin{aligned} f(m|d_{obs}) &= f(s, r, p, f|d_{obs}) = f(f|s_{posterior}, r_{posterior}, p_{posterior}, d_{obs}, f) \\ &= f(p|s_{posterior}, r_{posterior}, d_{obs}, p) \\ &= f(p|s_{posterior}, d_{obs}, p)f(s|d_{obs}, p) \end{aligned}$$

From the equation above, the joint uncertainty quantification is equivalent to a sequential uncertainty quantification. Furthermore, the uncertainty quantification of a model component is conditioned to the near wellbore region data and posterior models of the previous components. Unlike the standard DF of a sequential model decomposition technique, the posterior realizations  $p$  and prior realizations  $f$  will aid in determining a conditional distribution  $f(f|p_{posterior})$ ; subsequently, we assess this using near wellbore region observations  $d_{obs}$  of  $f$ .

Moreover, due to the high dimensionality of the model variables, distance-based generalized sensitivity analysis (DGSA) method [36,37] is performed to investigate the effect of model variables  $m$  on the data variables and select a subset with  $m$  parameters

that have the largest impact on the data variables. The main advantage of DGSA is that it does not require a functional form and it is easy to use and requires relatively low amount of simulations [19]. For more details, see [36–38].

### 2.6. Uncertainty Reduction Analysis

The uncertainty reduction analysis is considered as one of the important elements of decision assessment, and it is employed after the three methods of direct forecasting are completed. In this paper, we have conducted uncertainty reduction in two metric interests, including CO<sub>2</sub> mass and leakage.

Taking  $M_c$  as CO<sub>2</sub> mass and the prior probability density function (PDF) of  $M_c$  as  $P(M_c)$ .  $P(M_c)$  is evaluated based on the prior reservoir models. We refer to the number of uncertainty distributions  $U(P(M_c))$  in a CO<sub>2</sub> mass  $P(M_c)$  as the following:

$$U(P(M_c)) = P_{90}P(M_c) - P_{10}P(M_c) \quad (14)$$

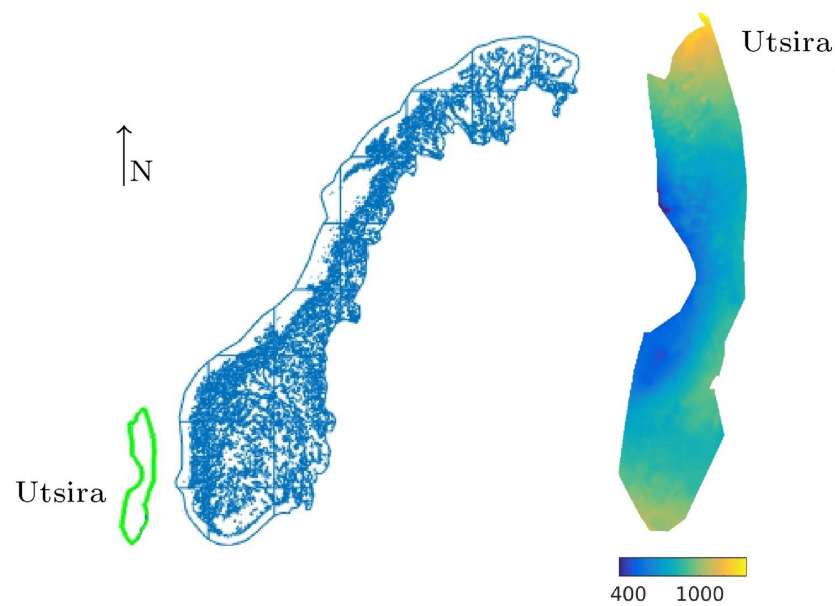
Here,  $P_{90}P(M_c)$  and  $P_{10}P(M_c)$  are the 90th and 10th percentiles, respectively. The posterior PDF of  $P(M_c|d_{obs})$  is computed using DF techniques. Therefore, the uncertainty reduction,  $U_R$ , is specified as the difference between prior uncertainty and posterior uncertainty in the model outcomes:

$$U_R = U(P(M_c)) - U(P(M_c|d_{obs})) \quad (15)$$

## 3. Materials

### 3.1. Model Description

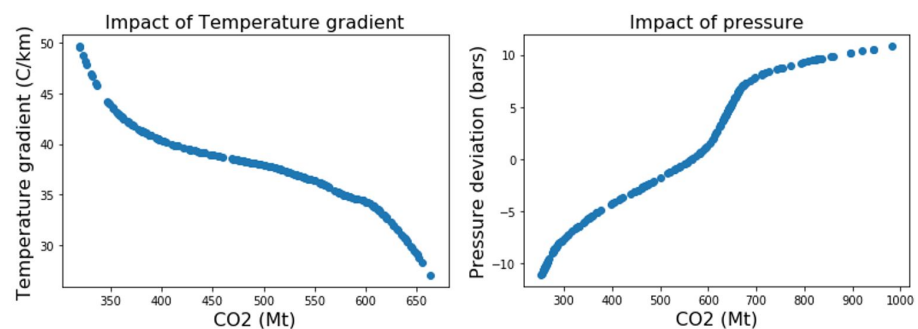
We test the performance level of the proposed methodology in the Utsira sand, which is a saline reservoir located beneath the central and northern North Sea as displayed in Figure 1. In this location, there are over 20 reservoir formations (producing oil and gas fields, abandoned oil and gas fields and geological formations such as saline aquifers). We simply use the reservoir dataset provided by the Norwegian Petroleum Directorate (NPD), which only consists of top-surface and thickness maps and average rock properties. Utsira formation consists of weakly consolidated sandstone with interlayered shale beds that act as baffles for the upward migration of the injected CO<sub>2</sub>, and it has an average top-surface depth of almost 800 m below the seabed (within the range of 300–1400 m). The storage capacity of the Utsira system is estimated to be 16 Gt, with a prospectivity of 0.5–1.5 Gt [39]. The boundaries of the aquifers are considered open. An open boundary means that there is communication between the aquifer and anything that lies adjacent to it, be it another aquifer or the sea bottom. The corresponding permeabilities in the Utsira geomodel range from 0.5 to 2.5 darcys. Another study Singh et al. [40] suggested that permeability could represent within the range of 1.1–5 darcys. Furthermore, in the NCS public datasets, there is no information about possible leakage through open boundaries or through the caprock. We acknowledge that these are important factors, but despite these limitations we have decided to use the Utsira available data to demonstrate BEL framework and discuss its advantages and potential benefits in future CCS operations. It is important to emphasize that in our study, some of the injected CO<sub>2</sub> can leave the computational domain during the simulation; these are considered as leaked volumes. Nonetheless, this cannot be the resulting CO<sub>2</sub> that has leaked back into the atmosphere; it will in most instances continue to migrate beyond the simulation model inside the rock volume.



**Figure 1.** Utsira formation. Location of along the Norwegian Continental Shelf (**left**). Maps of geomodel depths in meters (below the seabed) (**right**) [9].

### 3.2. The General Setup

A total number of  $N = 200$  prior geological realizations were generated using normal Gaussian distribution. There were uncertainties in terms of porosity, permeability, caprock elevation, temperature, and pressure. Following the case study by Nilsen et al. [3], which tested the sensitivity of  $\text{CO}_2$  migration to many input parameters, it was found that porosity differences would influence the total volume of rock that the plume comes into contact with. Increasing the thickness of the pore decreases the overall volume of rock occupied by the plume, reducing the migration so that the plume does not move far. Permeability impacts the behavior of  $\text{CO}_2$  plume flow by changing its speed and direction, creating a thinner plume that reaches further upslope. As shown in Figure 2, uncertain aquifer temperature and pressure may also affect the  $\text{CO}_2$  density, which further impacts plume migration and storage ability estimates.



