# Quantitative evaluation of the joint effect of uncertain parameters in CO2 storage in the Sleipner project, using data-driven models

Ahmadinia, M., Shariatipour, S. M., Andersen, O. & Nobakht, B.



**Original citation & hyperlink:**
Ahmadinia, M, Shariatipour, SM, Andersen, O & Nobakht, B 2020, 'Quantitative evaluation of the joint effect of uncertain parameters in CO2 storage in the Sleipner project, using data-driven models', International Journal of Greenhouse Gas Control, vol. 103, 103180. https://dx.doi.org/ 10.1016/j.ijggc.2020.103180

DOI 10.1016/j.ijggc.2020.103180
ISSN 1750-5836

Publisher: Elsevier

# Quantitative evaluation of the joint effect of uncertain parameters in $CO_2$ storage in the Sleipner project, using data-driven models

Masoud Ahmadinia[1],*, Seyed M. Shariatipour[1], Odd Andersen[2], Behzad Nobakht[3]

[1] Centre for Fluid and Complex Systems, Coventry University, UK

[2] SINTEF Digital, Mathematics and Cybernetics Department, Oslo, Norway

[3] TÜV SÜD National Engineering Laboratory, Glasgow, UK

* Correspondence: Ahmadinm@uni.coventry.ac.uk

**Abstract**

Several researchers have studied the Sleipner model to understand the inherent flow physics better, to find a satisfactory match of the $CO_2$ plume migration. Various sources of uncertainty in the geological model and the fluid have been investigated. Most of the work undertaken on the Sleipner model employed the one factor at a time (OFAT) method and analysed the impact of uncertain parameters on plume match individually. In this study, we have investigated the impact of some of the most cited sources of uncertainties including porosity, permeability, caprock elevation, reservoir temperature, reservoir pressure and injection rate on $CO_2$ plume migration and structural tapping in the Sleipner. We tried to fully span the uncertainty space on Sleipner 2019 Benchmark (Layer 9) using a vertical-equilibrium based simulator. To the best of our knowledge, this is the first time that a study has focused on the joint effect of six

uncertain parameters using data-driven models. This work would raise our scientific understanding of the complexity of the impact of the reservoir uncertainty on $CO_2$ plume migration in a real field model. The caprock elevation was shown to be the most important parameter in controlling the plume migration (overall importance of 26%) followed by injection rate (24%), temperature (22%), heterogeneity in permeability (13%), pressure (9%) and porosity (6%).

## 1. Introduction

Global warming, which is primarily caused by the rapid increase in carbon dioxide ($CO_2$) emission, is one of the major issues of our time (Onoja et al., 2019). One of the proposed solutions to tackle this problem is carbon capture, utilisation and storage (CCUS) which has the potential to decrease greenhouse emissions up to 85% by 2050 (Edenhofer, 2015; IEA, 2012; Viebahn, Vallentin, & Höller, 2015). The best formations for the safe long-term storage of $CO_2$ are usually sedimentary rocks with appropriate porosity and permeability to prevent the gas from escaping (Rutqvist, 2012). Likewise, other parameters such as volume, temperature and pressure, heterogeneity (which affects the sweep efficiency), caprock permeability, formation thickness, the presence of reactive minerals, $CO_2$ solubility in brine, seismic fault potential, stress regime, injectivity and fracture formation should also be considered in any storage site selection process (Grataloup et al., 2009; Wei et al., 2013). Note that although we have learnt a lot from different storage pilot projects worldwide, the reality, however, is that the geological formations are generally very heterogeneous, and their properties vary significantly with location.

The Sleipner CCS project, operated by Equinor, is acknowledged to be the first storage project on a commercial scale (Ghosh, Sen, & Vedanti, 2015; Torp& Gale, 2004). The operation costs

were approved to be financially reasonable as the Norwegian government was exempt by almost the same amount of carbon tax (Kaarstad, 1992). The project started in 1996 using a saline aquifer located at a depth between 800 m to 1,000 m beneath the sea and the $CO_2$ is provided by a nearby natural gas processing field (Arts et al., 2004; Head et al., 2004). The storage formation Utsira is late Cenozoic, a 200-250 m thick sandstone that has stored approximately around 17 Mt of $CO_2$ since 1996 (Ringrose& Oldenburg, 2018), while the caprock formation is a Nordland shale with a thickness of 200-300 m. About three years after the start of the project, the plume migrated through thin layers of shale and stopped beneath the cap-rock (Arts et al., 2004). These layers help the dissolution process as they increase the interaction time between the brine and the $CO_2$ (Bachu, 2000).

Although nothing suggests the $CO_2$ at the Sleipner storage site may escape to the atmosphere, it is, however, essential to managing the risk of leakage through all stages of the storage process (Nooner et al., 2007). Moreover, to make better decisions, a reliable estimation of storage capacity and plume dynamic behaviour is needed. For this purpose, it is also essential to quantify any uncertainties in the model. There have been numerous studies (Allen et al., 2018; Cavanagh, Andrew J., Haszeldine, & Nazarian, 2015; Chadwick RA& Noy DJ., 2010; Chadwick, RA et al., 2009; Hermanrud et al., 2012; Hodneland et al., 2019; Nilsen et al., 2017) to find a match between the $CO_2$ flow in the Utsira formation resulting from simulations with the one from seismic surveys.

Gravity-driven flow was shown to be an essential factor in $CO_2$ plume migration in the Sleipner storage site (Cavanagh, Andrew J. et al., 2015). The researchers applied a capillary flow model to the Sleipner model to solve the overprediction of viscous effects in gravity segregated systems. While a better match was achieved in the northern part of the plume, the upslope plume migration was, however, overestimated, explained as being due to neglecting viscous effects, acting against the gravity drive. Their study, therefore, highlights the importance of

gravity-driven flow and suggests using models with sufficient vertical grid resolution or reduced orders models, such as vertical equilibrium. In order to improve the match, higher permeabilities (6 and 10 Darcy) for the Utsira formation has also been considered (Chadwick RA& Noy DJ., 2010; Chadwick, RA et al., 2009). The authors also studied cases with anisotropic permeabilities and increased the temperature to 36 $^{o}$C, the history matching result, however, was unsatisfactory.

A history matching study was previously performed for the Sleipner model (Nilsen et al., 2017). The results suggested that the plume outline is governed by caprock, permeability, and density influence, the $CO_2$-brine contact shape and porosity and injection rate affect the plume volume. Impact of uncertainties in temperature and fluid impurities on $CO_2$ migration in the Sleipner was investigated (Hodneland et al., 2019). While the impact of $CO_2$ impurity on plume migration was negligible, the study showed that raising the average storage site temperature to 46$^{o}$C improves the history matching results. A recent history matching study on Sleipner 2019 benchmark model (Ahmadinia, M.& Shariatipour, 2020) showed an improvement of around 8% in the plume match resulted by an absolute elevation calibration of 3.23 meters in the caprock. Calibrating porosity, permeability, $CO_2$ density and injection rate all together resulted in 5% improvements in the match, and once caprock elevation was adjusted too, the match increased by 16%.

Most of the previous studies undertaken on the Sleipner model took one factor at a time (OFAT) approach (Allen et al., 2018) in which the response to one parameter is investigated, and the rest are kept at their initial value. Some other works also considered a limited number of parameters in their sensitivity analyses (Hodneland et al., 2019). In order to fully span the uncertainty space, to the best of our knowledge, this is the first time the focus will be on the joint effect of six important parameters, to show their impact on the overall $CO_2$ migration and trapping in Sleipner. The motivation for the current work is to analyse the impact of individual

uncertain parameters on $CO_2$ storage process while interacting with other parameters. The uncertain parameters in this study are chosen from previous works based on their impact on the $CO_2$ plume migration. Mainly we are interested in the importance of caprock morphology in the range lower than the seismic detection limit. The limitation in the seismic maps introduces errors in the geological models used in reservoir simulation studies. This study indicates to some extent this limitation would affect the final results (*i.e.* plume outline) in the Sleipner model. The current work also helps to understand which of the addressed uncertain parameters for Sleipner model in literature should be prioritised and calibrated more carefully to improve the match. A similar study can be performed on other models to find the impact of uncertain parameters on final results and minimise the mismatch between simulation and observed data by improving the geological, operational and fluid properties.

Several sources are considered including uncertainties in the geological model (porosity, permeability and caprock morphology), aquifer property (pressure and temperature) and operational condition (volume entry rate into layer 9). We have generated ten thousand samples of the six uncertain parameters within their reported ranges in literature for Sleipner model Layer 9 and ran a simulation for each set of input parameters. The study was performed on the most recent Sleipner Benchmark simulation grid, 2019 (Santi A., Furre A.K., & Ringrose P., 2020). Considering the computational cost, we decided to perform the forward simulations using the VE modelling approach implemented in MRST-co2lab (Nilsen, Lie, & Andersen, 2016a; Nilsen, Lie, & Andersen, 2016b). co2lab is an add-on module providing a family of computational tools specially developed to study the long-term $CO_2$ storage in large-scale aquifer systems. The MRST-co2lab performance was compared with full 3-D simulations in previous studies (Ahmadinia, Masoud et al., 2019; Nilsen et al., 2011), and compared reasonably. Each forward simulation of $CO_2$ injection in Layer 9 for 12 years, took about 30 seconds using the VE approach; while using the same computational power configuration the

simulation could take up to 10 and 14 hours in black-oil and compositional simulation tools, respectively. In order the make this study feasible in terms of computational time, we used a cluster system to run 80 parallel simulations at a time. Random forest (RF) and decision tree (DTree) (Dumont et al., 2009; Tin Kam Ho, 1995) models available in Scikit-learn machine learning library in Python programming language are employed to find the importance of each parameter in the plume shape. The input is a matrix of ten thousand by six (temperature, rate, porosity, heterogeneity in permeability and porosity, pressure and caprock elevation) and the output is a matrix of ten thousand by four (The dice coefficient for four time steps).

## 2. Material and method

### 2.1. Geomodel

Sleipner (Santi A. et al., 2020) is a $CO_2$ storage site located off the western coast of Norway, in the Utsira formation. This site has been actively used by Equinor ASA since 1994, with a goal to prevent $CO_2$ emission associated with the gas production from the same region. The site has been monitored from its inception, and a comprehensive set of documents including well-logs, baseline seismic and time-lapse seismic data showing the plume migration extension has been established. The 2019 Benchmark model is the first complete 3D model of Sleipner. The model covers eight reservoir zones within the Utsira formation (L1 to L8) interbedded with eight laterally continuous intra-formational shale layers each of them about 1m thick. Layer 9, is located right above L8 separated by a relatively thick (~7.6m) shale layer. The seismic lateral and vertical resolutions are about 12.5 m (Santi A. et al., 2020) and 8 m (Chadwick, RA et al., 2004), respectively. Figure 1 shows the cross-sectional model for the Sleipner saline aquifer. In the present study, we include the caprock layer, L9 sand layer (red zone in Figure 1) and a continuous shale layer on the bottom of L9. The caprock and shale

layers are considered to be impermeable to fluid flow. The sand layer is characterised by a porosity between 27 to 40 % (Lothe& Zweigel, 1999; Pearce, Kemp, & Wetton, 1999) and horizontal permeability of 1100 to 5000 mD (Lindeberg et al., 2001).

