Deliverable number: 5.4

Deliverable title: Workflow for reliability assessment

Organisation(s)	NGI
Author(s)	Jung Chan Choi, Yu Feng, Elin Skurtveit
Reviewer	Lars Grande, Steve Pearson, Tine Larsen
Type of deliverable	Report
Dissemination level	SHARP internal use
WP	5
Issue date	12.Feb.2024
Document version	Rev0

Keywords: Uncertainty, Fault stability assessment, Probability approach, Bayesian Update, Sensitivity analysis

Summary:

This report presents a workflow for probabilistic fault stability assessment in CO2 storage, emphasising uncertainty quantification, parameter ranking, and Bayesian updating for critical inputs. Using North Sea's Horda platform cases, including the Vette fault zone in Smeaheia, this study demonstrates the value of the suggested methods. This study highlights the usefulness of the probability of failure (Pf) as a reliable measure for stability assessment, particularly when traditional methods present conflicting results. It also highlights the effectiveness of the Sobol sensitivity analysis for input ranking and understanding parameter interactions, which is crucial for resource prioritisation and monitoring strategy in early field development. Furthermore, the report illustrates how the Bayesian approach can enhance the accuracy of stress prediction by leveraging data from geologically similar sites and acknowledging site-specific heterogeneity, significantly aiding in geomechanical risk assessments.

1 Introduction

This report summarises a workflow for probabilistic fault stability assessment in CO2 storage, emphasising uncertainty quantification, parameter ranking, and Bayesian updating for critical inputs, developed through SHARP WP5 (EC Project no. 691712).

1.1 Background

The heterogeneous nature of geological processes and the limited availability of measurable surface information trigger some degree of uncertainties in CO2 storage evaluation. Ongoing studies in SHARP also clearly indicate significant variations in in-situ stress conditions f (Grande et al., 2022) and mechanical behaviours (Grande et al., 2023) from the SHARP work on stress drivers and outliers. This observation raises a necessity for robust uncertainty quantification methods for the storage assessment, particularly in an early stage of field screening. As an initial effort to develop a workflow for quantitative risk assessment, SHARP report D5.1 (Bozorgzadeh, 2022) briefly introduces a method to quantify uncertainties involved in spatially averaged inputs required for geomechanics modelling. Following the previous initiative, this report summarises a probabilistic framework using a reliability-based approach to quantify the impact of the uncertainties involved in stress modelling and mechanical properties on the integrity of the CO2 sequestration site.

The developed workflow is anticipated to support other WPs in SHARP and in CCS projects to come. The outcomes of an initial analysis can be used to re-evaluate the key stress modelling inputs/uncertainties and their influence on subsequent predictions for WP1. Understanding the rank of the relative importance of inputs could also be used for planning for lab and field tests, which have been focused on WPs 2 and 3, respectively, and monitoring strategy for WP4. Then, the site-specific evaluation, including the probability of stability, could also be valuable input for upcoming deliverables on quantitative modelling of CO2 storage containment risks by WP5.3 and scientific guidance for quantifying risk by WP5.4.

1.2 Structure of the report

This report provides a comprehensive workflow, including probabilistic stability assessment and advanced sensitivity analysis to identify critical inputs, as well as novel techniques to reduce the uncertainty of critical input parameters.

Section 2 introduces a theory and method that is used for the workflow for probabilistic fault stability assessment together with a demonstration case for the Vette fault in the Smeaheia area. The section includes methods that can assess the stability of faults in a probabilistic manner and the variance-based sensitivity method, which can quantify the contribution of input uncertainties to the uncertainty in a probabilistic fault stability assessment. Then, the suggested method is demonstrated using a case study for the Smeaheia area.

Section 3 introduces a method that can reduce uncertainties of the identified critical parameters using the Bayesian hierarchical modelling (BHM) approach. This section will start with a brief explanation of how BHM enables logical borrowing of nearby historical stress data by accounting for between-site heterogeneity/similarity compared to the customary complete pooling approach. Then, the novelty of the method will be showcased through a case study for the Aurora and Smeaheia CO2 storage sites.

Then, Sections 4 and 5 will summarize the findings and limitations of the suggested workflow and address a further study at the end.

2 Probabilistic fault stability assessment

Fault stability is critical for the early screening process for CO2 storage sites. A quick and accurate way to assess it becomes more important. One of the challenges in screening unstable fault zones is how to handle uncertainties in input parameters and their impact on assessment results. Particularly, there is a lack of understanding of the relative importance of uncertainties in these parameters. This section introduces a theory and method that can be used for the workflow for probabilistic fault stability assessment, which is developed through SHARP WP5.

2.1 Fault stability assessment methods

Fault reactivation can occur when the stress acting on the fault exceeds its strength. Thus, fault reactivation analysis is a quantitative process comparing the fault stress to its strength. The stress acting on faults can be determined by decomposing surrounding principal stress conditions as illustrated in Figure 2-1a, and then the estimated normal and shear stresses acting on the faults can be compared to its failure criteria visually at the Mohr-Coulomb diagram as illustrated in Figure 2-1b. Stresses around the faults are affected by operations in reservoirs, like the production of oil and gas (pressure reduction) or CO2 injection (pressure increase). Particularly, during the CO2 injection, the increased pore pressure tends to move the Mohr circles toward the failure envelopes and can make the stresses around faults more critical with respect to failure criteria. Thus, the importance of robust assessment of fault reactivation analysis becomes more important for the CO2 injection site than the conventional depletion cases.

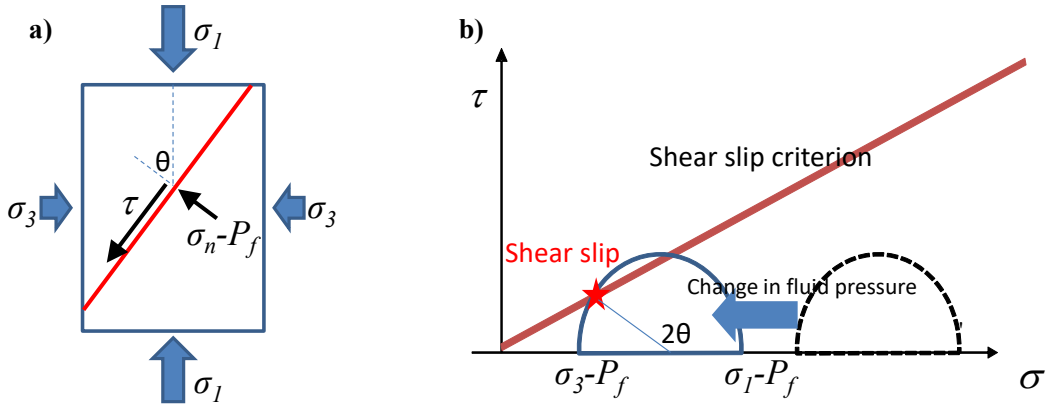

Figure 2-1 *(a) Schematic presentation of stresses on a fault plane and (b) illustration of the change in effective stresses due to injection on Mohr diagram.*

Stability assessment criteria provide quantitative thresholds or measures that we can use to predict the potential for fault reactivation under induced or natural changes in stress and pore pressure. This section introduces some criteria utilized in seismology and geomechanics for CO2 or geothermal reservoirs.