**Figure 2.** Impact of pressure and temperature gradient in  $\text{CO}_2$  storage capacity.

Moreover, we assume that the Utsira reservoir has one injection well at 1012 m depth. Then, an injection rate of 10 Mt per year is considered for a period of 40 years, followed by a 3000-year migration (postinjection) period. Every flow simulation is performed by using the open-source software MRST- $\text{CO}_2$  lab developed by SINTEF [41], the Department of Applied Mathematics.  $\text{CO}_2$  lab Computational tools in MRST was specifically designed for studying the long-term and large-scale storage of  $\text{CO}_2$ .



## 4. Results and Analysis

### 4.1. Scenario 1: Uncertainty Reduction Using Direct Forecasting

**Prior Model.** A set of  $N = 200$  prior reservoir models is generated by using Monte-Carlo to the prior distributions. The selected number of models make certain that the prior distributions are adequately sampled. In all cases, a “reference” model, which is not incorporated in the set of  $N$  prior models, is considered. The prior models are forward modeled using a MRST-CO<sub>2</sub> lab over 3000 years. The CO<sub>2</sub> saturation data is collected at near wellbore region during the 40-year injection period. We intend to assess the quantity of CO<sub>2</sub> mass during the postinjection period and the corresponding CO<sub>2</sub> leakage at the end of time tracking period (3000 years). The prior distribution of modeled of the data variables for the injection well as well as the forecasts are shown in Figures 3 and 4. From both figures, we notice a large amount of uncertainties is involved.

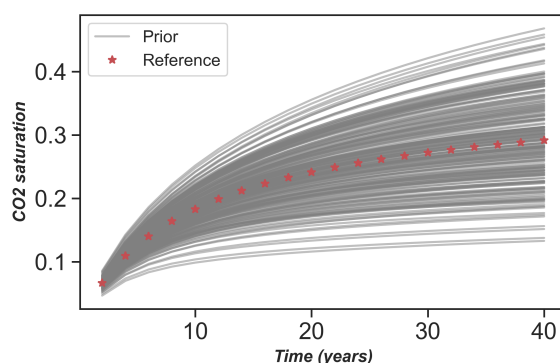


Figure 3. Prior measurement data variables.

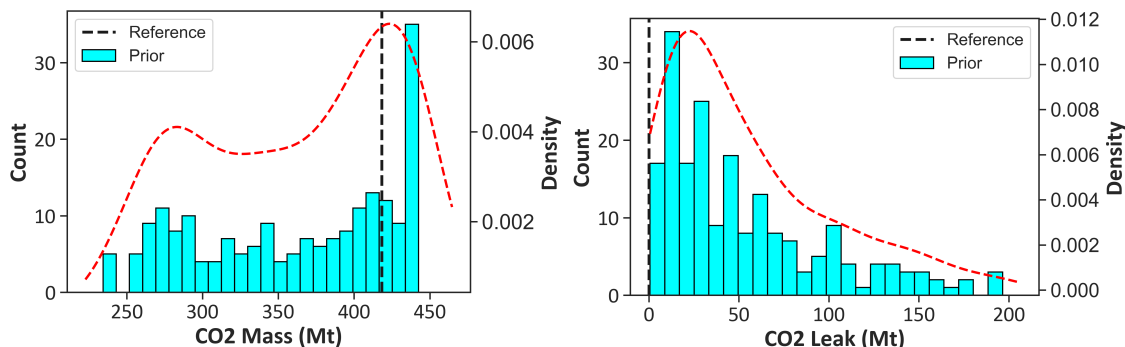
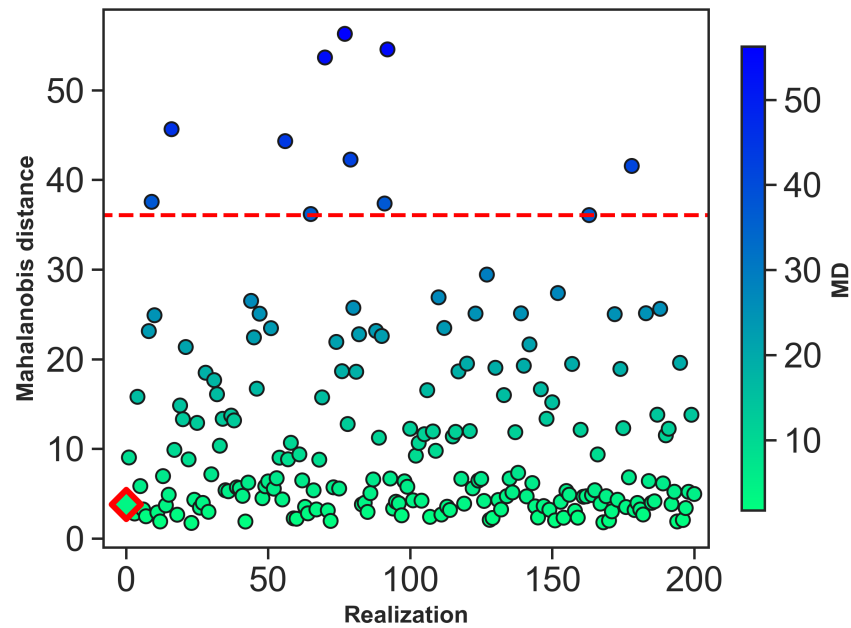


Figure 4. Prior distribution of prediction data variables—3000 years. The red dashed line is the prior probability density function.

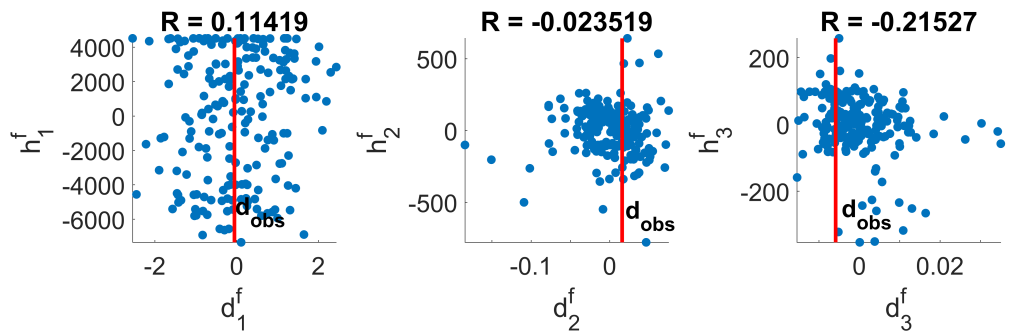
**Falsification.** To assess the quality of the prior models, data variables (CO<sub>2</sub> saturation) of the injection well of 200 prior models are used with  $d_{obs}$  by employing the MD outlier detection. The MD of  $d_{obs}$  is found to be 2.261, which is below the 95-percentile threshold, which suggests that the prior model is correct. Figure 5 shows the comparison of MD with  $d_{obs}$  and MD with 200 prior models.