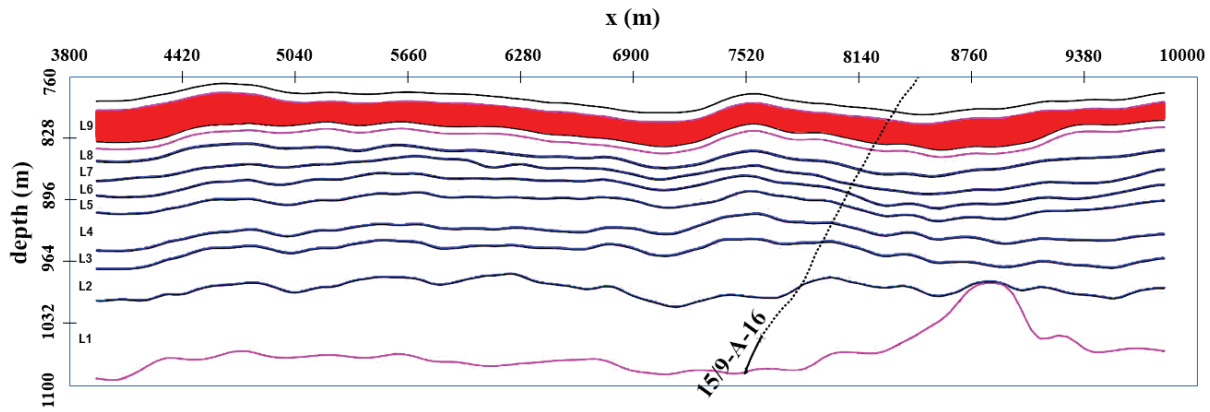


**Figure 1.** Cross-sectional model for the Sleipner saline aquifer. Purple line: bottom of the reservoir; blue lines: intra-formational shale layers; red zone: the only sand layer considered in this study (L9); dotted line: injection well in the original model; 15/9-A-16: injection well.

The injection well (15/9-A-16) surface coordinates in the model are (436137.42 m, 6470282.86 m) (Santi A. et al., 2020). The injection rates (volume entry rate into layer 9) for the reference case are the same as in (Nilsen et al., 2017) and are listed in Table 1. The rates presented in the table are representative of the $CO_2$ entry rate into Layer 9. The model is initially fully saturated with brine (33500 ppm), at hydrostatic balance.

Simulations are based on a two-phase black-oil formulation where gas can dissolve into water. Density and viscosity are functions of pressure and temperature using the sampled tables taken from the CoolProps open-source package (Bell et al., 2014). The residual saturation of 0.11 and 0.21 was assigned to the brine and $CO_2$, respectively (Singh VP et al., 2010). The $CO_2$ is injected for 12 years (1999 to 2010) and dissolution is considered in the study. Dissolved $CO_2$ is the dominant phase in the first couple of days after the start of injection, as the injected $CO_2$ encounters undersaturated brine (with respect to the $CO_2$). The fraction of dissolved $CO_2$ after

12 years approaches 9% in the base case model, which is within the reported range (*i.e.* less than 10%) in literature (Cavanagh, Andrew, 2013). Mineralisation trapping is likely to become effective in a timescale of hundreds to thousands of years (Wilkinson et al., 2009). Therefore, due to the time scale of the study and also computational cost, the mineralisation trapping has been neglected.

**Table 1.** $CO_2$ reservoir volume entry rates

| year | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rate ($m^3$) $\times 10^6$ | 0.03 | 0.08 | 0.14 | 0.21 | 0.31 | 0.44 | 0.62 | 0.87 | 1.18 | 1.57 | 2.06 | 2.65 |

Simulations are performed using the co2lab module in MRST (Nilsen et al., 2016b). Further information about the parameters used in this study is listed in Table 2.

**Table 2.** Model information

| **Parameter** | **Value** | **Reference** |
|---|---|---|
| Porosity | 0.27 – 0.4 | (Holloway et al., 2000; Lothe& Zweigel, 1999) |

| | | |
|---|---|---|
| Permeability (mD) | $(1.1 – 5) \times 10^3$ | (Lindeberg et al., 2001) |
| Number of cells (NX×NY×NZ) | 64×118×47 | (Santi A. et al., 2020) |
| Cell dimensions (m) (DX×DY×DZ) | 50×50×2 | (Santi A. et al., 2020) |
| Area ($km^2$) | 18 (6×3) | Seismic depth map |
| Seafloor temperature (ºC) | 7 | (Nilsen et al., 2017) |
| Brine viscosity (Pa.s) at 31ºC and 81bars | $8\times10^{-4}$ | (Nilsen et al., 2017; Singh VP et al., 2010) |
| $CO_2$ viscosity (Pa.s) at 31ºC and 81bars | $6\times10^{-5}$ | (Nilsen et al., 2017; Singh VP et al., 2010) |
| $CO_2$ density ($kg/m^3$) at 31ºC and 81bars | 687 | Coolprops data (Bell et al., 2014) |
| Water density ($kg/m^3$) at 31ºC and 81bars | 1020 | (Bickle et al., 2007a) |
| Brine compressibility ($Pa^{-1}$) at 81bars | $4.37\times10^{-10}$ | (Nilsen et al., 2017) |
| $CO_2$ compressibility ($Pa^{-1}$) at 81bars | $4.37\times10^{-9}$ | (Nilsen et al., 2017) |
| Rock compressibility ($Pa^{-1}$) at 81bars | $1.00\times10^{-10}$ | (Nilsen et al., 2017) |

### 2.2. Model uncertainty:

In the following, further details about the uncertain parameters are provided:

***Temperature and pressure***

One source of uncertainty addressed in previous studies on Sleipner model is $CO_2$ density (Alnes, Eiken, & Stenvold, 2008; Alnes et al., 2011; Cavanagh, Andrew J.& Haszeldine, 2014; Zhu et al., 2015) which is a function of pressure and temperature. We have considered temperature and pressure as uncertain parameters in the study.

The Sleipner aquifer is characterised with high porosity, permeability and lateral extension. This suitable combination, has resulted in negligible pressure build-up (< 0.1 Mpa) since the beginning of the injection phase (Chadwick, RA, Zweigel et al., 2004; Williams, Gareth& Chadwick, 2012). There is no downhole pressure gauge in the model (Boait et al., 2012; Furre et al., 2017). While wellhead pressure data have remained stable, however, since $CO_2$ at the wellhead is at the liquid/gas state, it is not possible to relate the pressure data at the wellhead to bottom-hole without having the gas/oil ratio (Furre et al., 2017). In this study, the reservoir pressure is assumed to be initially hydrostatic, and its uncertainty is addressed using a term (DP) within the range of -4 to 4 bars.

Temperature uncertainty in the Sleipner model has been a topic of discussion in the literature (Hodneland et al., 2019). Changing the temperature would primarily affect the $CO_2$ density and viscosity, and therefore its buoyancy and mobility. There is not any accurate data available for the temperature in the Sleipner due to the absence of any downhole gauge (Alnes et al., 2011; Eiken et al., 2011; Furre et al., 2017). A wide range of values was suggested for temperature in the Sleipner model, including 34 to 40ºC (Hermanrud et al., 2012) and 28.6 to 40.6 (Hodneland et al., 2019). Cavanagh and Haszeldine (Cavanagh, Andrew J.& Haszeldine, 2014) suggested two temperature values of 31ºC and 37ºC at 800 metres depth for the Sleipner layer 9. One of the most accurate data for the temperature in the Sleipner model is achieved from the gauges placed in water producing well in the Volve field, 10km north of the Sleipner injection well (Alnes et al., 2011). Using the temperature data from the Volve field, the temperature in the vicinity of Sleipner as a function of depth can be expressed as (Alnes et al., 2011):

$T=31.7ºC/km \times z + 3.4ºC\ (\pm 0.5\ ºC)$ **Equation. 1**

Where z denotes the depth measured in km. Based on Equation 1 which has been used in previous studies on Sleipner (Williams, GA, Chadwick, & Vosper, 2018; Williams, Gareth A.& Chadwick, 2017) the average temperature in Sleipner is around 31 ºC. In this study, we

added term DT to Equation 1 to consider the uncertainty in temperature in the model. Ten thousand random values of DT between -4 to +4 is considered to modify the aquifer temperature.

Changing temperature and pressure has an impact on $CO_2$ dissolution. With the temperature increased by 4$^oC$ in the base case model (at constant pressure condition), $CO_2$ solubility was reduced by around 1.1%. On the other hand, by decreasing pressure by 4 bars at isothermal condition, solubility dropped by around 0.2%. Temperature ranges used in previous studies on the Sleipner are listed in Table 3.

**Table 3.** Temperature range considered for the Sleipner model in previous studies on L9.

| **Study** | **Temperature ($^oC$)** | **Reference depth** |
|---|---|---|
| (Hodneland et al., 2019) | 28.6 to 40.6 | Average depth |
| (Chadwick RA& Noy DJ., 2010) | 29 | Top surface |
| (Alnes et al., 2011) | 32.2 | Top surface |
| (Bickle et al., 2007b) | 35 | Average depth |
| (Singh VP et al., 2010) | 35 | Top surface |
| (Allen et al., 2018) | 35 | Top surface |
| (Cavanagh, Andrew J. et al., 2015) | 35 | Average depth |
| (Cavanagh, Andrew J.& Haszeldine, 2014) | 31 & 37 | Average depth |
| (Baklid, Korbol, & Owren, 1996) | 37 | Average depth |

| | | |
|---|---|---|
| (Hermanrud et al., 2012) | 34 to 40 | Average depth |
| Current study | 27 to 35 | Average depth |

***Injection rate (volume entry rate into layer 9)***

The Sleipner original model is made up of nine layers each separated with a thin shale layer (Figure 1), and the plume is injected at a depth of 1010.5 metres (L1) below sea level (Singh VP et al., 2010). In this study, we are only modelling Layer 9 (L9), and the coordinate of the entry point in L9 is considered to be the same as of L1. In real case storage process, once injected, the plume encounters and passes through eight intra-formational shale layers (which are neglected in the current study) before reaching L9. The shale layers would result in 10-20 m thick $CO_2$ layers (beneath each shale layer), vertically stacked and extended by hundreds of metres laterally (Gregersen& Johannessen, 2001; Zweigel et al., 2004). Despite the accurately mapped areal distribution of the $CO_2$ plume, its flow behaviour is still subject to uncertainties (Cavanagh, Andrew J.& Haszeldine, 2014).

Moreover, the mechanisms of vertical migration (diffusion, migration points or both) and also the number and location of vertical migration points are uncertain (Nilsen et al., 2017; Zhu et al., 2015). Vertical migration here refers to the flow from intra-formational thin shale layers to the above sand layer. The volume entry rate into layer nine used in this study is taken from previous works (Nilsen et al., 2017) and represent an anticipation of the amount of $CO_2$ which is entering Layer 9 and is subjected to uncertainty because we are not sure how much of the injected $CO_2$ in L1 reaches to L9 in reality. Ten thousand random rate multipliers (RM) between 0.7 to 1.3 are applied to the benchmark's volumetric rate to include the uncertainty of the entry rate into Layer 9.