The allowable injection pressure P_c, which is also known as critical pressure perturbation (Wiprut and Zoback, 2000), represents the change in pore pressure ΔP that triggers a shear failure. The P_c can be defined as a horizontal distance between the stress acting on the faults and the failure envelope, as illustrated in Figure 2-2. Since it underestimates the allowable injection pressure due to its simplified stress path during the injection, which is a horizontal distance, it is often used as a conservative evaluation under a normal stress faulting regime. The P_c equation is expressed as follows:

$$\mathrm{P_c} = \sigma'_{n,fault} - \frac{\tau_{fault} - C_0}{\mu'} \tag{2.1}$$

Where $\sigma'_{n,fault}$ and τ_{fault} are the effective normal and shear stresses acting on fault, respectively, C_0 is the cohesion, and μ' is the effective friction coefficient. The effective friction coefficient μ' can also be esitmated using a tangent of effective friction angle φ' (i.e., $\mu' = \tan\ \varphi'$).

The mobilized shear strength τ_{mob} is defined as the ratio of the mobilized shear stress to its maximum resistance or strength, particularly to its friction coefficient (Choi, et al., 2023; Hettema, 2022). When the fault is stressed by its surroundings, the shear stress acting on the faults can increase only until its maximum possible resistance, which is shear strength. As the shear stress increases, the mobilized shear strength can thus reach to maximum 1.0. In geotechnical and structural engineering, the mobilized shear strength is widely used as a measure for stability evaluation that can show the relative distance between the stress and the strength (Ching and Phoon, 2013; Mesri and Shahien, 2003; Wong et al., 2007). Also, this term is an inverse of safety factor against a shear failure. The mobilized shear strength μ'_{mob} can be defined as:

$$\tau_{mob} = \frac{\mu'_{mob}}{\mu'} = \frac{\left[\dfrac{\tau_{fault}}{\dfrac{C_0}{\mu'} + \sigma'_{n,fault}}\right]}{\mu'} \tag{2.2}$$

For the cohesionless case, where $\mathrm{C_0} = 0$, the mobilized friction coefficient becomes the same as the slip tendency (ratio of shear stress to normal stress).

The allowable injection pressure is related to a future risk because it predicts the pore pressure increases needed for shear failure. In contrast, the mobilised shear strength evaluates the current stability status by comparing the current stress state to the material strengths. Thus, P_c can be used as a criterion for future conditions, while the mobilised shear strength can be used to evaluate the current status. However, in this study, we have also used the allowable injection pressure as a criterion for assessing the current status's probability. For that purpose, we have only used binary stability, which is fail or stable, as a criterion when the probability is calculated.

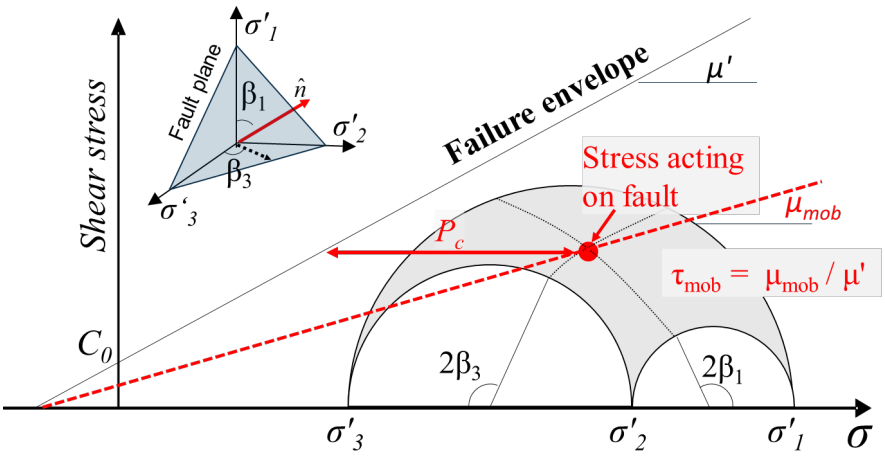

Figure 2-2 *Conceptual meaning of mobilized shear strength (τ_{mob}) and allowable injection pressure (P_c) illustrated in the 3-D Mohr circle diagram. The symbols in the figure are defined in the text (Choi et al., 2023)*

2.2 Probabilistic assessment methods

2.2.1 Workflow

Probabilistic fault stability assessment calculates the likelihood of fault reactivation or failure under uncertain geological and operation conditions. As illustrated in Figure 2-3, the assessment begins with defining a model for stability assessment. The model can be either a simple analytical solution, which is briefly covered in a previous section 2.1, or a complex numerical model. Selection of the model should be based on the availability of data during various stages of field development, the governing mechanisms of interest, and the overall risk level of interest. Then, essential inputs, including geomechanical parameters and operational conditions, need to be identified. SHARP report D5.1 (Bozorgzadeh, et al., 2022) briefly addresses methods to quantify the input uncertainties. In addition, Section 3 will introduce how the Bayesian approach can enhance the accuracy of input prediction.

Computational costs are one of the challenges in probabilistic assessment. Unlike deterministic analysis, which may require only a single calculation or a few, probabilistic analyses require extensive computations to cover a wide range of scenarios. Thus, the workflow involves a decision of model approximation. If the scale of the model exceeds a computational demand, some level of model approximation is essential to reduce the computational costs.

Subsequently or parallelly, a sensitivity analysis can be conducted to pinpoint the most influential parameters affecting fault stability. This analysis is crucial for understanding which factors to monitor and manage more closely to reduce uncertainties in model outputs or to approximate the model robustly (Razavi et al., 2021). Section 2.2.3 will briefly introduce a method that can quantify the contribution of input variance to the output variances.

Then, the probability of failure can be calculated through various reliability methods. This report briefly introduces a Monte Carlo Simulation and First Order Reliability method in Section 2.2.2.

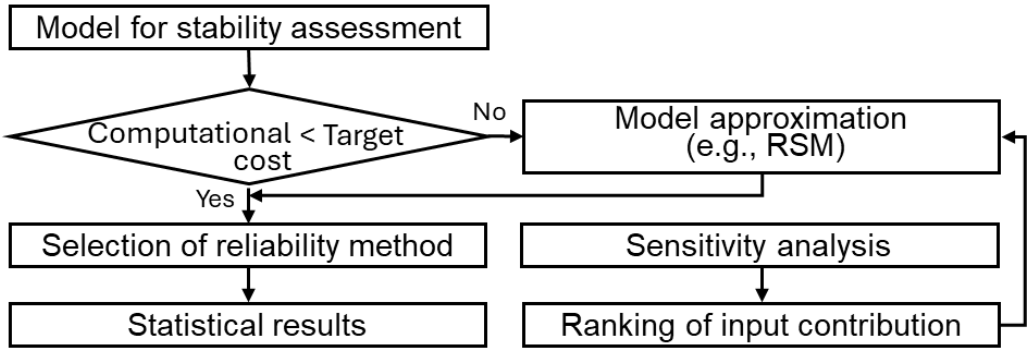

Figure 2-3 Workflow for probabilistic fault stability assessment

2.2.2 Reliability methods

2.2.2.1 Monte Carlo Simulation

Monte Carlo simulation (MCS) is a computational method that can simulate the response of randomly generated inputs. Particularly, this method is useful when assessing a problem with significant uncertainties in model inputs (e.g., material properties in subsurface materials). MCS randomly generates N inputs following its probabilistic distribution and performs a deterministic calculation using a model $f(X)$ for each set of random inputs where $X = [x_1, x_2, x_3, \dots, x_n]$. Then, the distribution of outputs can be calculated by aggregating the results of the deterministic calculation of each random set. For a stability assessment, MCS can calculate the probability of failure by counting the cases showing failure as expressed as:

$$\frac{1}{N}\sum_{m=1}^{N}[f(X_m) = Fail]$$

Where m denotes the mth sample, and [] is the Iverson bracket notation that can convert statements into 1 for True and 0 for False. For a fault stability assessment, the failure is determined by a criterion such as the mobilised shear strength exceeding a unity (i.e., $\tau_{mob} > 1$) or the allowable injection pressure smaller than a zero (i.e., $P_c < 0$).