**Dimension Reduction and Linearization.** To establish a relationship between the data and forecast variables, it is first necessary to ensure low dimensionality in both variables. For this purpose, we perform FPCA on the data variables  $d$  and  $h$  by selecting the principal components (PCs) that preserve 90 % variance. Accordingly, three dimensions are retained for both the data and forecast variables (CO<sub>2</sub> mass and CO<sub>2</sub> leak). The choice of the three dimensions is based on a compromise—it is important to keep as much variance as possible while ensuring maximum reduction of the problem’s dimensionality. Thereafter, CCA is conducted to the reduced data and prediction sets to maximize linearity between the reduced data and forecast. As shown in Figures 6 and 7, the relationship between the components in the functional domain is not linear; the application of CCA subsequently

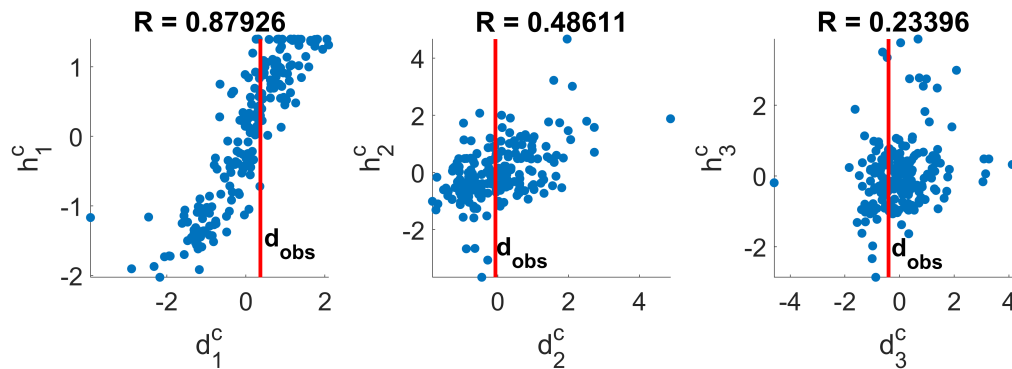
increases the correlation between the components in the canonical space, except the third dimension as displayed in Figure 6—CCA fails to establish a unique linear relationship.



**Figure 5.** Prior falsification using Mahalanobis Distance (MD). The red square is the MD for  $d_{obs}$ . Circle dots Refer the MD Results of 200 data variable samples, and the red dashed line is the 95th percentile of the Chi-Squared distributed MD.



(a) PCA correlation analysis.

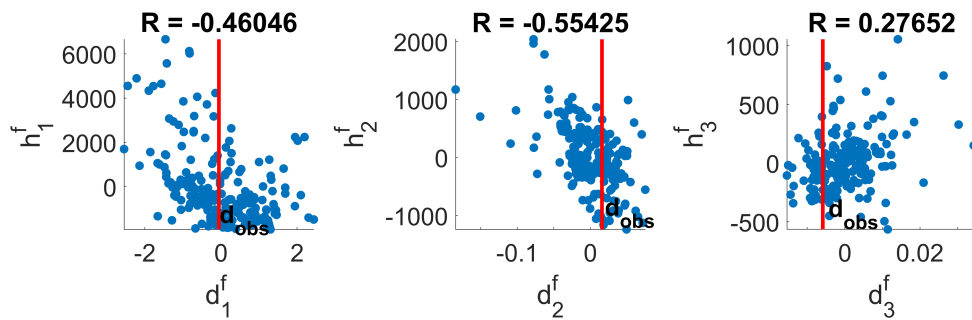


(b) PCA correlation analysis in canonical space.

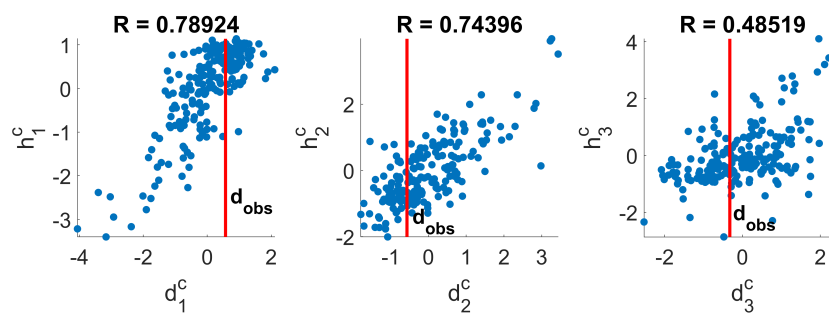
**Figure 6.** Functional components correlation analysis. Red lines correspond to the observed ( $CO_2$  mass).

**Reconstruct Posterior Model.** After a linear correlation in low dimensions has been established, we calculate the posterior distribution of the forecast components. First, we

use the linear Gaussian regression equation that has been explained in one of the previous sections, for which  $h^c$  must be first transformed using a normal score to get  $h_{gauss}^c$ . Thus, Gaussian regression generates a multivariate normal posterior  $f(h_{gauss}^c | d_{obs}^c)$  which can be easily used to sample forecast components that are conditioned to  $d_{obs}^c$ . Second, we apply modified ES-MDA explained in one of the previous sections to generate the posterior distribution of forecast variables  $h^c$ . Moreover, we used  $Na = 4MDA$  iterations. It must be noted that we have added random Gaussian noise to  $d_{obs}^c$ , with a mean of zero and a standard deviation of 10%.



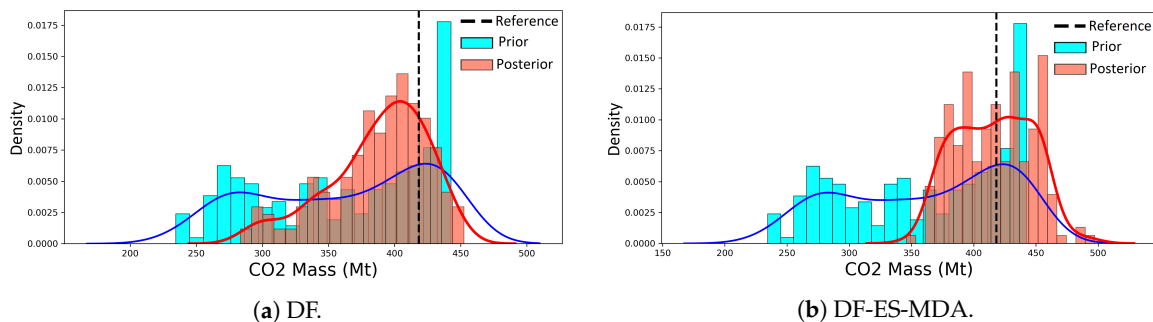
(a) PCA correlation analysis.



(b) PCA correlation analysis in canonical space.

**Figure 7.** Functional components correlation analysis. Red lines correspond to the observed (CO<sub>2</sub> Leak).

Once the posterior distribution of the prediction in the latent dimension space is established, it can be easily sampled and transformed back into the original initial space, where the posterior distribution of the prediction is shown in Figures 8 and 9; we notice that the DF with Gaussian regression techniques predicts a larger uncertainty range for both CO<sub>2</sub> mass and leakage after 3000 years compared to DF with ES-MDA, for which results are reasonable and data match is excellent, the uncertainty bands are reduced for both CO<sub>2</sub> mass and leakage at the end of 3000 years. The results stipulate that the proposed DF-ESMDA is more robust than the original DF. Both methods are fast in terms of computation, but they require running reservoir simulations of the prior ensemble, which definitely consumes a lot of the computational time.



(a) DF.

(b) DF-ES-MDA.