***Porosity, permeability and caprock elevation***

The reported ranges for porosity and permeability data of Sleipner Layer 9 are 0.27-0.4 (Holloway et al., 2000; Lothe& Zweigel, 1999) and 1100-5000 mD (Lindeberg et al., 2001), respectively. In this study, ten thousand permeability realisations are generated using a lognormal distribution approach within the reported range. Porosity realisations are then generated from permeability data using the Kozeny-Carman correlation (Carman, 1937). The typical seismic resolution is around 10 m and topography variations below this resolution are referred to as rugosity (Jones et al., 2009; Pringle et al., 2010). The reported seismic vertical resolution for the Sleipner model is 8 m (Chadwick, RA et al., 2004). In order to investigate the importance of the topography variations below the seismic detection range, ten thousand realisations of top surface elevations within the range of $\pm 5$ m are considered using Gaussian random fields.

Figure 2 shows an illustrative (not representative of actual uncertainty ranges) of data for porosity, permeability and caprock elevation perturbations within the mentioned ranges. Note that since we are using the VE model, by porosity and permeability, we are referring to their column-wise averaged values. Therefore, the impact of their variation on plume migration might be underestimated in comparison to a full 3D model.

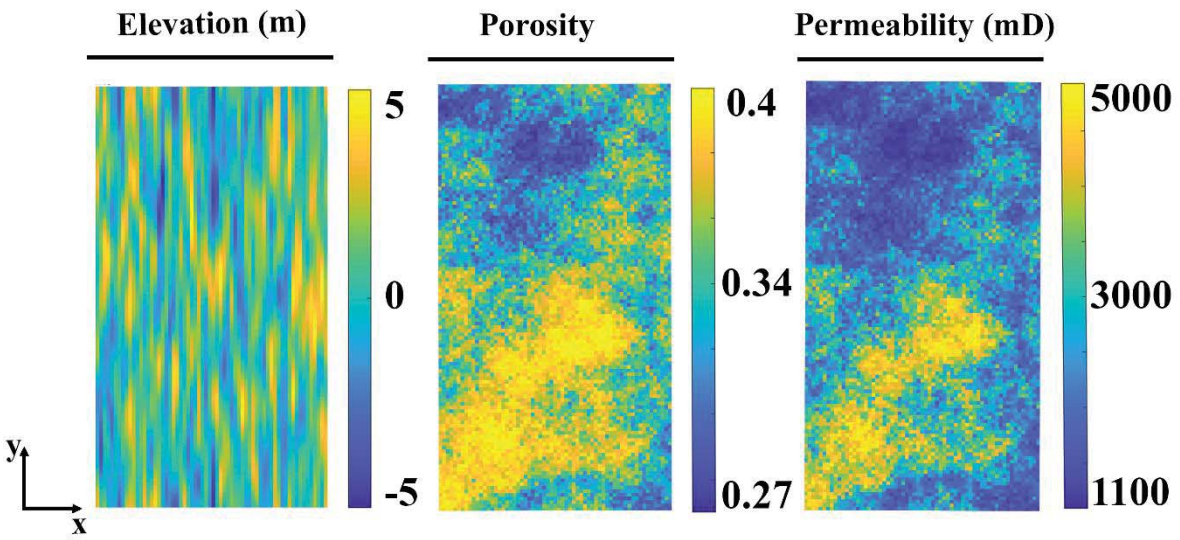


**Figure 2.** Samples of realisations representing the elevation (m), porosity and permeability (mD) distribution in the model. Figures show the top view of the grid (x and y-direction).

### 2.3. Simulation approach

In this study, we ran ten thousand sets of simulation with two phases, $CO_2$ and formation brine using the co2lab module in MRST (Nilsen et al., 2016b), which is based on the vertical equilibrium simulation. The impact of uncertainty in caprock topography, reservoir pressure, reservoir temperature, porosity and permeability heterogeneity and volume entry rate into layer nine are studied on plume migration and structural trapping. In order to treat the parameters equally, in each of the ten thousand simulation runs, the six parameters are randomly selected within their allocated range at once, to provide the inputs of the simulations.

The results from a similar VE setup were reasonable when compared with full 3D simulations models in previous studies using synthetic (Ahmadinia, Masoud et al., 2019) and Sleipner (Cowton et al., 2018; Nilsen et al., 2011) models. Therefore, the VE method is used in this study to decrease the computational cost. Two base assumptions of VE modelling are the following: First, the hydrostatic equilibrium between brine and $CO_2$ is pre-assumed throughout the simulation. Due to the difference between the fluid densities, the gravity segregation

process occurs significantly rapidly, and fluids form two separate layers, in comparison to the lateral plume migration. For injection rates and formation thickness typical for geological carbon storage sites, the VE model is likely to be valid for formation permeabilities higher than about 100 mD to have fast vertical segregation of fluids (Court et al., 2012).

Moreover, the injected plume may migrate several kilometres in the horizontal direction with minimal vertical movement (Shariatipour, Seyed M., Pickup, & Mackay, 2016). This makes the second assumption, stating that the vertical flow migration can be considered negligible compared with the horizontal one (Nordbotten& Celia, 2011). In a VE simulation model, the problem is reduced to 2-D, providing the modeller to allocate the computational cost to increase the lateral resolution beyond what would be otherwise feasible in 3-D simulations. The MRST implementation of a VE model is written based on black-oil based formulations with upscaled models for capillary pressure and mobility.

Note that vertical heterogeneity in permeability is ignored in the basic VE models. Consequently, in this study, we do not include intra-layer flow in the simulation here, and by permeability, we are referring to horizontal permeability only. However, in a study undertaken by (Møyner& Nilsen, 2017; Møyner, Andersen, & Nilsen, 2018), the authors presented a multi-layer VE approach with full 3-D simulations locally where needed. The errors introduced by VE modelling can, in many cases, be lower than the errors resulting from low lateral resolution to make the 3-D simulations computationally feasible (Nilsen et al., 2016a). Readers are referred to (Nordbotten& Celia, 2011) for a detailed description of vertical equilibrium models.

Hodneland et al. (Hodneland et al., 2019) showed that for a particular set of assumptions (thin plumes moving upwards under a sloping caprock), $CO_2$ migration velocity is given by:

$$V \cong \frac{k(\rho_w - \rho_g) g \sin\theta}{\mu_g \Phi}$$ **Equation 2**

where $V$ represent the fluid velocity, $\theta$ is the caprock tilt angle and $\mu_g$ is $CO_2$ viscosity. According to Equation 2, an increase and drop in the values of permeability and $CO_2$ density, respectively, results in higher migration speed. Higher permeability means less resistance to flow. Moreover, reducing the $CO_2$ density results in a higher driving force for the fluid migration, and the fluids tend to migrate from a zone with higher density to one with a lower value. Within the temperature range used in this study and the average pressure of 83 bars in layer 9 (Hodneland et al., 2019), based on the data provided in the literature (Bachu, 2003), increasing the temperature results in lower viscosity and density, thus increasing migration velocity. In this study, we have used both plume shape similarity (section 2.4) and structural trapping (section 2.5) as quantities to analysis the importance of uncertain parameters on $CO_2$ migration and trapping process.

### 2.4. Plume similarity

Several methods have been previously utilised to quantify the similarity of the plume migration of two different geological models. Some researchers (Han et al., 2011; Manceau& Rohmer, 2014; Manceau& Rohmer, 2016) have compared the location of the plume's centre of mass with a reference point, such as injection point or plume centre of mass of the base model. Another method is the Sørensen–Dice coefficient (SDC), a statistic used to quantify the similarity of two discrete samples (Dice, 1945; Sørensen, 1948),

$$\mathrm{SDC} = \frac{2|X \cap Y|}{|X| + |Y|}$$ **Equation 3**

This method has been recently used to compare the similarity of the simulated and observed $CO_2$ footprint at the Ketzin (Lüth, Ivanova, & Kempka, 2015) and Sleipner (Allen et al., 2018; Hodneland et al., 2019) storage sites. In the current study, X represents the plume outline from the simulation at the desired time, and Y is the observed footprint generated from the seismic data at the same time. Therefore, SDC equals twice the overlapping area, divided by the summation of plume outlines (Equation 3). SDC ranges between 0 and 1, where an SDC equals to 1 corresponds to identical samples.

To better understand the underlying relationship between our target variables SDC and uncertain variables, we must employ a reliable data-driven model that unveils linear/non-linear dependence in data. This enables us to quantify later the importance of each uncertain variable (*i.e.* input variables of the data-driven model) based on their contribution to the predicted target values. The model inputs are caprock elevation, temperature, pressure, porosity and permeability heterogeneity and injection rate (volume entry rate into layer 9), while the output (or target) variables are SDC values in different years. Knowing which model is appropriate for a given scenario is not always understood and requires more than one data-driven method to be trained on any supplied dataset. Therefore, we initially fit a baseline Linear Regression (LR) model to a training set, predict an unseen test set (25% of the entire data set), and then compared the predicted target values against the observed data. The baseline LR model reached an R-squared of about 0.3, which is very poor. We then use the following models to improve the baseline model prediction:

K-nearest neighbours (KNN): a neighbours-based regression model that performs the learning process based on the proximity of K closest training examples of each query point, where K is a user-defined constant (Goldberger et al., 2004).

Decision Trees (DTree): a tree-based model that sets up decision rules inferred from the observed data. Decision-tree learners can generate over-complex trees that fail to reliably generalise to unseen data (Dumont et al., 2009).

Random Forests (RF): an ensemble method to link the predictions of several decision trees to improve the predictive capability of each estimator while minimising the risk of overfitting (Tin Kam Ho, 1995).

The training/testing set used for the baseline (LR) model is used to compare the predictive power of our KNN, DTree and RF models, and the results are listed in Table 4.

**Table 4.** Comparison between the employed data-driven models.

| Model | Mean absolute error | Mean squared error | R-squared |
|---|---|---|---|
| LR | $6.7423\times10^{-3}$ | $2.8866\times10^{-4}$ | $3.4532\times10^{-1}$ |
| KNN | $2.8709\times10^{-2}$ | $1.5298\times10^{-2}$ | $7.9643\times10^{-1}$ |
| RF | $3.1603\times10^{-5}$ | $1.7620\times10^{-7}$ | $9.9912\times10^{-1}$ |
| DTree | $8.6138\times10^{-6}$ | $1.5492\times10^{-6}$ | $9.9967\times10^{-1}$ |

Table 4 clearly shows that DTree and RF models outperform the LR and KNN models by a large margin, with the RF model having R-squared of 0.9991. Therefore, we solely keep our DTree and RF models for evaluating variable importance. As for variable importance, simple models strongly represent themselves and are highly interpretable as they are based on simple rules. However, complex statistical models, such as ensemble methods, are not easy to explain. Instead, an interpretable approximation of the original statistical model can be used to represent

an explanation model. To address this problem, we used a unified structure for interpreting predictions, namely the SHAP (SHapley Additive exPlanations) method introduced by (Lundberg& Lee, 2017) in 2017. The SHAP framework identifies the class of additive feature importance methods and finds a solution in this class that quantifies variable importance. SHAP relates to the family of models called "additive feature attribution methods" where the real variables are replaced by additive variables in that the explanation model for variable importance is formulated as a linear function of additive binary features. The exact solution to SHAP values is computationally expensive. They can, however, be approximated by combining different additive feature attribution methods. SHAP also provides each feature with an importance value for a particular prediction. (Lundberg& Lee, 2017) demonstrated that SHAP is better adjusted using human intuition and more robustly distinguish between model output classes than several existing methods.