Although MCS is a relatively straightforward and simple method, the computational cost can be high, particularly for cases with a large-scale deterministic model or low probable response. However, the computational cost can be optimised by adopting advanced techniques such as Latin hypercube sampling (McKay et al., 1979) or Importance sampling (Kloek and van Dijk, 1978). Also, the deterministic model can be approximated using a response surface method or other surrogate models (Sudret et al., 2017).

2.2.2.2 First-order reliability methods with Response surface methods

The First Order Reliability Method (FORM) (Hasofer et al., 1973) overcomes some limitations of Monte Carlo Simulation (MCS) primarily in terms of computational efficiency. MCS requires a large number of simulations to accurately estimate failure probabilities, especially for low-probability events, which can be computationally intensive. FORM, by transforming

the problem into a standard normal space and linearising the limit state function, simplifies the computation. This approach reduces the number of required simulations while still providing a reliable estimate of failure probability and sensitivity factors, making it more practical for complex engineering problems with multiple variables.

The starting point for FORM is the definition of the limit function $G(X)$, where X is the vector of basic random variables. The limit function must be defined such that $G(X) > 0$ means satisfactory limit and $G(X) \leq 0$ means failure. If the joint probability density function of all random variables $F_x(X)$ is known, then the probability of failure P_f is given by

$$P_f = \int_L F_x(X)\ \mathrm{d}X$$

where L is the domain of X where $G(X) \leq 0$. In general, the above integral cannot be solved analytically. In the FORM approximation, the vector of random variables X is transformed to the standard normal space U, where U is a vector of independent Gaussian variables with zero mean and unit standard deviation, and where $G(U)$ is a linear function. The probability of failure P_f is then (where $P[...]$ means "the probability that…"):

$$P_\mathrm{f} = P[G(U) < 0] \approx P[\sum_{i=1}^{n} \alpha_i U_i - \beta < 0] = \Phi(-\beta)$$

where α_i is the direction cosine of random variable Ui, β is the distance between the origin and the hyperplane $G(U) = 0$, n is the number of basic random variables X, and Φ is the standard normal distribution function. The vector of the direction cosines of the random variables (α_i) is called the vector of sensitivity factors, and the distance β from the origin to the "design point" is the reliability index.

The failure modes used as limit state functions for the fault stability analyses could be the mobilized shear strength larger than unity. Although the critical injection pressure is used to predict a future change of the pore pressure increases needed for shear failure, the binary output indicating the fail or stable can also be used as an indicator showing the current stability condition. Thus, the condition satisfying the critical injection pressure less than 0 is also used as a limit state function for the fault stability analysis.

If the scale of the model for the limit state function exceeds a computational demand, some level of model approximation, such as a response surface method (Box and Draper, 1987), is required to reduce the computational costs. As illustrated in Figure 2-4, the following steps are included in the probabilistic analysis:

- Quantify the uncertainty in input parameters and, if relevant, in the analysis method.
- Express the limit state function (in this case with the RSM approximation).
- Do the FORM analysis (comparable to millions of Monte Carlo simulations).
- Calculate the failure probability (P_f), reliability index (β) and sensitivity of P_f to each uncertain parameters.
- Check the coordinates of the "design point", which is the point on the limit state surface in the standard normal space that is closest to the origin, to ensure that the coordinates are consistent with the modelled situation.

The components of the normalized gradient vector at the design point can also be used as a sensitivity factor, which is known as the 'alpha factor. The magnitude of these factors indicates the relative importance of each variable in terms of its contribution to the failure probability.

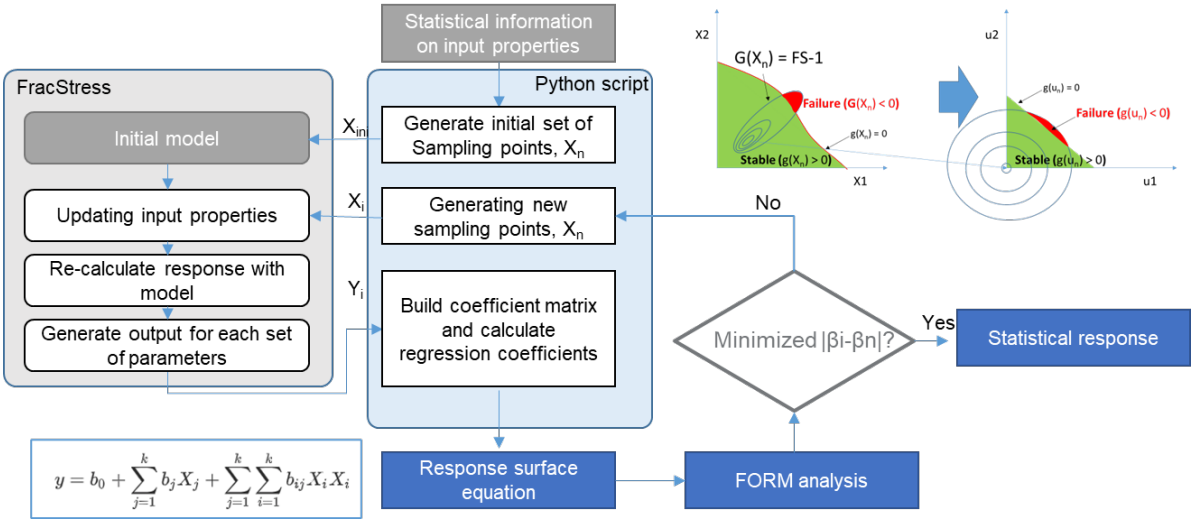

Figure 2-4 Workflow for FORM with response surface method

An illustration of the "design point" and a graphical representation of the reliability index β are given in Figure 2-5. The FORM approximation involves (1) transforming a general random vector into a standard Gaussian vector, (2) locating the point of maximum probability density (most likely failure point, "design point", or simply the β-point) within the failure domain, and (3) estimating the probability of failure as $P \approx \Phi(-\beta)$, in which $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function. The square of the direction cosines or sensitivity factors (α_i^2), which sum is equal to unity, quantifies in a relative manner the contribution of the uncertainty in each random variable X_i to the total uncertainty. Applications of the FORM approach to offshore geotechnical problems are given (Lacasse and Nadim, 2007).