**Figure 8.** Reconstruct posterior CO<sub>2</sub> mass.

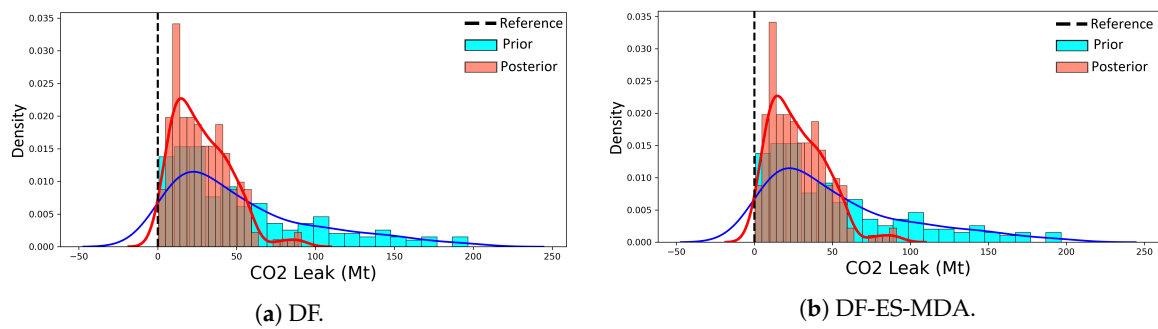


Figure 9. Reconstruct posterior CO<sub>2</sub> leak.

#### 4.2. Scenario 2: Uncertainty Reduction Using Direct Forecasting on a Sequential Model Decomposition

We use the same generated prior model used in Scenario 1, but as discussed in the previous section, we replace the prediction variable  $h$  with geological model variable  $m$  to obtain  $f(m|d_{obs})$ .

**Dimension Reduction.** We perform PCA on the model variable  $m$ , which consists of permeability, porosity, temperature, and pressure; we select the PCs that preserve 90% variance. As displayed in Figure 10, 102 dimensions are retained for both permeability and porosity, and 165 dimensions are kept for temperature and pressure, respectively.

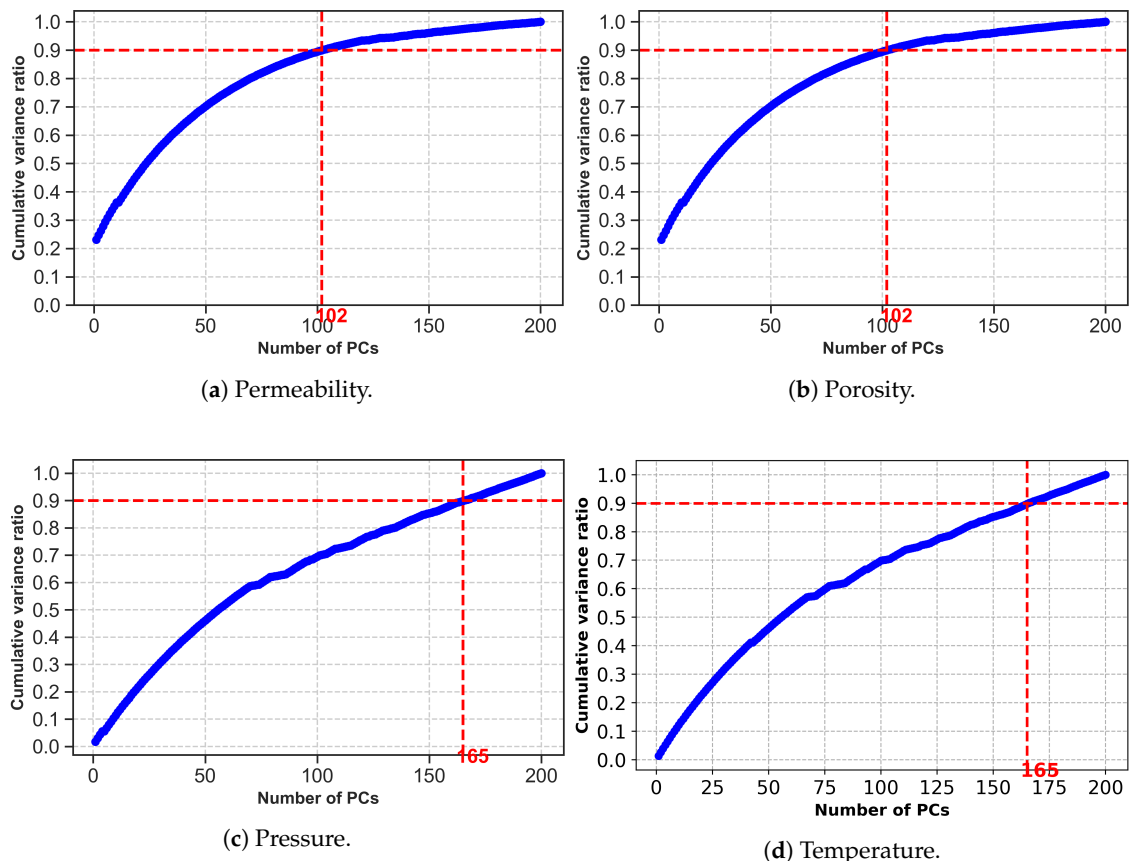


Figure 10. Cumulative sum of the PCA eigenvalues for the model  $m$  variables.

**Global Sensitivity Analysis.** In the next step, we would intend to find which PCA components impact the data prediction such that we can purpose a strategy for reducing the uncertainty of prediction variables. We apply DGSA based on a Euclidean distance to assess global sensitivity. Figure 11 outlines the main effects on a Pareto plot in which DGSA identifies the nonsensitive (measure of sensitivity < 1) and sensitive (measure of sensitivity > 1)

effects. In total, 18 sensitive principal components exist from the pressure spatial model, 22 for temperature, 11 for porosity, and 16 for permeability. These sensitive principal component and global variables scores are now assigned for uncertainty quantification.