### 2.5. Structural trapping estimation

Structural traps (ST), the most immediately available trapping mechanism, corresponding to the local maxima of the top surface play a key role in $CO_2$ storage. In this trapping mechanism, the $CO_2$ plume is prevented from further upward migration after reaching caprock (Shariatipour, S. M., Pickup, & Mackay, 2016). MRST-co2lab implements several algorithms to identify the ST in the sealing caprock without any flow simulation. Due to its low computational cost, this method can be used in large models. "Spill path" refers to the path $CO_2$ follows beneath the caprock assuming infinitesimal flow. Once injected, the $CO_2$ plume tends to move upwards and fill the traps/ridges below the caprock. When a trap has been filled to its spill point, any additional $CO_2$ spills and possibly lead to the neighbour trap (Nilsen, Lie, & Andersen, 2015). Individual traps are connected by spill paths, like the lakes being connected

by rivers in surface hydrology. The static ST capacity in terms of $CO_2$ mass is estimated using Equation 4 (Allen et al., 2018).

$$ST\ capacity\ (kg) = \int_{\Omega} \rho V \phi (1 - S_{rw}) \qquad \textbf{Equation 4}$$

where $\rho$ denotes for $CO_2$ density at aquifer condition (kg/m$^3$), $V$ is trap volume (m$^3$), $\phi$ is porosity, $S_{rw}$ is residual water saturation and the integrate is over the boundary of $\Omega \subset R^n$. Note that for traps with the same pore volume ($V\phi$) but located at a different depth, ST capacity differs due to $CO_2$ density variation. Readers are advised to refer to (Nilsen et al., 2015) for more details about ST capacity and spill-point analysis.

## 3. Results and discussion

In this section, we analyse two simulations with the best and worst average (in four time steps) plume match, together with the ones with the minimum and maximum ST capacity. Moreover, the importance of each of the uncertain parameters on plume outline is investigated considering the joint effect of all parameters and also one factor at a time (OFAT) approach. Figure 3 shows the case with the best match between the plume outline from simulation results (1$^{st}$ row) and seismic (2$^{nd}$ row). The figure also visualises the overlapped area of the plume outlines (3$^{rd}$ row). Note that the presented results are generated from one set of uncertain parameters which have the highest average SDC in 2001, 2004, 2006 and 2010. The average SDC of the four time steps for each of the ten thousand simulations are calculated, and the results are sorted from best (highest average SDC) to worst (lowest average SDC) match. The plume outline at the end of each of the studied timestep is reported. The reason to focus on these four outputs is that the plume outline data for Sleipner model is available for limited timesteps of which we chose 2001, 2004, 2006 and 2010 in this study.

It should be noted that although the vertical seismic resolution is about 8 m, it is possible to estimate the thickness of topmost $CO_2$ below the seismic detection range using methods such

as structural analysis of the reservoir top (Chadwick RA& Noy DJ., 2010; Chadwick, R. A. et al., 2009; White et al., 2018) and spectral decomposition technique (Huang et al., 2015; White et al., 2015; White, Williams, & Chadwick, 2013; Williams, Gareth& Chadwick, 2012).

### 3.1.Best and worst matches

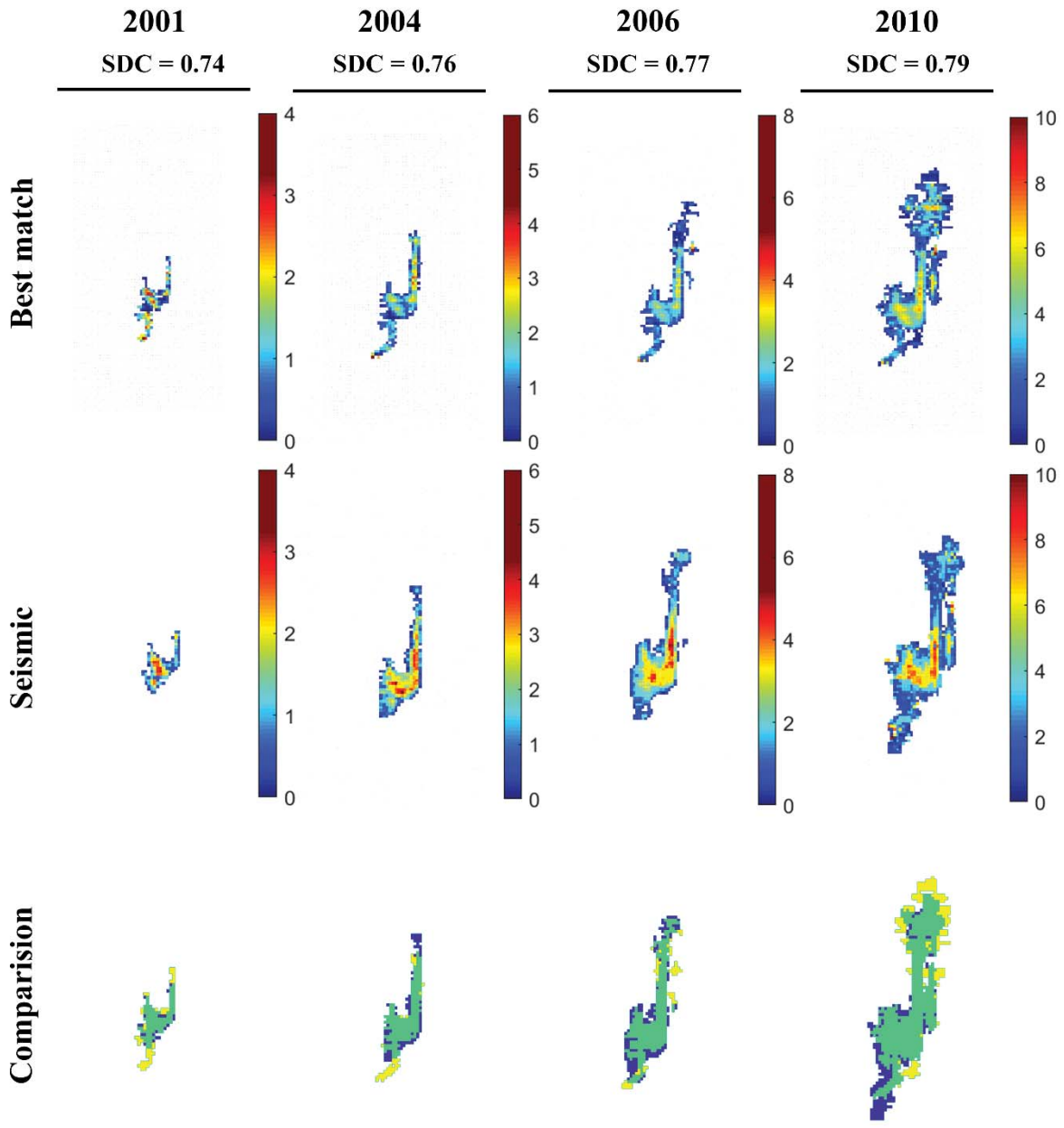


**Figure 3.** Plume outline from simulation results with the worst average match (1st row) and seismic (2nd row), together with their comparison (3rd row; dark blue: seismic, yellow:

simulated, green: overlapped) in 2001, 2004, 2006 and 2010. The legend shows the plume thickness (m).

The plume shape at different time steps shows a good match between the plume lateral extension outline from seismic and simulation studies. The presented results are, however, the best averaged outcome over four time steps from our ten thousand simulations and not necessarily the best possible match for the model. The uncertain input parameters for the simulation are listed in Table 5.

**Table 5.** Simulation parameters for the case with the highest average SDC.

| Parameter | Value |
|---|---|
| Average SDC of the studied time steps | 0.76 |
| DP (bar) | -3.21 |
| Rate multiplier (RM) | 0.86 |
| DT (°C) | 3.23 |
| $\phi_{avg}$ | 0.32 |
| $K_{avg}$ (D) | 2.85 |
| Average elevation change (m) | 0.43 |
| Average absolute elevation change (m) | 2.19 |

According to Table 5, a positive DT (3.23 °C) results in a better match. Using Equation 1, with an average depth of in Layer 9 (~ 818 m) and DT equals to 3.23°C, aquifer temperature is around 34°C. The reservoir pressure at the corresponding depth and with DP=-3.21 bars is around 78 bars. $CO_2$ density at this pressure and temperature condition, based on the Coolprops (Bell et al., 2014) data would be around 390 kg/m$^3$. Cavanagh (Cavanagh, Andrew J.& Haszeldine, 2014) and Zhu et al. (Zhu et al., 2015), considered an average $CO_2$ density of 355

kg/m$^3$ and 479 kg/m$^3$, respectively. A density of 391 kg/m$^3$ was also suggested by (Nilsen et al., 2017) in one of their calibrated set of parameters for Layer 9.

The Sleipner condition is close to the critical point (30.4 ºC and 73.8 bars), and $CO_2$ has a gas-like behaviour in a supercritical condition (Hodneland et al., 2019). Therefore, increasing the temperature results in a significantly lower density and consequently, a higher buoyancy force. Moreover, a higher temperature at pressures close to the average pressure of 83 bars in Layer 9 (Hodneland et al., 2019), results in a lower viscosity (Bachu, 2003) and consequently higher mobility. In this condition (higher temperature) the $CO_2$ plume conforms more accurately the caprock morphology. Increasing the temperature was previously (Hodneland et al., 2019) suggested improving the match between simulation and seismic surveys results. As it is shown in Table 5, an RM of 0.86 results in the best average plume match. Note that the results presented here are just one of the many possible "acceptable" results. Since the parameters are not entirely independent, a different set of input parameters might potentially lead to the same if not better results.

Table 5 also shows that realisations identified as the best match, happened to have average porosity and permeability values of 0.32 and 2.85 D respectively. An average absolute elevation change of around 2.19 m is observed in the case with the highest averaged SDC.

Similar to the previous case, Figure 4 shows the results for the case with the lowest averaged SDC over the studied time steps. The uncertain input parameters for the simulation are listed in Table 6.