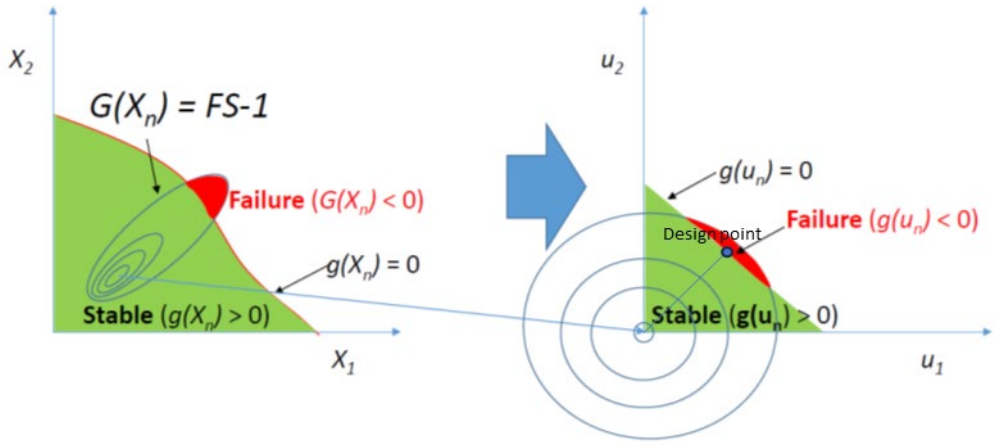

Figure 2-5 Illustration of First-Order Reliability Method (FORM) to calculate failure probability.

For a given probability distribution (e.g., normal, lognormal, triangular, exponential), there is a unique relationship between reliability index and failure probability. Figure D4 gives the relationship for a normally distributed function. For example, a reliability index (β-value) of 3.7 corresponds to a failure probability (P_f) of 10^{-4} and a β-value of 4.3 to a P_f of 10^{-5}.

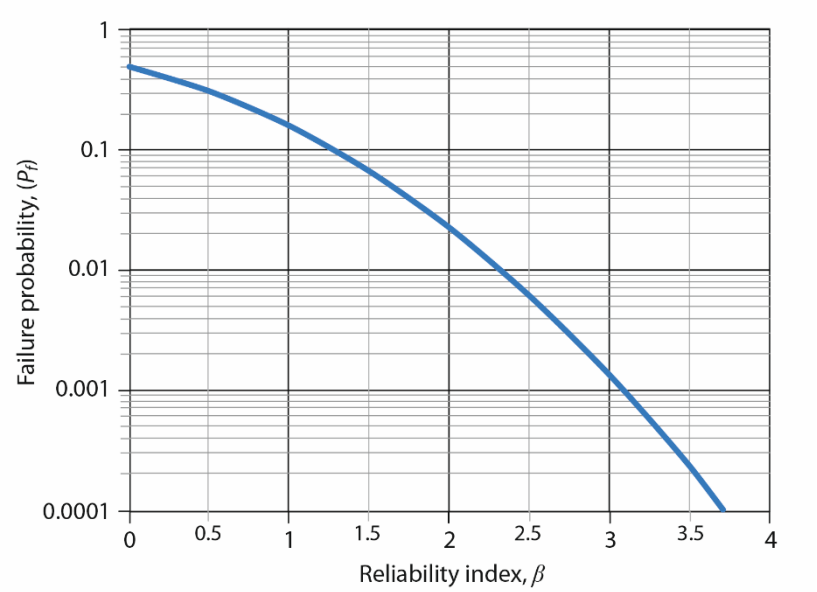

Figure 2-6 Relationship between probability of failure, P_f, and reliability index β (normal distribution).

2.2.3 Sensitivity analysis

Assessing fault stability is controlled by various factors, including the geometry of faults, initial stress conditions, mechanical properties of faults, operating conditions, etc. Thus, in order to stochastically assess the stability of faults, it is vital to have detailed information about the inputs required for the assessment. However, it is not easy or sometimes impossible to estimate the inputs required for the assessment accurately. Particularly, in an early stage of field development, due to limited information about the field condition, it is more challenging to estimate an accurate range of input parameters.

Sensitivity analysis can be a helpful tool that can assist in identifying the influence of input parameters and understanding the model when there is limited information on the various input parameters, such as in an early phase of field development. Particularly for a problem that needs to consider uncertainty in inputs using a stochastic or probabilistic analysis, it is very important to identify the contribution of input uncertainties to output uncertainties. Understanding the contribution of input uncertainties can reduce an unnecessary cost and effort to tune the less important parameters and consequently can optimize the financial cost for field development.

Variance-based sensitivity analysis, also often referred to as the Sobol method (Sobol′, 2001), is an efficient tool that can quantify the contribution of inputs to a variance in output (e.g., distribution of mobilized shear strength or allowable injection pressure). Since the variance-based sensitivity evaluates a sensitivity by varying entire parameters simultaneously, it is also known as global sensitivity analysis, in contrast to a local sensitivity method which changes parameters ‘one at a time’(Saltelli and Annoni, 2010). The Sobol method can evaluate the

sensitivity caused either by a single parameter or by the interaction of several parameters. Thus, it can effectively identify the relative importance of inputs even for the cases where inputs are widely spread and outputs are determined by interaction with other inputs. This section will briefly introduce the concept of variance-based sensitivity analysis. Then, the results of Sobol sensitivity analysis for the Smeaheia fault stability assessment will be presented in section 2.4.

In the Sobol sensitivity analysis, the total variance V of the model output Y is decomposed into components attributed to each input variable (Sobol′, 2001). The first-order effect S_i of an input variable X_i is defined as the contribution to the output variance by varying X_i alone, holding other inputs constant. This is calculated as:

$$S_i = \frac{V_{X_i}\left(E_{X_{\sim i}}(Y|X_i)\right)}{V(Y)}$$

Where $X_{\sim i}$ notation indicates the set of all variables except X_i, V_{X_i} is the variance across X_i and $E_{X_{\sim i}}(Y|X_i)$ is the conditional expectation (or mean) of Y when all input variables except for X_i are held constant at their mean values.

The contribution of the interaction between parameter X_i and X_j can be quantified by the second-order effect S_{ij}. The second order index indicates the additional variance in Y due to interaction between inputs. The total effect index S_{Ti} includes its first-order effect and all interaction effects. The total effects S_{Ti} can be calculated by:

$$S_{Ti} = 1 - \frac{V_{X_{\sim i}}\left(E_{X_i}(Y|X_{\sim i})\right)}{V(Y)}$$

This equation means the total effect index is one minus the ratio of the variance in the model output caused by all inputs except X_i to the total variance in the model output $V(Y)$.

The interaction effect can be captured by the difference between the total effect index ST and the first-order index S1, which is a contribution of a single parameter to the output variance. When the ST value is significantly higher than the S1 value for a parameter, it suggests that the parameter doesn't just influence the output on its own, but also has substantial interactions with other parameters. These interactions contribute additional variance to the output, which is captured in ST but not in S1. In simpler terms, a big difference between ST and S1 indicates that the parameter's impact on the output is amplified or modified when combined with changes in other parameters.