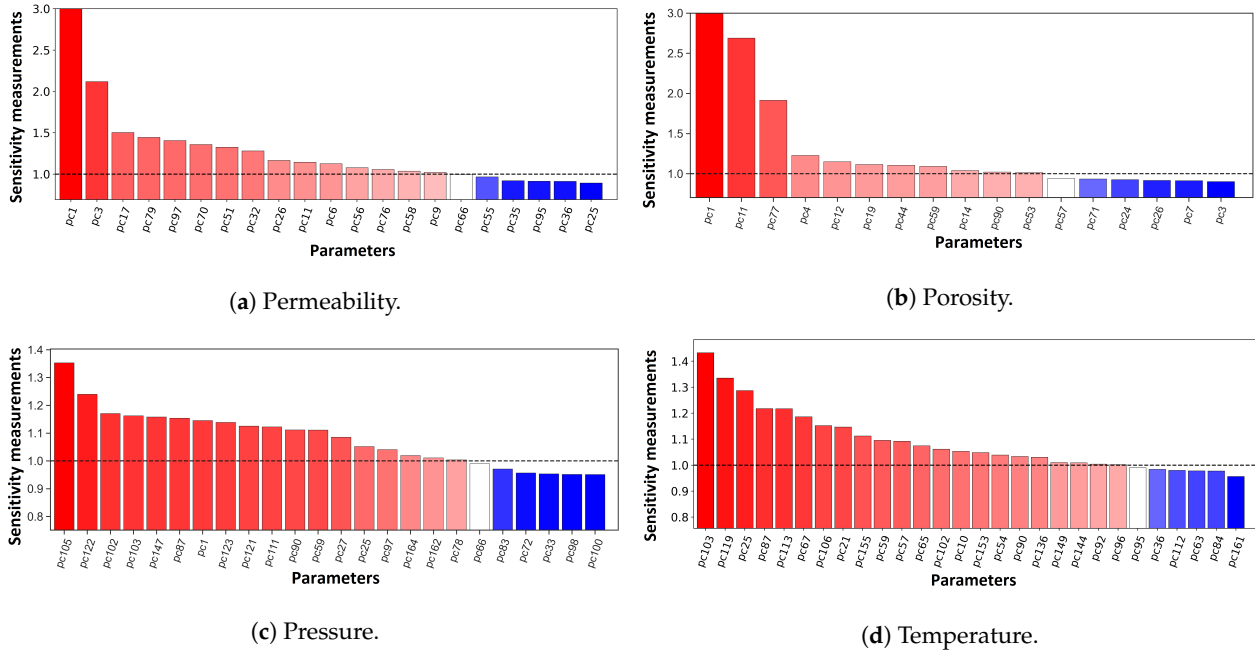


Figure 11. Global sensitivity of model parameters to measured data.

**Linearization.** After all sensitive model variables have been mapped into a lower-dimensional space, we require the application of CCA to establish a useful relationship between model variables and data variables. Figure 12 indicates that the primary canonical components  $d$  and  $m$  exhibit much stronger correlations.

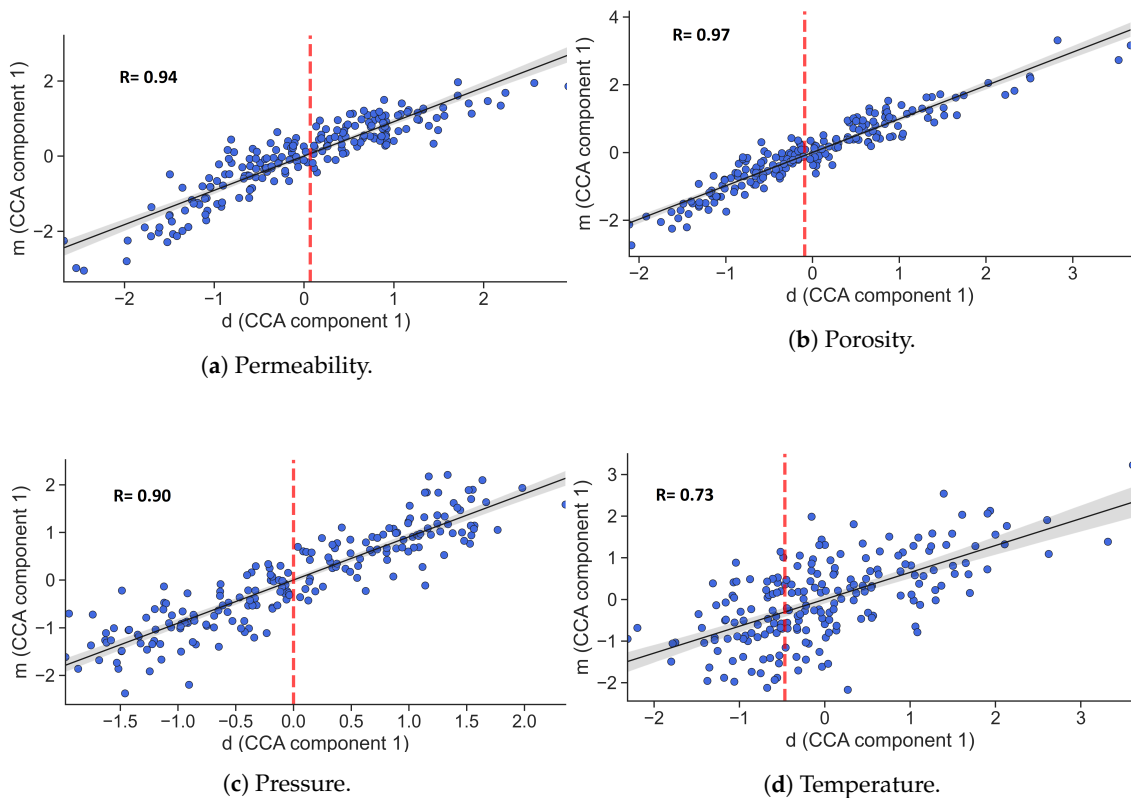
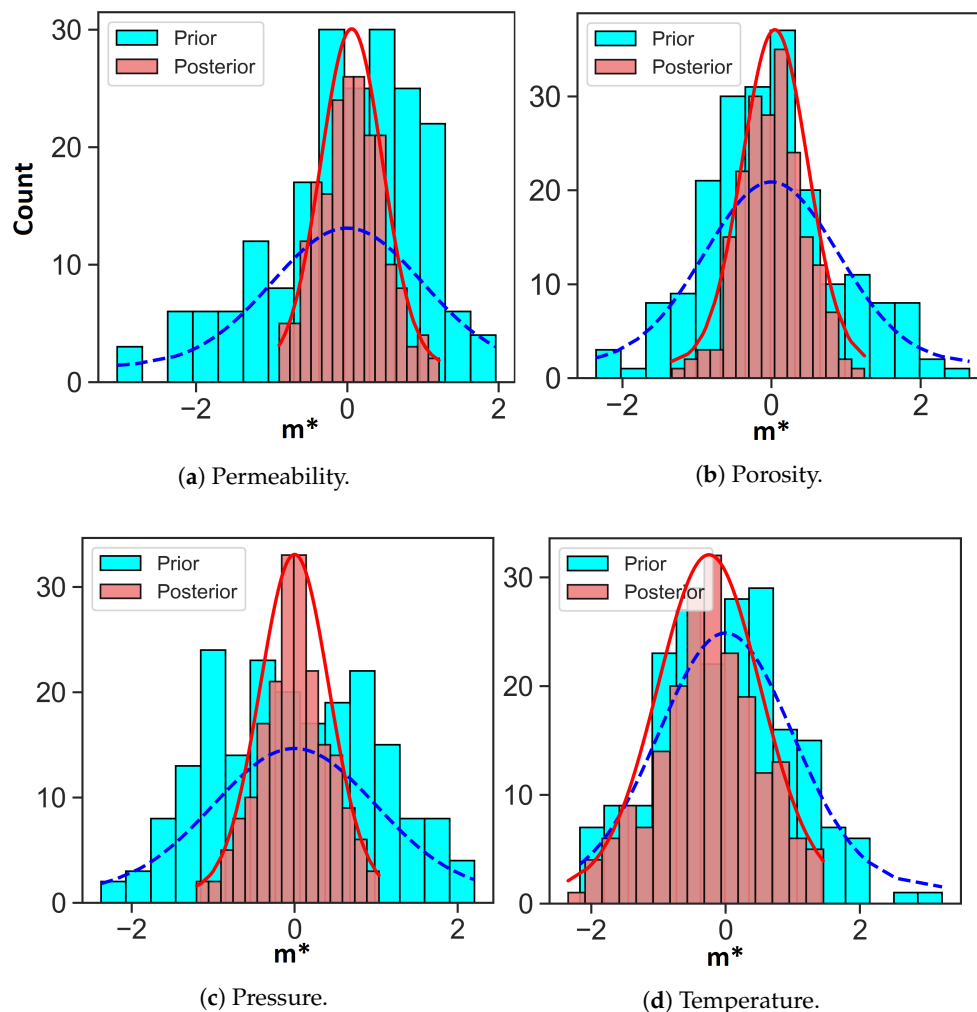


Figure 12. First canonical covariates of data and model variables. Red dashed lines correspond to the observed data.