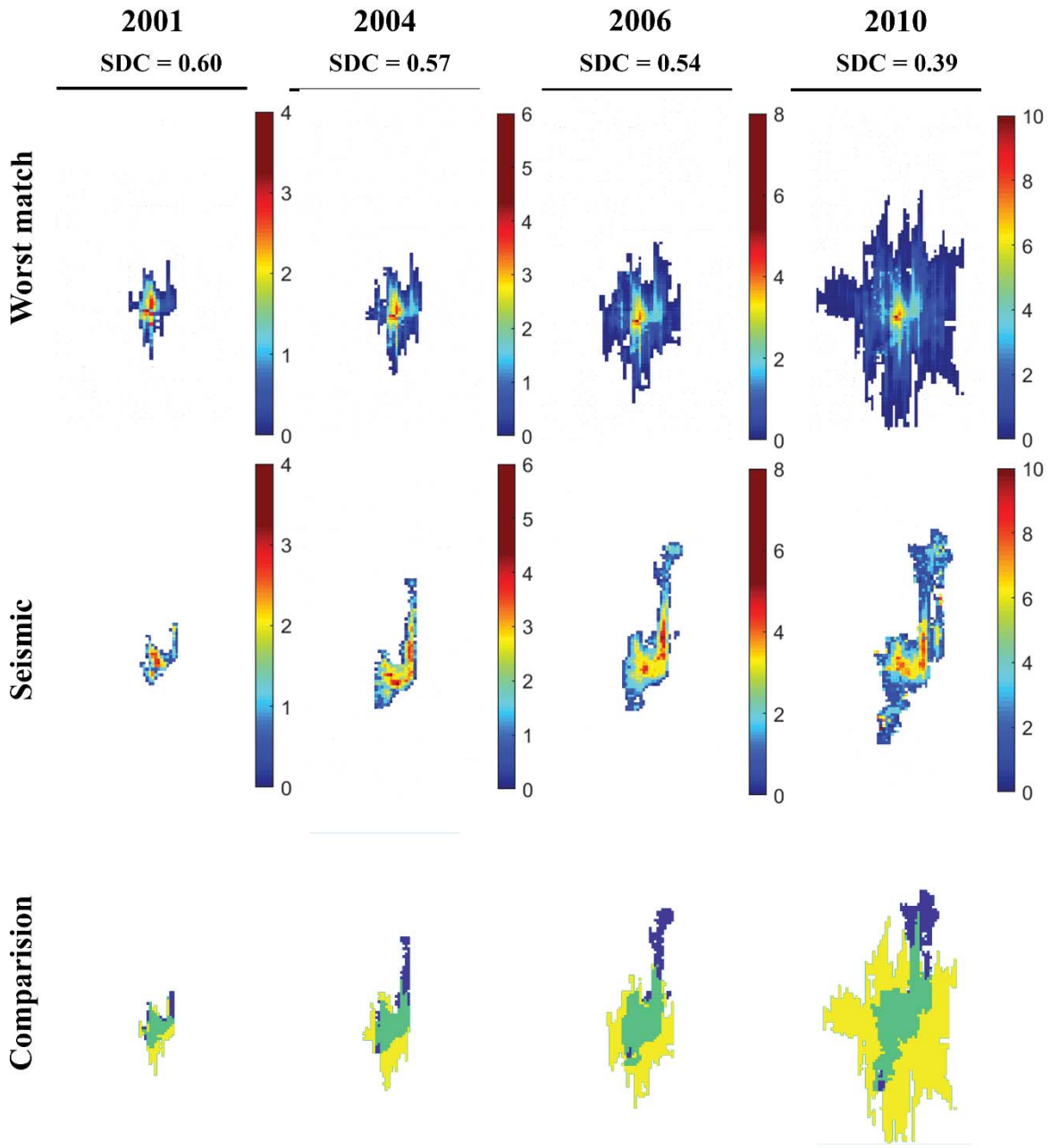


**Figure 4.** Plume outline from simulation results with the worst average match (1st row) and seismic (2nd row), together with their comparison (3rd row; dark blue: seismic, yellow: simulated, green: overlapped) in 2001, 2004, 2006 and 2010. The legend shows the plume thickness (m).

According to Figure 4 (3rd row), all the plume outlines from the simulation results, are larger than the ones from the seismic studies. It can be justified by the rate multiplier, which is around

its maximum possible value (Table 6). The case with minimum average SDC in 2001, 2004 and 2006 has a DT of -1.31 ºC. The impact of uncertain parameters on plume migration changes throughout the simulation time and the results presented here shows the average impact on plume match. The problem we are dealing with in the Sleipner is complex; therefore, a different set of parameters might account for the best match in each time step. We will use a data-driven modelling approach in Section 3.3. to find the contribution of each parameter in $CO_2$ plume migration more precisely.

**Table 6.** Simulation parameters for the case with the lowest average SDC.

| Parameter | Value |
|---|---|
| Average SDC of the studied time steps | 0.52 |
| DP | 3.21 |
| Rate multiplier (RM) | 1.28 |
| DT (oC) | - 1.31 |
| $\phi_{avg}$ | 0.35 |
| $K_{avg}$ (D) | 3.17 |
| Average elevation change (m) | 0.27 |
| Average absolute elevation change (m) | 3.41 |

### 3.2. OFAT approach

As mentioned earlier, the following set of values are assigned to the uncertain parameters in this study:

- DT: -4 to 4 ºC.
- DP: -4 to 4 bars.

- RM: 0.7 to 1.3.
- Porosity: 0.27 to 0.4.
- Permeability: 1.1 D to 5 D.
- Caprock elevation: -5 to +5 m.

In this section, we employ the well-known OFAT approach, by analysing the response of the model to the change in individual uncertain parameters, while keeping the rest of inputs at their initial state. One hundred values (and realisations for the cases of porosity, permeability and elevation) are considered within the allocated range. The results are presented in Figure 5, with the initial values of the model highlighted in purple. It is clear that while changing one parameter at a time, the plume has a better match in 2001, and the simulated plume outline becomes less similar to the one from seismic as injection continues.

It is hard to find a similar trend between the results of the same parameters in various time steps. For example, looking at the overall trend for rate, while in 2010 an increase in the RM rapidly decreases the matching accuracy, the same phenomenon in 2001, 2004 and 2006 has a less significant effect on SDC.

As claimed in work undertaken by (Nilsen et al., 2017), caprock topography is one of the main factors in shaping the plume at Sleipner. There is no visible relationship between the employed realisation for caprock elevation and the calculated SDC in different time steps. The results in 2004, 2006 and 2010, however, show more resemblance than 2001. In the current study, the elevation is changing within -5 to 5 metres, with an average absolute elevation change of around 2 meters. We can see even this small variation, which is a quarter of the vertical resolution (8 meters), can significantly impact the $CO_2$ plume migration behaviour.

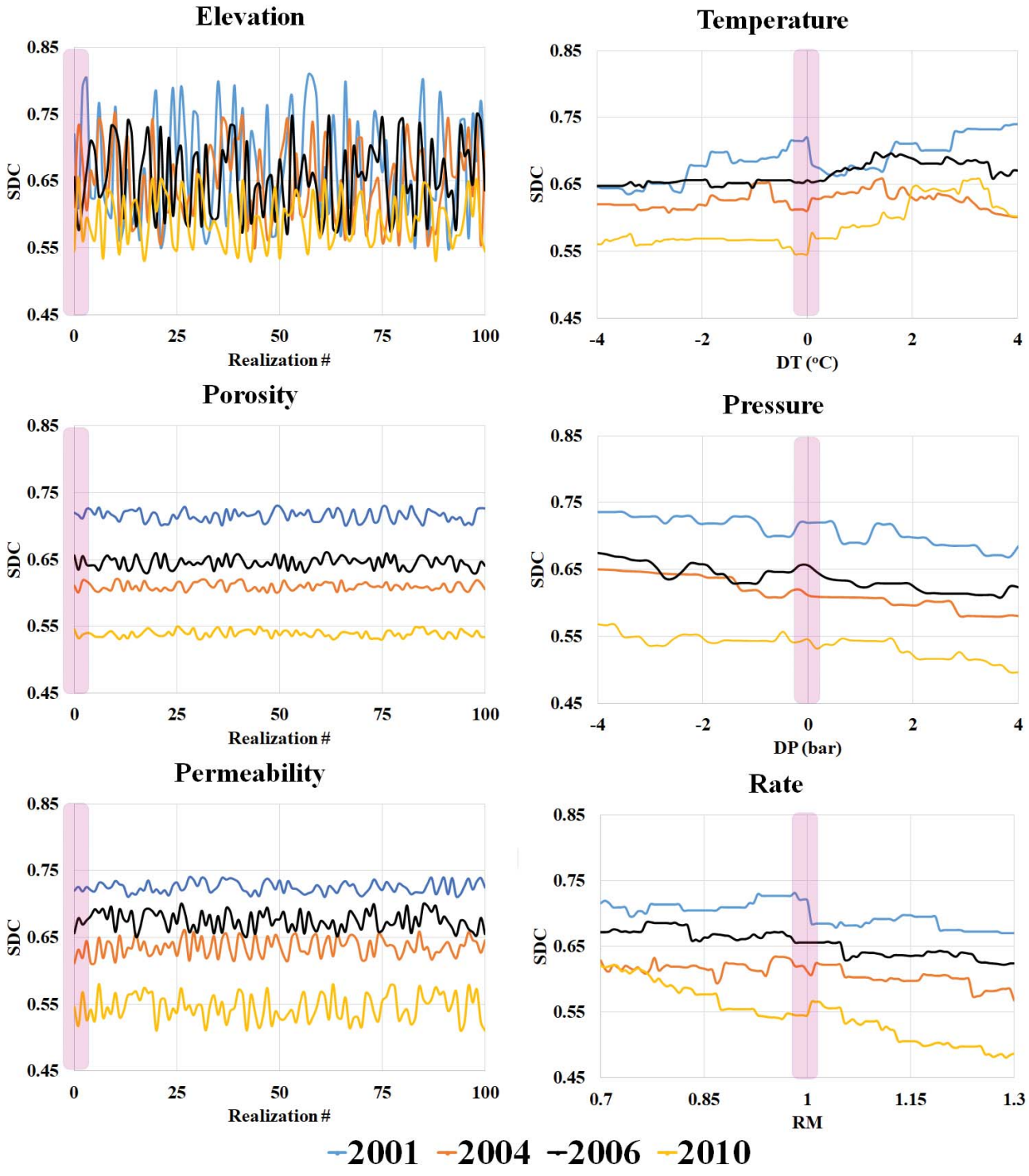


**Figure 5.** Calculated SDC for each of the input parameters using an OFAT approach. Initial values of the model are highlighted in purple. Blue, orange, black and yellow lines represent the results for the years 2001, 2004, 2006 and 2010, respectively.

Note that the overall constant trend of SDC for DT (between -4 to 0) or RM (between 0.7 to 1) in 2004 and 2006 does not mean that the plume outline has remained constant while these parameters are changing as it is possible to have two different outlines with the same SDC. This can be observed in Figure 6, showing the plume outline in an example case for various values of DT while keeping other parameters constant. The figure clearly shows that not only plume extensions are not identical, but also the $CO_2$ plume seems to become thicker by increasing the DT. This could be addressed as one of the limitations of the Dice method as it considers the difference in the uncommon elements. In other words, if circle A and B have an area equivalent to 90% and 110% of circle C, respectively, we will have $SDC_{A\,\&\,C} = SDC_{B\,\&\,C} = 0.90$. Moreover, the plume thickness is not considered in the similarity measure, which is a limitation of seismic interpretation, providing only a 2D surface (Hodneland et al., 2019).