2.3 Demonstration case

Vette Fault Zone (VFZ) in the Smeaheia area, offshore Norway, was chosen as a demonstration case for this study. As illustrated in Figure 2-7, the Smeaheia area is located on the eastern part of the Horda Platform offshore Norway, part of the north-south-trending structural high on the eastern side of the Viking Graben. The Smeaheia area is a fault-bounded structure to the east,

north and west. The VFZ is a western bound of a structural trap of Alpha closure and a west-dipping normal fault with a prominent wedge-shaped growth section during the deposition of the Cromer Knoll Group (Mulrooney et al., 2020; Wu et al., 2021). The stability of VFZ has been evaluated by various researchers (Choi et al., 2023; Michie et al., 2021; Rahman et al., 2021; Skurtveit et al., 2018). Although the main messages of the previous studies for VFZ are that the risk of failure is low, there is still a need for a better understanding of the relative influence of the input parameters on the stability of the faults and their interactions on the fault stability. Thus, we have selected the VFZ as a case study for a probability assessment with Sobol sensitivity.

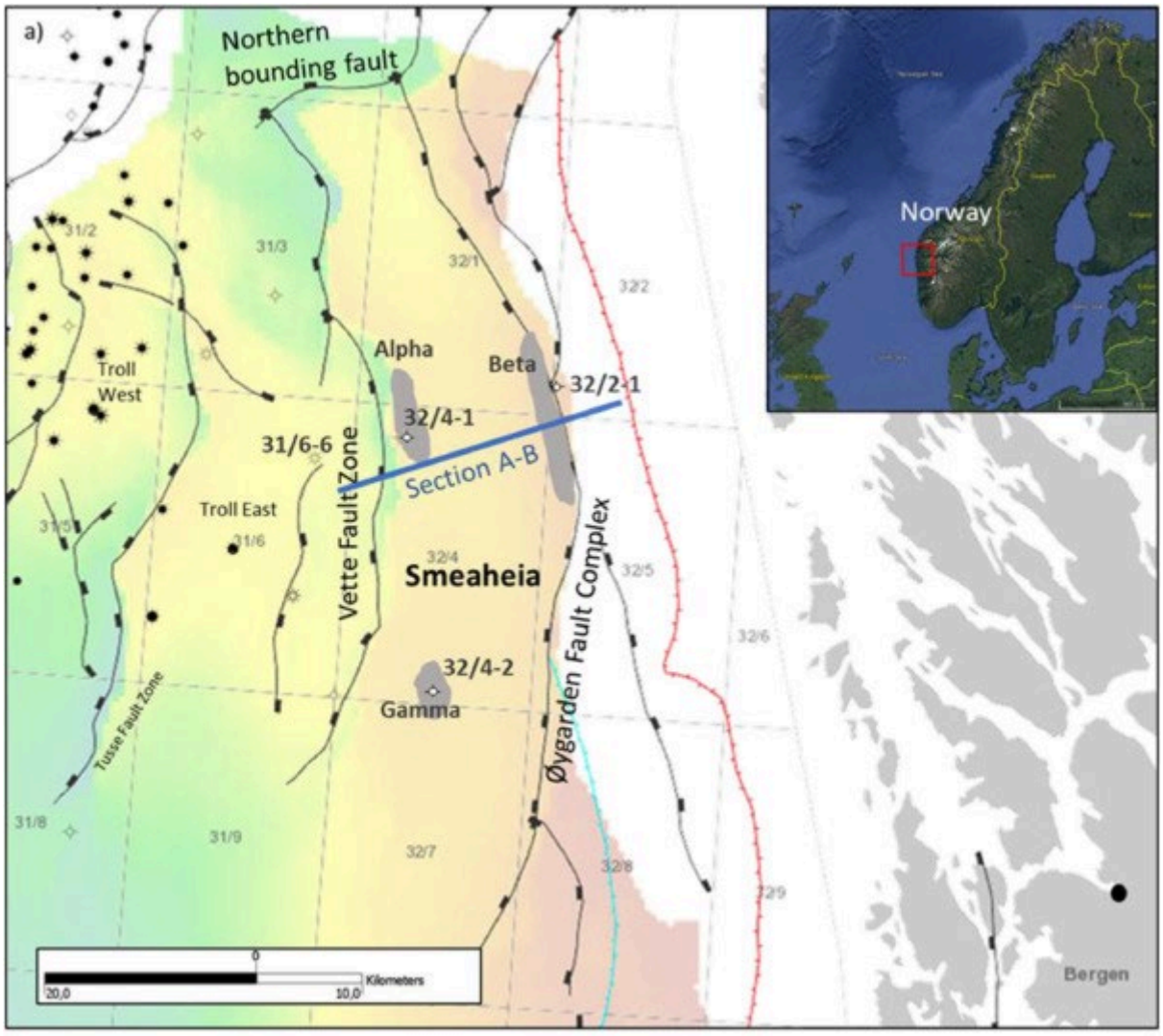

Figure 2-7 Location of the Smeaheia fault blocks on the Horda Platform (modified from Choi et al., 2023)

13 parameters were selected as random variables for a probabilistic analysis and its sensitivity assessment. Table 2-1 summarizes the input parameters considered in this study. 5 parameters (i.e., overburden gradient *OBG*, pore pressure gradient *PPG*, Coefficient of earth pressure at rest *K0*, Ratio of the maximum horizontal stress to the minimum horizontal stress *Sh_anisotropy_ratio*, and Azimuth of maximum horizontal stress *Azimuth_SH*) were related to the initial stress conditions. Fault geometries were determined by strike and dip, and the strength

parameters were selected as cohesion and friction angle. In order to consider the injection-induced stress changes, 3 parameters (i.e., Biot's coefficient α and reservoir stress paths for vertical and horizontal directions γ_h and γ_v) were considered for the study. The range of the inputs was selected from the data used for VFZ fault stability analysis (Choi et al., 2023; Michie et al., 2021; Rahman et al., 2021; Skurtveit et al., 2018) and the publicly available Smeaheia dataset (Gassnova, 2021). In addition, a recently published data set from the Horda platform (Thompson et al., 2022a) was used as a supplementary dataset. When the prior information of the input statistic is uncertain, the ranges in literature were assumed to be within the range of P10 and P90 with various distribution types. After the preliminary sensitivity, the sensitivity of assumptions in distribution were carried out to the influential parameters. For the probabilistic assessment, it is assumed that all input parameters are statistically independent. The independent assumption is primarily due to the lack of comprehensive understanding on the correlation between input parameters. Also, the sensitivity analysis will more focus on the relationship between inputs and the output, rather than the potential correlation between inputs.