**Reconstruct Posterior Model.** Once the linear correlation is maximized in low dimensions, it becomes easy to sample the posterior distribution and transform back lower-dimensional scores into original permeability, porosity, temperature, and pressure dimension scores. Figure 13 depicts the posterior distribution model realizations by comparing it to the following prior model, indicating that the model uncertainty range has reduced. We compare the score of both prior and posterior distribution along the two sensitive PCs with the highest score. From Figure 14, we notice that the prior samples' uncertainty has remarkably reduced. Note that the uncertainty quantification includes all the PCs sensitive score variables.

Figure 15 compares the Empirical CDF of the ensemble means of the sampled posterior log-perm, porosity, pressure, and temperature to their counterparts in the prior models. The results suggest a slight change on the distribution posterior model. Moreover, the uncertainty reduction is achieved, as the posterior samples are conditioned to the data variables of the well that are held in within the prediction domain. For verifying this results, we forward the posterior samples model  $m$  for simulation and extract the CO<sub>2</sub> mass and CO<sub>2</sub> leak posterior samples, and indeed, the posterior prediction distribution from evidential analysis accordingly reduces the uncertainty on the CO<sub>2</sub> mass and CO<sub>2</sub> leak as displayed in Figure 16; hence, this provides the input information required on the distribution of the data regarding CO<sub>2</sub> leakage at the end of migration tracking time (3000 years). From Table 1, it can be observed that integration of DF with ES-MDA would result in higher uncertainty reduction of CO<sub>2</sub> leakage (29.82–66.40 Mt) than the other techniques at both 40 years (after which we stop injection) and 3000 years.



**Figure 13.** Posterior and prior distributions of model variables (first canonical components).

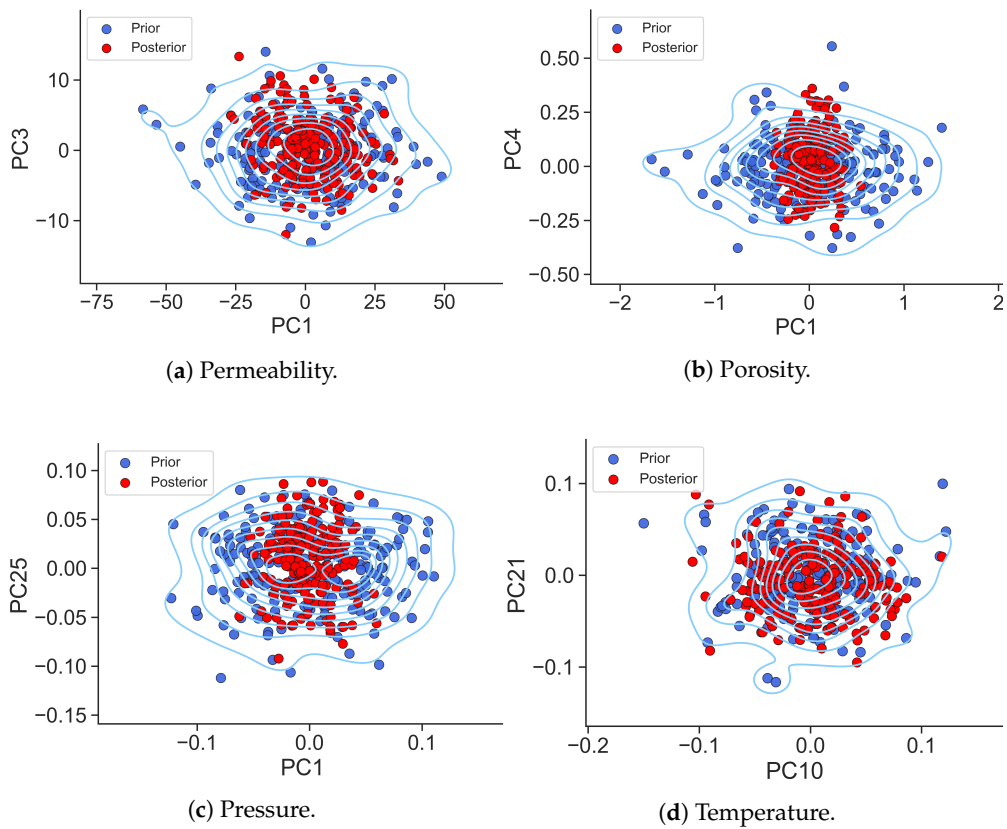


Figure 14. Prior and posterior distribution of the scores of the two sensitive PCs with the highest variances.

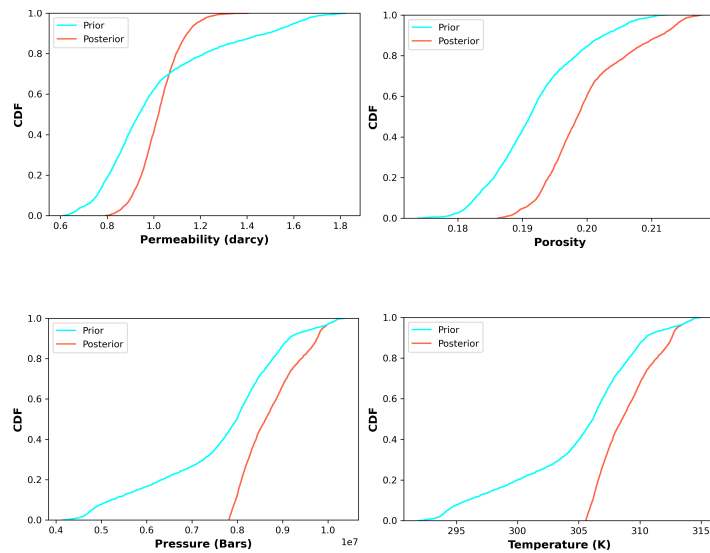
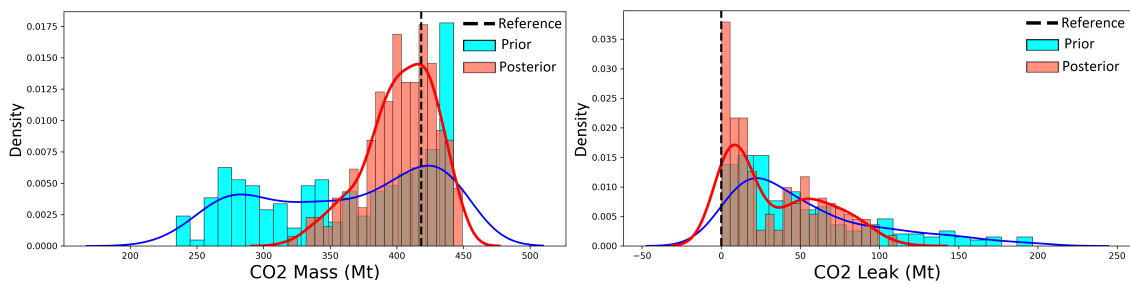


Figure 15. Empirical CDF computed from ensemble means of prior and posterior parameters.

Table 1. Uncertainty Reduction (UR) of CO<sub>2</sub> leak (Mt).