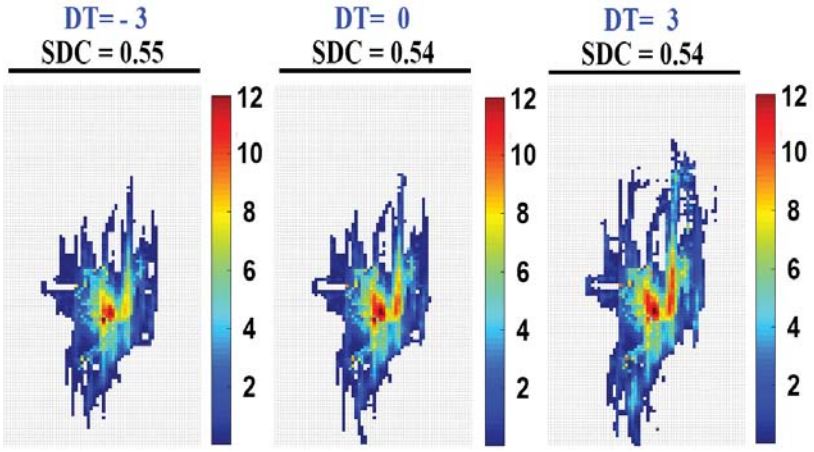


**Figure 6.** Plume outline in an example case for various values of DT. The legend shows the plume thickness (m).

Figure 5 shows that the plume match is affected by permeability distribution in the model which is in agreement with the results of the work undertaken by Nilsen et al. (Nilsen et al., 2017). They used adjoint-based sensitivities, adjusting parameters in a way to minimise the mismatch between the observation and simulation. The optimised permeability data set in their work, however, had an average one order magnitude higher than the original value reported for Layer 9. While the reported range of permeability for Layer 9 is between 1.1 and 5 Darcy, the authors proposed an average permeability of 21.7 Darcy to improve the plume match. In the current study, we are not changing the range of reported permeability (and porosity) data. Therefore, the impact of permeability on overall plume match might be less significant than their work. Uncertainties in porosity show a smaller impact on plume dynamics than uncertainties in permeability and caprock elevation. While the level of perturbations applied to these parameters affects their degree of impact, however, similar results were observed in work performed by (Allen et al., 2018) where even an increase of ±50% of the original average porosity had minimal impact on the match in the Utsira model.

Results of increasing DT showed an overall improvement in the match. The analysis of the worst and best average match also showed that higher temperature would result in a better match in plume for Sleipner model. An increase in temperature in the Sleipner model results in a lower density and viscosity and consequently, higher mobility (Hodneland et al., 2019). Figure 5 also shows that reducing pressure results in a better match which could be due to lower density and viscosity. While the OFAT approach in this study for temperature shows a slight improvement in plume match by increasing the temperature, a recent study on Sleipner model showed that raising storage temperature would significantly improve the match (Hodneland et al., 2019). Hodneland et al. (Hodneland et al., 2019) used the previous Sleipner model; therefore, the base case of their model has a different caprock elevation than the one used in this study. Moreover, the distribution of porosity and permeability data was also different.

Here, we examined the impact of uncertainty in porosity, permeability and caprock elevation in calculated SDC for temperature, using the OFAT approach. For this regard, we repeated the simulations illustrated in Figure 5, using different sets of realisations for porosity, permeability and caprock elevation (within the allocated range and using the same distribution approach as in Section 2.2). Later, the response of the model to the change in temperature was analysed (one parameter at a time). The results showed different trends in SDC than the one observed in Figure 5 for temperature. The analysis was performed for rate, and a similar result was achieved. It is therefore difficult to make a general statement on the impact of a parameter on the overall match. The lack of agreement in the trends shows the limitation of the OFAT approach as due to the complexity of the problem, the presence of other sources of uncertainty affects the results. In the next section, we introduce a data-driven approach to investigate the contribution of individual parameters, in the presence of other sources of uncertainty, more precisely.

### 3.3. The joint effect of parameters

Percentage overall variable importance plots for both RF and DTree models approximated by SHAP are presented in Figure 7. All SHAP values are obtained by averaging over five trials with the different starting point (multiple restarts) to ensure reliability and reproducibility of estimations. Both RF and DTree models consistently identify elevation, injection rate, temperature, permeability and pressure as the most important parameters when averaged over the four years under study. Heterogeneity in porosity is the lowest-impact variable for both RF and DTree models. Percentage variable importance is also computed for every variable at each year (see the number in the boxes). A higher percentage shows the dominant impact of the

parameter on the SDC or in other words, a better match between the simulated and observed plume outlines. For instance, the variable "injection rate" has the highest impact on model prediction in 2010. Here we discuss the trends observed in the RF model (which are consistent with the DTree model as well).

Results clearly show that the impact of each parameter might change throughout the simulation. For instance, while the elevation is the dominant factor in 2001 (38%), its impact becomes less significant later in 2010 (14.43%). Meanwhile, the importance of injection rate seems to increase with time and its percentage predictor importance changes from 17.83% in 2001 to 32.40% in 2010. One justification is that the injection rate is overshadowing the impact of other uncertain parameters in later years. This happens because the volume entry rate into layer 9 is not constant and increases with time. Injection rate is the only parameter which impacts the mass flow rate in aquifer directly. Since we are using multipliers for this parameter, a constant amount is not applied throughout the simulation. As the injection rate in the model increases, its impact becomes more significant as well. In the case of temperature, pressure and caprock elevation, the weight of adjustment in these parameters is constant during the simulation while adding/subtracting a value within the same range over the simulation.

The results indicate the caprock elevation as the most important parameter in controlling the plume outline in Sleipner model. Note that the average absolute change on caprock elevation in this study is about 2m with a maximum and minimum of 5m and -5m, respectively. Although there have been several sources of uncertainties reported in the literature for determining the best plume match, the impact of caprock morphology, however, seems to be underestimated as it has average importance of about 26%. We used an elevation change in a range of about half of the reported seismic resolution in the Sleipner, and the impact is yet significant. Permeability and porosity contribute to changing the shape of the plume outline with overall percentage importance of around 13% and 6% respectively.

Uncertainty in pressure has overall percentage importance of around 9%. Pressure and temperature both have an impact on viscosity and density. Based on the Coolprops data (Bell et al., 2014), an initial condition of T=31ºC and P=81bars results in a density of around 687 kg/m$^3$. Considering values of -4 ºC and +4 ºC for DT (T=27ºC and T=35ºC), while keeping the pressure constant results in a density of 754 kg/m$^3$ and 491 kg/m$^3$ respectively (*i.e.* 263 kg/m$^3$ change in density). The corresponding density for cases with DP of – 4 bars to 4 bars in an isothermal condition becomes 710 kg/m$^3$ and 650 kg/m$^3$, respectively (*i.e.* 60 kg/m$^3$ change in density). This clearly shows that within the pressure and temperature ranges in Sleipner model, density in more sensitive to changes in temperature than the pressure which would justify the higher overall importance percentage of temperature than pressure.

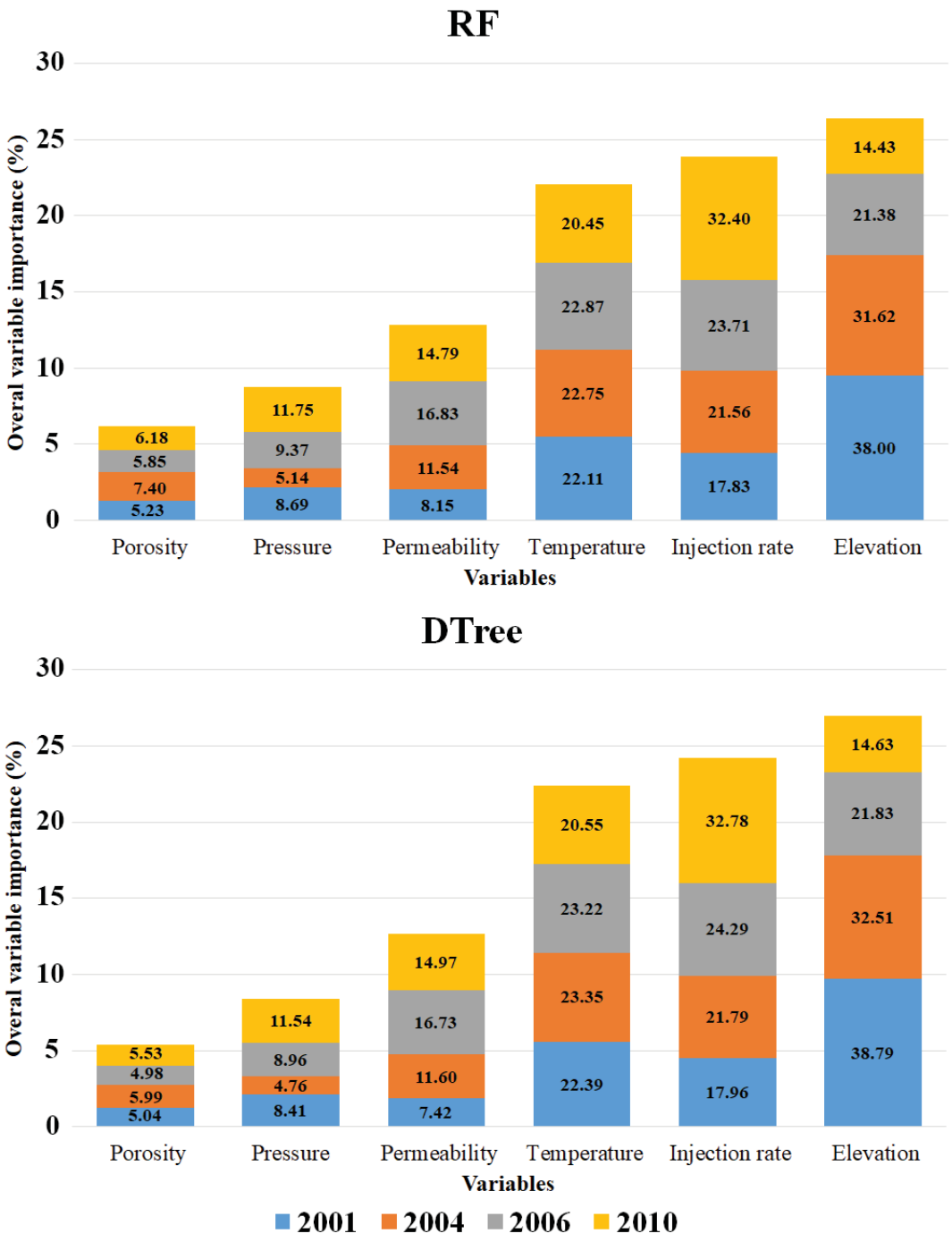


**Figure 7.** Percentage overall variable importance plots for DTree and RF models approximated by SHAP

### 3.4. ST capacity estimation

The parameters in cases with the highest and lowest ST capacity are listed in Table 7. As the $CO_2$ density increases, the structural traps can contain more $CO_2$ ($mass = density \times$

$volume\ (constant)$). As mentioned earlier, a lower temperature and higher pressure in the Sleipner aquifer condition would result in a higher density and consequently higher ST capacity. The findings in this study agree with previous works (Allen et al., 2018; Bachu, 2003) and the minimum and maximum ST capacity corresponds to the DT of 3.78 and -3.95, respectively. As expected, a larger (and positive) caprock elevation change and porosity increase the structural trapping. The cases with minimum and maximum ST capacities had a DP of -3.01 and 3.23 bars, respectively.