Table 2-1 Input parameters used for the study

Variables	Unit	Mean	St.Dev.	CoV	Distribution type	References
OBG	**MPa/m**	2.10E-02	1.50E-03	7.1%	Normal	0.019 to 0.022 from Table 3-2 in Grande et al., 2023 (SHARP DV1.2)
PPG	**MPa/m**	1.03E-02	1.03E-03	10.0%	Normal	Close to hydrostatic with an error of less than 10% from Section 3.4.1 in Grande et al., 2023 (SHARP DV1.2)
K0	-	5.00E-01	1.00E-01	20.0%	Normal or Uniform*	0.41-0.63 from Table 3-4 in SHARP DV1.2
Azimuth_SH	**deg**	9.09E+01	2.13E+01	23.4%	Normal	87 to 267 N◦E but are strongly concentrated around 90 N◦E (East-West) is reported in Thompson et al. 2022a
sh_anisotropy_ratio	-	1.00E-01	1.00E-01	100.0%	Log-Normal (shifted)	1.01-1.27 from Thompson et al.2022a
Strike	**deg**	1.69E+02	2.92E+01	17.3%	Normal	Choi et al, 2023
Dip	**deg**	4.20E+01	1.00E+01	23.8%	Log-Normal	35-70 from Michie et al., 2021
Cohesion	**MPa**	1.00E+00	1.00E+00	100.0%	Log-Normal or Uniform*	0-2 Mpa from Thompson et al. 2022
Friction_angle	**deg**	3.00E+01	3.90E+00	13.0%	Log-Normal	25-35 from Thompson et al., 2022a
Reservoir stress paths for horizontal directions, γ_h	-	5.40E-01	1.40E-01	25.9%	Normal	effective horizontal stress path coefficinet ranges from -0.28 to - 0.64 from Choi et al, 2023
Reservoir stress paths for vertical directions, γ_v	-	0.00E+00	6.00E-02	-	Normal	effective vertical orizontal stress path coefficient has mean of -1.0 and SD of 3-6% from Choi et al, 2023
Biot's coefficient, α	-	9.00E-01	1.00E-01	11.1%	Uniform*	Thompson et al., 2022a estimated the range of 0.82-0.95 based on drained bulk modulus and Poisson's ratio. Other studies assumed it as

						1.0 (Choi et al., 2023, Rahman et al., 2022)
Depth of faults	**m**	900-1400			Deterministic	Gassnova 2021

* For the uniform distribution, the range is defined as [mean-std.,mean+std.]

2.4 Results

2.4.1 Probability of failure

Table 2-2 summarizes the results from a Monte Carlo Simulation (MCS) conducted to evaluate the critical injection pressure and mobilized shear strength at various depths and pressure changes. The table presents both deterministic and probabilistic results, reflecting the mean values as well as the 10th (P10), 50th (P50), and 90th (P90) percentiles of the distributions, alongside the probability of failure (Pf). The ΔPP in the table stands for the change in pore pressure caused by CO2 injection. As expected, an increase in injection pressure (i.e., $\Delta PP > 0$) decreases the critical injection pressure. However, the effect on critical injection pressure is less pronounced than expected. For instance, at a depth of 890m, an increase of 2.0 MPa in ΔPP leads to a reduction in the mean critical injection pressure of only ~0.5 MPa, which is from 6.30 MPa to 5.84 MPa, a change smaller than the applied pressure increment. This observation suggests that the computed critical injection pressures may be more conservative than actual critical values. In contrast, mobilized shear strength demonstrates a counter-intuitive trend. It becomes more stable (lower mobilized shear strength values) with increasing ΔPP. Similar counterintuitive results is also observed in the deeper depth. The increasing depth from 890m to 1400m leads to higher mean critical injection pressure showing more stable condition, but the mobilized shear strength values show opposite trends. This conflicting behaviour in critical injection pressure and mobilized shear strength underlines the need for a more representative measure of stability, as these metrics alone may not fully capture the complex dynamics of subsurface systems.

The calculated probability of failure (Pf) shows consistent values regardless of the specific stability criterion because it counts unstable cases. In deterministic assessments, critical injection pressure and mobilized shear strength both serve as indicators of proximity to failure. However, their different bases can lead to apparent inconsistencies in assessing how close a system is to failure. Despite this, they may still concur on a binary outcome—whether a system has failed. The probabilistic method gains its strength from the use of consistent 'fail' or 'stable' outputs across different failure criteria. This consistency in binary outputs, regardless of the definition, allows for a consistent evaluation of system stability. Furthermore, by analysing the distribution of these binary results, the probabilistic approach can quantify the proximity of failure. The consistency and proximity of failure present an advantage of probabilistic assessment over deterministic methods.

The calculated Pf indicates a general increase with depth, aligning more closely with the trends observed in mobilized shear strength. However, the impact of ΔPP on Pf varies with depth; while increasing injection pressure at shallower depths seems to increase the risk of failure, deeper systems exhibit a reduced Pf with increased ΔPP. As illustrated by the stress paths in Figure 2-8, although the same increase in ΔPP are considered with the same material properties, different depth can lead to different stress paths and associated differences regarding fault stability. Particularly, the assumption of a fixed horizontal stress path used in the allowable injection pressure may result in an error in the prediction of future stability. The consistent increase in Pf with depth, despite the contrasting effects on critical injection pressure and mobilized shear strength, underscores the importance of utilizing Pf as a key measure in fault stability assessments. However, this also highlights the need for a better understanding of the

individual contributions of these inputs and their complex interactions to effectively interpret and manage subsurface stability risks.

Table 2-2 Calculated results from MCS

Depth [msl]	ΔPP [MPa]	Critical_injection_pressure [MPa]						Mobilized shear strength [-]					
		P10	P50	P90	Mean	Std.	Pf	P10	P50	P90	Mean	Std.	Pf
890m	0.0	3.61	6.30	9.12	6.30	2.13	2.77E-04	0.14	0.32	0.52	0.32	0.15	2.77E-04
	1.0	3.47	6.02	8.82	6.09	2.07	2.94E-04	0.12	0.29	0.50	0.30	0.15	2.94E-04
	2.0	3.29	5.79	8.47	5.84	2.00	3.88E-04	0.11	0.26	0.48	0.28	0.15	3.88E-04
1400m	0.0	4.95	8.74	13.09	8.90	3.14	4.81E-04	0.15	0.33	0.56	0.35	0.16	4.81E-04
	2.0	4.69	8.35	12.43	8.47	2.99	4.64E-04	0.12	0.30	0.53	0.32	0.16	4.64E-04

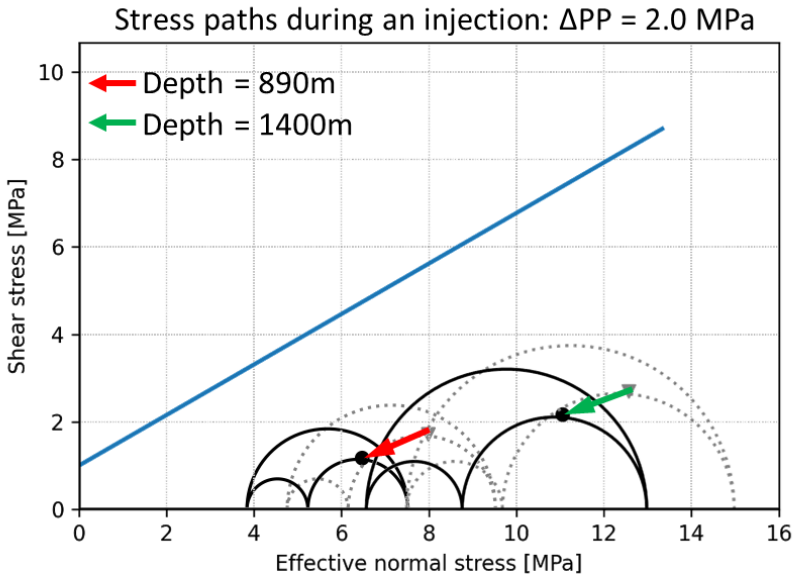

Figure 2-8 Stress paths at 890m (red) and 1400m (green) depths during a 2.0 MPa pore pressure increase, based on median parameters (P50).