Methods	UR—40 Years	UR—3000 Years
DF	26.11	51.563
DF-ES-MDA	<b>29.82</b>	<b>66.40</b>
DF-SMD	28.35	56.83



**Figure 16.** Posterior distribution of CO<sub>2</sub> mass and leakage at 3000 years using DF-SMD. Black dashed lines correspond to the observed data.

## 5. Discussion and Concluding Remarks

This paper makes a contribution by showing a novel approach to quantify uncertainty during the injection of CO<sub>2</sub> for its storage and migration in deep saline aquifers by applying a Bayesian evidential learning (BEL) framework that involves falsification, global sensitivity analysis, and direct forecasting (DF). We presented a new DF implementation coupled with ES-MDA. The proposed DF-ES-MDA was compared with the original DF proposed in [21,23] and DF with sequential model decomposition in [22]. Both of the original methods mitigate the uncertainty reduction to a linear problem by reducing the high dimensionality of the original data using PCA and CCA; then, we established a statistical relationship between the data and forecast for DF and among the model, data, and forecast for DF-SMD. This estimated relationship combined with Bayesian Gaussian regression is thus used to generate a statistical forecast of the interest quantities—in our study, CO<sub>2</sub> mass and leakage. The new implementation preserves the main advantage of the original DF—its ability to provide an ensemble of CO<sub>2</sub> mass and leakage forecasts without iterative data inversion or history matching problems that can be computationally expensive and difficult. The three methods are advantageous even though the time to execute the reservoir simulations for the prior models tends to be time consuming. We compared the DF-ES-MDA with the original DF and DF-SMD of a real field case. Moreover, we showed that the accuracy of the DF-ES-MDA was consistently enhanced and a higher degree of uncertainty reduction could be achieved.

However, some criteria must be addressed to ensure the high-quality formulation of the three methods, in that a key for successful BEL framework application is the definition of the prior model, which should retain geological realism, as an unrealistic large uncertainty range may impact the data-prediction relationship and minimize accuracy. As such, a multivariate outlier detection method is employed to examine the quality of the prior model distribution compared to the observed case. Furthermore, the dimension reduction method should be selected based on the nature of the variable itself. Accordingly, we observed that FPCA was practical in our study for smoothly diversified time-series dataset (CO<sub>2</sub> saturation around near wellbore region), while eigen-image analysis proved useful in reducing the dimension of the spatial maps, such as permeability, porosity, etc. Moreover, PCA was mainly chosen as it is simple and bijective. Notably, multiple dimension reduction techniques, such as auto-encoder [42] and Gaussian process latent variable models (GPLVM) [43,44], can be included in the BEL framework. Additionally, the choice of regression technique is guided by the type, dimension, and relationship of the measurements, data, and forecast variables (linear or nonlinear). Due to the high-dimensionality problems, parametric regression is usually chosen instead of nonparametric techniques, except that large number of prior samples are available [19]. This work could be improved and extended in several ways. It is important to note that for this study, we have only considered quantities such as CO<sub>2</sub> saturation through wellbores and their respective CO<sub>2</sub> mass and leakage. This approach can be applied to examine the effectiveness of monitoring and the monitoring duration to lower uncertainty in risk metrics, such as top-layer CO<sub>2</sub> saturation and plume mobility and seismic time-lapse data. Accordingly, it will also be useful to apply the DF procedures to more complex geological models, such as bimodal



channelized systems, which can be challenging for traditional (model-based) history matching methods, kernel density estimation [45], and extensions of CCA [46] can be included in the BEL framework to tackle more complex nonlinear inverse problems. Finally, using data space inversion (DSI), as described by Sun and Durlofsky [26], CO<sub>2</sub> leakage detection under uncertainty should also be considered.

**Author Contributions:** A.T. wrote the paper and contributed to tuning the model and analyzing the results. R.B.B. supervised the work and provided continuous feedback. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research is funded by Petromaks-2 project DIGIRES (RCN no. 280473), and the APC was funded by the library at University of Stavanger, Norway.

**Acknowledgments:** The author acknowledges financial support from the Research Council of Norway through the Petromaks-2 project DIGIRES (RCN no. 280473) and the industrial partners AkerBP, Wintershall DEA, ENI, Petrobras, Equinor, Lundin, and Neptune Energy. The author would also like to thank Stanford center for reservoir forecasting for providing Auto-BEL implementation code.

**Data Availability Statement:** Not Applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Harp, D.R.; Stauffer, P.H.; O'Malley, D.; Jiao, Z.; Egenolf, E.P.; Miller, T.A.; Martinez, D.; Hunter, K.A.; Middleton, R.; Bielicki, J. Development of robust pressure management strategies for geologic CO<sub>2</sub> sequestration. *Int. J. Greenh. Gas Control* **2017**, *64*, 43–59. [[CrossRef](#)]
- Jin, L.; Hawthorne, S.; Sorensen, J.; Pekot, L.; Kurz, B.; Smith, S.; Heebink, L.; Hergegen, V.; Bosshart, N.; Torres, J.; et al. Advancing CO<sub>2</sub> enhanced oil recovery and storage in unconventional oil play—Experimental studies on Bakken shales. *Appl. Energy* **2017**, *208*, 171–183. [[CrossRef](#)]
- Nilsen, H.M.; Lie, K.A.; Andersen, O. Analysis of CO<sub>2</sub> trapping capacities and long-term migration for geological formations in the Norwegian North Sea using MRST-co2lab. *Comput. Geosci.* **2015**, *79*, 15–26. [[CrossRef](#)]
- IPCC. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2014; p. 151.
- Hosa, A.; Esentia, M.; Stewart, J.; Haszeldine, H. Injection of CO<sub>2</sub> into saline formations: Benchmarking worldwide projects. *Chem. Eng. Res. Des.* **2011**, *89*, 1855–1864. [[CrossRef](#)]
- Michael, K.; Golab, A.; Shulakova, V.; Ennis-King, J.; Allinson, G.; Sharma, S.; Aiken, T. Geological storage of CO<sub>2</sub> in saline aquifers—A review of the experience from existing storage operations. *Int. J. Greenh. Gas Control* **2017**, *4*, 659–667. [[CrossRef](#)]
- Institute for Global Change. *Global Status of CCS: 2018*; Institute for Global Change: London, UK, 2018.
- Jordan, K.; Gary, T.; Jeffrey, P.; Hans, T.; Haroon, K.; Yen-Heng, H.C.; Sergey, P.; Howard, H. Developing a Consistent Database for Regional Geologic CO<sub>2</sub> Storage Capacity Worldwide. *Energy Procedia* **2017**, *114*, 4697–4709.
- Allen, R.; Nilsen, H.; Lie, K.A.; O, M.; Andersen, O. Using simplified methods to explore the impact of parameter uncertainty on CO<sub>2</sub> storage estimates with application to the Norwegian Continental Shelf. *Int. J. Greenh. Gas Control* **2018**, *75*, 198–213. [[CrossRef](#)]
- Dai, C.; Li, H.; Zhang, D.; Xue, L. Efficient data-worth analysis for the selection of surveillance operation in a geologic CO<sub>2</sub> sequestration system. *Greenh. Gases Sci. Technol.* **2015**, *5*, 513–529. [[CrossRef](#)]
- Oladyshkin, S.; Class, H.; Nowak, W. Bayesian updating via bootstrap filtering combined with data-driven polynomial chaos expansions: Methodology and application to history matching for carbon dioxide storage in geological formations. *Comput. Geosci.* **2013**, *17*, 671–687. [[CrossRef](#)]
- Sun, A.Y.; Nicot, J. Inversion of pressure anomaly data for detecting leakage at geologic carbon sequestration sites. *Adv. Water Resour.* **2012**, *44*, 20–29. [[CrossRef](#)]
- Sun, A.Y.; Zeidouni, M.; Nicot, J.; Lu, Z.; Zhang, D. Assessing leakage detectability at geologic CO<sub>2</sub> sequestration sites using the probabilistic collocation method. *Adv. Water Resour.* **2013**, *56*, 49–60. [[CrossRef](#)]
- Chen, B.; Harp, D.R.; Lin, Y.; Keating, E.H.; Pawar, R.J. Geologic CO<sub>2</sub> sequestration monitoring design: A machine learning and uncertainty quantification based approach. *Appl. Energy* **2018**, *225*, 332–345. [[CrossRef](#)]
- Chen, B.; Harp, D.R.; Lin, Y.; Lu, Z.; Pawar, R.J. Reducing uncertainty in geologic CO<sub>2</sub> sequestration risk assessment by assimilating monitoring data. *Int. J. Greenh. Gas Control* **2020**, *94*, 102926. [[CrossRef](#)]
- González-Nicolás, A.; Baù, D.; Alzraiee, A. Detection of potential leakage pathways from geological carbon storage by fluid pressure data assimilation. *Adv. Water Resour.* **2015**, *86*, 366–384. [[CrossRef](#)]
- Cameron, D.A.; Durlofsky, L.J.; Benson, S.M. Use of above-zone pressure data to locate and quantify leaks during carbon storage operations. *Int. J. Greenh. Gas Control* **2016**, *52*, 32–43. [[CrossRef](#)]