**Table 7.** Simulation parameters for the cases with minimum and maximum ST capacity.

| **Parameter** | **ST capacity** | |
|---|---|---|
| | **Min** | **Max** |
| ST capacity (Mt) | 1.46 | 3.68 |
| DP (bar) | -3.01 | 3.23 |
| DT (oC) | 3.78 | -3.95 |
| $\phi_{avg}$ | 0.30 | 0.35 |
| Average elevation change (m) | 0.25 | 1.06 |
| Average absolute elevation change (m) | 1.13 | 3.63 |

## 4. Conclusions

In this work, we focused on the joint effect of uncertain parameters in which their impact was believed to influence the overall $CO_2$ migration and trapping in the Sleipner 2019 benchmark model. For this purpose, we performed ten thousand forward simulations to analyse the importance of porosity and permeability heterogeneity, reservoir temperature, reservoir

pressure and caprock elevation of plume outline. To make the study computationally feasible, the simulations are performed using the VE approach. We disregarded internal layers and modelled the whole thickness of the aquifer as one layer as a single VE model. Upwards migration of $CO_2$ through internal layers was implicitly modelled through the rate multipliers. A more detailed study would involve applying a VE model to each internal layer which we consider worthy of future work.

The results showed that $CO_2$ density values of around 390 kg/m$^3$ improve the plume match in the Sleipner model. The caprock morphology was shown to be the most critical parameter in controlling the plume migration with the overall importance of 26% followed by the injection rate (24%), temperature (22%), heterogeneity in permeability (13%), pressure (9%) and porosity (6%). We would like to highlight that the best combination of parameters reported in this study is one of the many possible answers. As was shown in the results of the OFAT approach, the effect of a parameter on the plume outline can be different in the presence of another parameter which could be considered as one of the limitations of OFAT approach. For example, while previous studies showed that increasing temperature would result in a better match in Sleipner model, our results showed that this statement is not always valid and depends on the realisations used for caprock, porosity and permeability. A similar result was achieved for the rate, and various trends of SDC vs RM was observed after using different realisations for porosity, permeability and elevation. There are not any fixed correct sets of realisations for these data, and any distribution of porosity and permeability within the reported range and any elevation variations within the ranges lower than the seismic detection limit can be considered as a valid answer. Therefore, it is not suggested to make a general statement on the impact of a parameter on the plume match in Sleipner (and possibly other sites), based on the results from the OFAT approach.

The current work also helps to understand which of the addressed uncertain parameters for Sleipner model in literature should be prioritised and calibrated more carefully to improve the match. A similar study could be performed on any $CO_2$ storage or oil and gas site to find the importance of uncertain parameters before performing a history matching. After which, it is possible to minimise the mismatch between simulation and observed data more efficiently by improving the geological, operational and fluid properties. We hope this study helps Equinor to improve the Sleipner model in future releases. For instance, caprock morphology which was shown as the most important parameter in controlling plume migration in the current study, in the recent benchmark model results in a less accurate match in comparison to the previous model which needs further improvements.

## 5. Acknowledgements

The authors would like to thank the Centre for Fluid and Complex Systems at Coventry University for providing financial support for this project. The authors also wish to thank MathWorks for the use of Matlab, SINTEF Digital for the use of MRST and Equinor for providing the latest Sleipner benchmark model. Special thanks also to Dr Philip Costen for reviewing the paper and sharing his constructive comments with us.

## 6. References

Ahmadinia, M., & Shariatipour, S. M. (2020). Analysing the role of caprock morphology on history matching of Sleipner $CO_2$ plume using an optimisation method. *Greenhouse Gases: Science and Technology*.

Ahmadinia, M., Shariatipour, S. M., Andersen, O., & Sadri, M. (2019). Benchmarking of vertically integrated models for the study of the impact of caprock morphology on $CO_2$ migration. *International Journal of Greenhouse Gas Control, 90*, 102802.

Allen, R., Nilsen, H. M., Lie, K. A., Møyner, O., & Andersen, O. (2018). Using simplified methods to explore the impact of parameter uncertainty on $CO_2$ storage estimates with application to the Norwegian Continental Shelf. *International Journal of Greenhouse Gas Control, 75*, 198-213.

Alnes, H., Eiken, O., Nooner, S., Sasagawa, G., Stenvold, T., & Zumberge, M. (2011). Results from Sleipner gravity monitoring: Updated density and temperature distribution of the $CO_2$ plume. *Energy Procedia, 4*, 5504-5511.

Alnes, H., Eiken, O., & Stenvold, T. (2008). Monitoring gas production and $CO_2$ injection at the Sleipner field using time-lapse gravimetry. *Geophysics, 73*, WA155-WA161.

Arts, R., Eiken, O., Chadwick, A., Zweigel, P., Van der Meer, L., & Zinszner, B. (2004). Monitoring of $CO_2$ injected at Sleipner using time-lapse seismic data. *Energy, 29*, 1383-1392.

Bachu, S. (2003). Screening and ranking of sedimentary basins for sequestration of CO 2 in geological media in response to climate change. *Environmental Geology, 44*, 277-289.

Bachu, S. (2000). Sequestration of $CO_2$ in geological media: criteria and approach for site selection in response to climate change. *Energy conversion and Management, 41*, 953-970.

Baklid, A., Korbol, R., & Owren, G. (1996). Sleipner Vest $CO_2$ disposal, $CO_2$ injection into a shallow underground aquifer.

Bell, I. H., Wronski, J., Quoilin, S., & Lemort, V. (2014). Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. *Industrial & Engineering Chemistry Research, 53*, 2498-2508.

Bickle, M., Chadwick, A., Huppert, H. E., Hallworth, M., & Lyle, S. (2007a). Modelling carbon dioxide accumulation at Sleipner: Implications for underground carbon storage. *Earth and Planetary Science Letters, 255*, 164-176.

Bickle, M., Chadwick, A., Huppert, H. E., Hallworth, M., & Lyle, S. (2007b). Modelling carbon dioxide accumulation at Sleipner: Implications for underground carbon storage. *Earth and Planetary Science Letters, 255*, 164-176.

Boait, F., White, N., Bickle, M., Chadwick, R., Neufeld, J., & Huppert, H. (2012). Spatial and temporal evolution of injected $CO_2$ at the Sleipner Field, North Sea. *Journal of geophysical research: solid earth, 117*.

Carman, P. C. (1937). Fluid flow through granular beds. *Trans.Inst.Chem.Eng., 15*, 150-166.

Cavanagh, A. (2013). Benchmark calibration and prediction of the Sleipner $CO_2$ plume from 2006 to 2012. *Energy Procedia, 37*, 3529-3545.

Cavanagh, A. J., & Haszeldine, R. S. (2014). The Sleipner storage site: Capillary flow modeling of a layered $CO_2$ plume requires fractured shale barriers within the Utsira Formation. *International Journal of Greenhouse Gas Control, 21*, 101-112.

Cavanagh, A. J., Haszeldine, R. S., & Nazarian, B. (2015). The Sleipner $CO_2$ storage site: using a basin model to understand reservoir simulations of plume dynamics. *First Break, 33*, 61-68.

Chadwick RA, & Noy DJ. (2010). Chadwick RA; Noy DJ. History-matching flow simulations and time-lapse seismic data from the Sleipner $CO_2$ plume. *, 7*, 1171-1182.

Chadwick, R., Arts, R., Eiken, O., Kirby, G., Lindeberg, E., & Zweigel, P. (2004). 4D Seismic Imaging of an Injected C02 Plume at the Sleipner Field, Central North Sea. *Geological Society, London, Memoirs, 29*, 311-320.

Chadwick, R., Noy, D., Arts, R., & Eiken, O. (2009). Latest time-lapse seismic data from Sleipner yield new insights into $CO_2$ plume development. *Energy Procedia, 1*, 2103-2110.

Chadwick, R., Zweigel, P., Gregersen, U., Kirby, G., Holloway, S., & Johannessen, P. (2004). Geological reservoir characterization of a $CO_2$ storage site: The Utsira Sand, Sleipner, northern North Sea. *Energy, 29*, 1371-1381.

Chadwick, R. A., Noy, D., Arts, R., & Eiken, O. (2009). Latest time-lapse seismic data from Sleipner yield new insights into $CO_2$ plume development. *Energy Procedia, 1*, 2103-2110.

Court, B., Bandilla, K. W., Celia, M. A., Janzen, A., Dobossy, M., & Nordbotten, J. M. (2012). Applicability of vertical-equilibrium and sharp-interface assumptions in $CO_2$ sequestration modeling. *International Journal of Greenhouse Gas Control, 10*, 134-147.

Cowton, L., Neufeld, J., White, N., Bickle, M., Williams, G., White, J., & Chadwick, R. (2018). Benchmarking of vertically-integrated CO 2 flow simulations at the Sleipner Field, North Sea. *Earth and Planetary Science Letters, 491*, 121-133.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology, 26*, 297-302.

Dumont, M., MarÃ©e, R., Wehenkel, L., & Geurts, P. (2009). *Fast Multi-class Image Annotation with Random Subwindows and Multiple Output Randomized Trees.*

Edenhofer, O. (2015). *Climate change 2014: mitigation of climate change.* Cambridge University Press.

Eiken, O., Ringrose, P., Hermanrud, C., Nazarian, B., Torp, T. A., & Høier, L. (2011). Lessons learned from 14 years of CCS operations: Sleipner, In Salah and Snøhvit. *Energy Procedia, 4*, 5541-5548.

Furre, A., Eiken, O., Alnes, H., Vevatne, J. N., & Kiær, A. F. (2017). 20 Years of Monitoring $CO_2$-injection at Sleipner. *Energy procedia, 114*, 3916-3926.

Ghosh, R., Sen, M. K., & Vedanti, N. (2015). Quantitative interpretation of $CO_2$ plume from Sleipner (North Sea), using post-stack inversion and rock physics modeling. *International Journal of Greenhouse Gas Control, 32*, 147-158.

Goldberger, J., Roweis, S., Hinton, G., & Salakhutdinov, R. (2004). *Neighbourhood Components Analysis.*

Grataloup, S., Bonijoly, D., Brosse, E., Dreux, R., Garcia, D., Hasanov, V., Lescanne, M., Renoux, P., & Thoraval, A. (2009). A site selection methodology for $CO_2$ underground storage in deep saline aquifers: case of the Paris Basin. *Energy Procedia, 1*, 2929-2936.