2.4.2 Sensitivity analysis

Figure 2-9 shows the Sobol sensitivity indices (S1 for first-order effects and ST for total effects, which include interactions) for input variables affecting fault stability at 890m depth. Figure 2-9a) represent the scenario before injection and the figure b) shows the results after a 2MPa pressure increase. The results show that 'sh_anisotropy_ratio' (Horizontal total stress anisotropy) and 'K0' (in-situ effective horizontal stress ratio), and 'Cohesion' are key parameters with high ST values before the injection. For the post-injection case (Figure b), 'Cohesion' emerges as the most influential parameter. As addressed in the previous section 2.2.3, S1 values represent the independent effect of an input variable, while ST accounts both for the interaction with other variables and the independent contribution. Thus, the difference between S1 and ST typically indicates the contribution of interaction with other parameters. 'K0' and 'Cohesion' shows relatively small difference between S1 and ST, indicating its independent influence on the uncertainties in fault stability assessment. However, although 'Azimuth_SH' and 'strike' have relatively high ST, they have relatively larger differences

between S1 and ST values, indicating significant interaction effects. Also, 'sh_anisotropy_ratio' has higher difference between S1 and ST comparing to other influential parameters 'K0' and 'Cohesion'. The observed interaction contribution for directional parameters for stress 'Azimuth_SH' and fault geometry 'strike' as well as its magnitude 'sh_anisotropy_ratio' can be explained by a mechanism of stress acting on the fault plane. For example, if the fault is parallel to its maximum principal stresses, the stress acting on faults is purely governed by vertical and minimum horizontal stresses but not by the maximum horizontal stresses. In such cases, the uncertainties related to the maximum horizontal stress do not affect the fault stability. Thus, those uncertainties might become meaningful when they interact with other parameters. This observation highlights the novelty of Sobol' sensitivity analysis that can identify the relative importance of inputs and their interaction to the stability output.

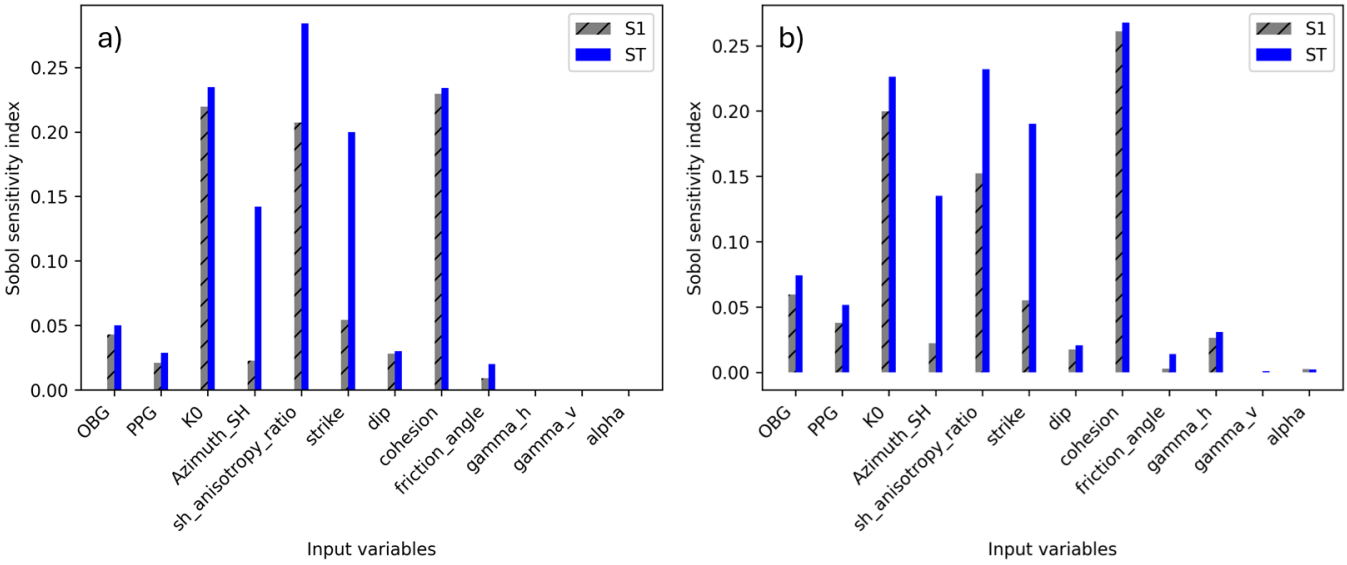

Figure 2-9 *Total and first order Sobol' sensitivity index for the case with depth of 890m for a) before injection and b) after 2MP of pressure increase*

Figure 2-10 compares the total Sobol sensitivity indices at two different depths, 890m and 1400m, each after a 2MPa pressure increase. As depth deeper, parameters associated with horizontal stresses, which are 'K0', 'Azimuth_SH', and 'sh_anisotropy_ratio', emerge as most influential. However, 'Cohesion', which shows the highest sensitivity at 890m with an ST of 0.2676, become less important at 1400m than the parameters related to horizontal stresses. In the previous MCS for Pf, the impact of ΔPP on Pf with depth were oppositely observed, which shows increasing the risk of failure with ΔPP at shallow depth but reduced Pf with ΔPP at deeper depth. The relative importance observed in Figure 2-10 may explain its trend. At the shallow depth, the strength parameter, particularly cohesion, governs the stability rather than stress acting on the faults. Thus, while injection-induced stress shifts a condition to a stress failure envelope, the injection-induced stress change does not influence the stability and can be more likely to make the fault stability more critical without having much room to change the stress. However, for the deeper depth, the stability of faults is mostly governed by stress parameters acting on the faults. In such a case, while injection-induced pore pressure changes the stress, the stress changes may have more room to change the relationship to its stability. That may partially explain counter-intuitive observation at the deeper depth.