18. Oliver, D.S.; Chen, Y. Recent progress on reservoir history matching: A review. *Comput. Geosci.* **2011**, *15*, 185–221. [[CrossRef](#)]
19. Scheidt, C.; Li, L.; Caers, J. *Quantifying Uncertainty in Subsurface Systems*, 1st ed.; John Wiley & Sons: Hoboken, NJ, USA, 2018.
20. Scheidt, C.; Renard, P.; Caers, J. Prediction-focused subsurface modeling: Investigating the need for accuracy in flow-based inverse modeling. *Math. Geosci.* **2015**, *47*, 173–191. [[CrossRef](#)]
21. Satija, A.; Caers, J. Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. *Adv. Water Resour.* **2015**, *77*, 69–81. [[CrossRef](#)]
22. Yin, Z.; Strebelle, S.; Caers, J. Automated Monte Carlo-based Quantification and Updating of Geological Uncertainty with Borehole Data (AutoBEL v1.0). *Geosci. Model. Dev.* **2019**, *13*, 651–672. [[CrossRef](#)]
23. Satija, A.; Scheidt, C.; Li, L.; Caers, J. Direct forecasting of reservoir performance using production data without history matching. *Comput. Geosci.* **2017**, *21*, 315–333. [[CrossRef](#)]
24. Athens, N.D.; Caers, J. A Monte Carlo-based framework for assessing the value of information and development risk in geothermal exploration. *Appl. Energy* **2019**, *256*, 113932. [[CrossRef](#)]
25. Hermans, T.; Lesparre, N.; De Schepper, G.; Robert, T. Bayesian evidential learning: A field validation using push-pull tests. *Hydrogeol. J.* **2019**, *27*, 1661–1672. [[CrossRef](#)]
26. Sun, W.; Durloufsky, L.J. Data-space approaches for uncertainty quantification of CO<sub>2</sub> plume location in geological carbon storage. *Adv. Water Resour.* **2019**, *123*, 234–255. [[CrossRef](#)]
27. Emerick, A.A.; Reynolds, A.C. Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **2016**, *55*, 3–15. [[CrossRef](#)]
28. De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D.L. The Mahalanobis distance. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 1–18. [[CrossRef](#)]
29. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying Density-Based Local Outliers. *Assoc. Comput. Mach.* **2000**, *29*, 2.
30. Liu, F.T.; Ting, K.M.; Zhou, Z. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422.
31. Schölkopf, B.; Williamson, R.; Smola, A.; Shawe-Taylor, J.; Platt, J. Support Vector Method for Novelty Detection. In Proceedings of the 12th International Conference on Neural Information Processing Systems, NIPS'99, Cambridge, MA, USA, 29 November–4 December 1999; pp. 582–588.
32. Springer. *Principal Component Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2002.
33. Hardoon, D.R.; Szedmak, S.; Shawe-Taylor, S. Canonical Correlation Analysis: An Overview with Application to Learning Methods. *Neural Comput.* **2004**, *16*, 2639–2664. [[CrossRef](#)] [[PubMed](#)]
34. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*, 1st ed.; SIAM: New Orleans, LA, USA, 2005.
35. Emerick, A.A. Analysis of the performance of ensemble-based assimilation of production and seismic data. *J. Pet. Sci. Eng.* **2016**, *139*, 219–239. [[CrossRef](#)]
36. Fenwick, D.; Scheidt, C.; Caers, J. Quantifying Asymmetric Parameter Interactions in Sensitivity Analysis: Application to Reservoir Modeling. *Math. Geosci.* **2014**, *46*, 493–511. [[CrossRef](#)]
37. Park, J.; Yang, G.; Satija, A.; Scheidt, C.; Caers, J. DGSA: A Matlab toolbox for distance-based generalized sensitivity analysis of geoscientific computer experiments. *Comput. Geosci.* **2016**, *97*, 15–29. [[CrossRef](#)]
38. Spear, R.C.; Hornberger, G.M. Eutrophication in peel inlet-II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Res.* **1980**, *14*, 43–49. [[CrossRef](#)]
39. Andersen, O.; Nilsen, H.M.; Lie, K.A. Reexamining CO<sub>2</sub> Storage Capacity and Utilization of the Utsira Formation. In Proceedings of the ECMOR XIV-14th European Conference on the Mathematics of Oil Recovery, Catania, Italy, 8–11 September 2014; pp. 1–18.
40. Singh, V.; Cavanagh, A.; Hansen, H.; Nazarian, B.; Iding, M.; Ringrose, P. Reservoir modeling of CO<sub>2</sub> plume behavior calibrated against monitoring data from Sleipner, Norway. In Proceedings of the SPE Annual Technical Conference and Exhibition, Florence, Italy, 19–22 September 2010.
41. Sintef. MRST-co2lab. Available online: <https://www.sintef.no/projectweb/mrst/modules/co2lab/> (accessed on 6 February 2021).
42. Wang, W.; Huang, Y.; Wang, Y.; Wang, L. Generalized Autoencoder: A Neural Network Framework for Dimensionality Reduction. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 496–503.
43. Lawrence, N.D. Learning for larger datasets with the Gaussian process latent variable model. In Proceedings of the Eleventh International Workshop on Artificial Intelligence and Statistics, San Juan, Puerto Rico, 21–24 March 2007.
44. Lawrence, N.D. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *J. Mach. Learn. Res.* **2005**, *6*, 1783–1816.
45. Lopez-Alvis, J.; Hermans, T.; Nguyen, F. A cross-validation framework to extract data features for reducing structural uncertainty in subsurface heterogeneity. *Adv. Water Resour.* **2019**, *133*, 103427. [[CrossRef](#)]
46. Lai, P.; Fyfe, C. A neural implementation of canonical correlation analysis, Neural Networks. *J. Mach. Learn. Res.* **1999**, *12*, 1391–1397.