Gregersen, U., & Johannessen, P. (2001). The Neogene Utsira Sand and its seal in the Viking Graben area, North Sea Saline Aquifer CO 2 Storage (SACS) project, Phase 2 Task 1.7 Geology. *Geological Survey of Denmark and Greenland Report, 100*, 1-2.

Han, W. S., Kim, K., Esser, R. P., Park, E., & McPherson, B. J. (2011). Sensitivity study of simulation parameters controlling $CO_2$ trapping mechanisms in saline formations. *Transport in Porous Media, 90*, 807-829.

Head, M. J., Riding, J. B., Eidvin, T., & Chadwick, R. A. (2004). Palynological and foraminiferal biostratigraphy of (Upper Pliocene) Nordland Group mudstones at Sleipner, northern North Sea. *Marine and Petroleum Geology, 21*, 277-297.

Hermanrud, C., Bolås, H. N., Teige, G., Nilsen, H., & KIær, A. (2012). Hermanrud C, Bolås HN, Teige GM, Nilsen HM, KIær AF. History-matching of $CO_2$ flow at Sleipner–new insight based on analyses of temperature and seismic data.

Hodneland, E., Gasda, S., Kaufmann, R., Bekkvik, T. C., Hermanrud, C., & Midttømme, K. (2019). Effect of temperature and concentration of impurities in the fluid stream on $CO_2$ migration in the Utsira formation. *International Journal of Greenhouse Gas Control, 83*, 20-28.

Holloway, S., Chadwick, R., Kirby, G., Pearce, J., Gregersen, U., Johannessen, P., Kristensen, L., Zweigel, P., Lothe, A., & Arts, R. (2000). Saline Aquifer $CO_2$ Storage (SACS)–Final report: Work Area 1 (Geology).

Huang, F., Juhlin, C., Han, L., Kempka, T., Norden, B., Lüth, S., & Zhang, F. (2015). Application of seismic complex decomposition on thin layer detection of the $CO_2$ plume at Ketzin, Germany.

IEA, P. (2012). Energy technology perspectives 2012: Pathways to a clean energy system.

Jones, R. R., McCaffrey, K. J. W., Clegg, P., Wilson, R. W., Holliman, N. S., Holdsworth, R. E., Imber, J., & Waggott, S. (2009). Integration of regional to outcrop digital data: 3D visualisation of multi-scale geological models. *Computers & Geosciences, 35*, 4-18.

Kaarstad, O. (1992). Emission-free fossil energy from Norway. *Energy Conversion and Management, 33*, 781-786.

Lindeberg, E., Zweigel, P., Bergmo, P., Ghaderi, A., & Lothe, A. (2001). Prediction of $CO_2$ distribution pattern improved by geology and reservoir simulation and verified by time lapse seismic. *Greenhouse gas control technologies, 372*, 377.

Lothe, A., & Zweigel, P. (1999). Saline Aquifer $CO_2$ Storage (SACS). Informal annual report 1999 of SINTEF Petroleum Research's results in work area 1 'Reservoir Geology'. *SINTEF*.

Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. , 4765-4774.

Lüth, S., Ivanova, A., & Kempka, T. (2015). Conformity assessment of monitoring and simulation of $CO_2$ storage: A case study from the Ketzin pilot site. *International Journal of Greenhouse Gas Control, 42*, 329-339.

Manceau, J., & Rohmer, J. (2014). Ranking importance of uncertainties for the assessment of residual and dissolution trapping of $CO_2$ on a large-scale storage site. *Energy Procedia, 63*, 3658-3664.

Manceau, J., & Rohmer, J. (2016). Post-injection trapping of mobile $CO_2$ in deep aquifers: Assessing the importance of model and parameter uncertainties. *Computational Geosciences, 20*, 1251-1267.

Møyner, O., Andersen, A., & Nilsen, H. M. (2018). Multi-model hybrid compositional simulator with application to segregated flow.

Møyner, O., & Nilsen, H. M. (2017). Multiresolution Coupled Vertical Equilibrium Model for Fast Flexible Simulation of $CO_2$ Storage. *Computational Geosciences 23.1 (2019): 1-20.*

Nilsen, H. M., Herrera, P., Ashraf, S. M., Ligaard, I. S., Iding, M., Hermanrud, C., Lie, K. A., Nordbotten, J. M., & Keilegavlen, E. (2011). Field-case simulation of $CO_2$-plume migration using vertical-equilibrium models. *Energy Procedia, 4*, 3801-3808.

Nilsen, H. M., Krogstad, S., Andersen, O., Allen, R., & Lie, K. A. (2017). Using sensitivities and vertical-equilibrium models for parameter estimation of $CO_2$ injection models with application to Sleipner data. *Energy Procedia, 114*, 3476-3495.

Nilsen, H. M., Lie, K. A., & Andersen, O. (2016a). Fully-implicit simulation of vertical-equilibrium models with hysteresis and capillary fringe. *Computational Geosciences, 20*, 49-67.

Nilsen, H. M., Lie, K. A., & Andersen, O. (2016b). Robust simulation of sharp-interface models for fast estimation of $CO_2$ trapping capacity in large-scale aquifer systems. *Computational Geosciences, 20*, 93-113.

Nilsen, H. M., 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.

Nilsen, H. M., Lie, K. A., Møyner, O., & Andersen, O. (2015). Spill-point analysis and structural trapping capacity in saline aquifers using MRST-co2lab. *Computers & Geosciences, 75*, 33-43.

Nooner, S. L., Eiken, O., Hermanrud, C., Sasagawa, G. S., Stenvold, T., & Zumberge, M. A. (2007). Constraints on the in situ density of $CO_2$ within the Utsira formation from time-lapse seafloor gravity measurements. *International Journal of Greenhouse Gas Control, 1*, 198-214.

Nordbotten, J. M., & Celia, M. A. (2011). *Geological storage of $CO_2$: modeling approaches for large-scale simulation.* John Wiley & Sons.

Onoja, M. U., Ahmadinia, M., Shariatipour, S. M., & Wood, A. M. (2019). Characterising the role of parametric functions in the van Genuchten empirical model on $CO_2$ storage performance. *International Journal of Greenhouse Gas Control, 88*, 233-250.

Pearce, J., Kemp, S., & Wetton, P. (1999). *Mineralogical and Petrographical Characterisation of a 1m Core from the Utsira Formation, Central North Sea: British Geological Survey Report WG/99/024.* British Geological Survey.

Pringle, J. K., Brunt, R. L., Hodgson, D. M., & Flint, S. S. (2010). Capturing stratigraphic and sedimentological complexity from submarine channel complex outcrops to digital 3D models, Karoo Basin, South Africa. *Petroleum Geoscience, 16*, 307-330.

Ringrose, P., & Oldenburg, C. (2018). Mission Innovation task force reports on enabling Gigatonne-scale $CO_2$ storage. *First Break, 36*, 67-72.

Rutqvist, J. (2012). The geomechanics of $CO_2$ storage in deep sedimentary formations. *Geotechnical and Geological Engineering, 30*, 525-551.

Santi A., Furre A.K., & Ringrose P. (2020). Sleipner 2019 Benchmark Model and Supporting documentation. *Equinor ASA*.

Shariatipour, S. M., Pickup, G. E., & Mackay, E. J. (2016). Simulations of $CO_2$ storage in aquifer models with top surface morphology and transition zones. *International Journal of Greenhouse Gas Control, 54, Part 1*, 117-128.

Shariatipour, S. M., Pickup, G. E., & Mackay, E. J. (2016). Investigation of $CO_2$ storage in a saline formation with an angular unconformity at the caprock interface. *Petroleum Geoscience, 22*, 203-210.

Singh VP, Cavanagh A, Hansen H, Nazarian B, Iding M, & Ringrose PS (2010). Singh VP; Cavanagh A; Hansen H; Nazarian B; Iding M; Ringrose PS. Reservoir modeling of $CO_2$ plume behavior calibrated against monitoring data from Sleipner, Norway.

Sørensen, T. J. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. København, I kommission hos E. Munksgaard.

Tin Kam Ho (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition, 1*, 278-282 vol.1.

Torp, T. A., & Gale, J. (2004). Demonstrating storage of $CO_2$ in geological reservoirs: The Sleipner and SACS projects. *Energy, 29*, 1361-1369.

Viebahn, P., Vallentin, D., & Höller, S. (2015). Prospects of carbon capture and storage (CCS) in China's power sector – An integrated assessment. *Applied Energy, 157*, 229-244.

Wei, N., Li, X., Wang, Y., Dahowski, R. T., Davidson, C. L., & Bromhal, G. S. (2013). A preliminary sub-basin scale evaluation framework of site suitability for onshore aquifer-based $CO_2$ storage in China. *International Journal of Greenhouse Gas Control, 12*, 231-246.

White, J. C., Williams, G. A., Grude, S., & Chadwick, R. A. (2015). Utilizing spectral decomposition to determine the distribution of injected $CO_2$ at the Snøhvit Field. *Geophysical Prospecting, 63*, 1213-1223.

White, J. C., Williams, G., Chadwick, A., Furre, A., & Kiær, A. (2018). Sleipner: The ongoing challenge to determine the thickness of a thin $CO_2$ layer. *International Journal of Greenhouse Gas Control, 69*, 81-95.

White, J. C., Williams, G. A., & Chadwick, R. A. (2013). Thin Layer Detectability in a Growing CO2 Plume: testing the Limits of Time-lapse Seismic Resolution. *Energy Procedia, 37*, 4356-4365.

Wilkinson, M., Haszeldine, R. S., Fallick, A. E., Odling, N., Stoker, S. J., & Gatliff, R. W. (2009). $CO_2$–mineral reaction in a natural analogue for $CO_2$ storage—implications for modeling. *Journal of Sedimentary Research, 79*, 486-494.

Williams, G., Chadwick, R., & Vosper, H. (2018). Some thoughts on Darcy-type flow simulation for modelling underground $CO_2$ storage, based on the Sleipner $CO_2$ storage operation. *International Journal of Greenhouse Gas Control, 68*, 164-175.

Williams, G. A., & Chadwick, R. A. (2017). An improved history-match for layer spreading within the Sleipner plume including thermal propagation effects. *Energy Procedia, 114*, 2856-2870.

Williams, G., & Chadwick, A. (2012). Quantitative seismic analysis of a thin layer of $CO_2$ in the Sleipner injection plume. *Geophysics, 77*, R245-R256.

Zhu, C., Zhang, G., Lu, P., Meng, L., & Ji, X. (2015). Benchmark modeling of the Sleipner $CO_2$ plume: Calibration to seismic data for the uppermost layer and model sensitivity analysis. *International Journal of Greenhouse Gas Control, 43*, 233-246.

Zweigel, P., Arts, R., Lothe, A. E., & Lindeberg, E. B. (2004). Reservoir geology of the Utsira Formation at the first industrial-scale underground $CO_2$ storage site (Sleipner area, North Sea). *Geological Society, London, Special Publications, 233*, 165-180.