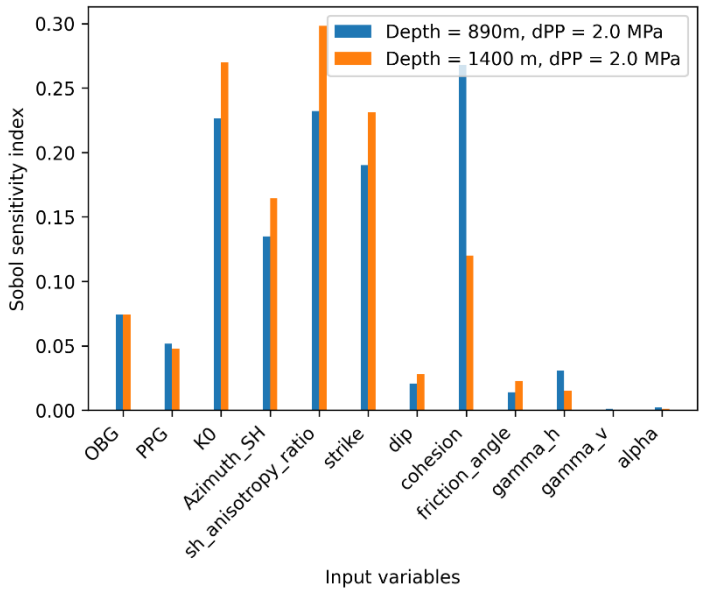

Figure 2-10 Effect of depth variation on total Sobol sensitivity index

2.5 Discussion

The results of this case study support the consensus of low probability of failure for the Vette fault zone, aligning with recent studies (Choi et al., 2023; Michie et al., 2021; Rahman et al., 2021; Skurtveit et al., 2018; Thompson et al., 2022a). However, although previous studies for Vette fault zone concluded similar stability ranges, those studies have focused on different aspects: Skurtveit et al. (2018) emphasised cohesion, (Michie et al., 2021) on fault-picking uncertainty, and Choi et al. (2023) on the reservoir stress path coefficient. Consequently, it was difficult to get an integrated overview considering the influence of other factors. However, the results of the sensitivity index suggested in this study (e.g., cohesion and horizontal stress parameters for this demonstration case) can clarify the parameters to be prioritised for further investigation or monitoring. The novelty of this study is thus the strategic use of sensitivity results with Pf for resource prioritisation in early field development. This information provides a clearer understanding of fault mechanics and reduces unnecessary costs and efforts to tune the less important parameters.

3 Bayesian modelling for site-specific stress prediction

3.1 Background

Reliable prediction of the in-situ minimum horizontal stress σ_h is crucial for the safe development and operation of CO_2 storage, as σ_h is an essential input parameter for assessing containment and induced seismic risks in the screening of CO_2 storage sites (Thompson et al., 2022aa; Wu et al., 2021, 2022). To infer the minimum horizontal stress at any given depth at a site, the customary approach is to perform a deterministic regression analysis of stress data as a function of depth and treat the line of best fit as the predicted site-specific stress profile (Andrews et al., 2016; Andrews and de Lesquen, 2019; Thompson et al., 2022b). This regression approach is the practical way to deal with the horizontal stresses in the industry,

however, in reality the horizontal stresses are depth dependent influenced by the lithology variations (DV3.2) and stress history including burial history, unloading from erosion and glacial loading and potential tectonics (SHARP reports DV1.1b and DV1.2.).

Stress data are often highly limited at CO_2 storage sites and may be even completely unavailable in the early site screening stages. For example, the Aurora CO_2 storage site under construction in the Norwegian North Sea has only five σ_h measurements available at best obtained using different measurement methods. Such limited stress data may significantly underrepresent the true depth-wise stress distribution at a storage site, and when coupled with other sources of uncertainty (e.g., measurement error and spatial variability), can reduce reliability in the site-specific stress predictions. In addition, stress uncertainty is the required input information in the more rational probabilistic risk assessment framework (e.g., probabilistic fault stability assessment in this report), because CO_2 storage integrity is mostly governed by weak locations (i.e., on faults or caprocks) that are related to conservative (high or low) estimates rather than the best estimates of uncertain geomechanical parameters. Hence, it is crucial to quantify and reduce uncertainty in site-specific stress prediction for CO_2 storage.

A natural approach to reduce stress prediction uncertainty is to augment the limited site-specific stress data for statistical analysis with prior information from other sources, such as nearby stress data, engineering judgement, geological indicators (e.g., faulting stress regime, frictional faulting theory) and indirect stress measurements (e.g., borehole breakouts, well logging and drilling-induced tensile fractures). Borrowing prior stress information may be inevitable when site-specific stress data are non-existent. Nevertheless, integrating prior information into statistical analysis of site-specific stress data is neither automatic nor straightforward. When facing limited data for a site of interest, the common practice is to either directly use the published stress trends from other sites/regions having richer data or expand the coverage area to include data from nearby sites. Such practice is usually made when different sites involved have similar geological settings (e.g., lithology, pore pressure and structures and segments if faulted areas) that allow assuming similar mean stress trends between sites. However, such semi-subjective borrowing of historical stress data, although effective in many cases, may lead to arbitrarily overconfident stress predictions when stress heterogeneity between sites is non-negligible.

The customary deterministic statistical analysis of stress data neither quantifies uncertainties about stress predictions nor offers a rigorous approach to integrating historical data for reducing the uncertainties. Unfortunately, there also seems a lack of studies in the literature on rigorous statistical integration of historical data into site-specific stress prediction in the context of CO_2 storage. Bayesian inference has been widely used as a rigorous and powerful statistical approach for quantifying uncertainty, as well as combining information from different sources via informative prior distributions. Hence, historical stress data may be integrated into Bayesian analysis of site-specific data in the form of prior distributions, with stress uncertainties being quantified and updated as the posterior distributions (Bao and Burghardt, 2022; Feng et al., 2023, 2021, 2020). Since posterior distributions from one Bayesian analysis can be used directly, or with modification, as prior distributions for a future analysis, informative priors are often constructed by Bayesian modelling of historical and other relevant data. When developing prior distributions for site-specific stress prediction, it may be tempting to follow the common

practice to perform a holistic Bayesian analysis of all stress data combined, yet such approach—statistically known as *complete pooling*—may result in overconfident borrowing of prior information and hence misleading stress predictions in that it ignores the stress heterogeneity between sites.

This section presents a *Bayesian hierarchical modelling (BHM) approach for quantifying and reducing uncertainties in site-specific stress predictions* and uses the Aurora and Smeaheia CO_2 storage sites to showcase how BHM enables logical borrowing of nearby historical stress data by accounting for between-site heterogeneity/similarity compared to the customary *complete pooling* approach. The proposed methodology provides more reliable uncertainty characterisation of stress predictions for the sake of robust probabilistic risk assessment related to CO_2 storage.

3.2 Study area and stress data

On the Norwegian continental shelf, extensive databases of field injection tests have been accumulated from drilling in oil and gas projects and sorted mainly into three categories; formation integrity tests (FITs), leak-off tests (LOTs), and extended leak-off tests (XLOTs) While XLOT data are considered the most accurate for estimation of minimum horizontal stress among the three types of injections tests, they are highly limited in numbers due to the great cost and normally kept confidential within energy companies. The Norwegian Petroleum Directorate (NPD) has made the LOT and FIT data publicly available through https://factpages.sodir.no/. Since LOT data literally account for the entire NPD database and are also considered more accurate than FIT, this report will not include the latter type of stress data. A collection of LOT data from the database for the entire Norwegian continental shelf has been made through previous studies (Choi et al., 2019), and further sorting and systematisation were done for selected wider Horda platform study area under SHARP DV1.2. The DV1.2 report discusses the influence of geology and stress history to explain the variations in observed stress trends and individual measurement points. For more discussion on uncertainties related to estimating the stress from the three types of injection test data, see the short summary in report DV5.1 and more detailed in the original references Raaen et al., 2006 and Andrews et al., 2016.

Figure 3-1 shows the study area (i.e., Section A and Section B) of the SHARP project surrounding the Aurora and Smeaheia storage sites in the Norwegian North Sea, and Section A as the main study area includes 10 sites (or fields): Martin Linge, Oseberg, Oseberg East, Huldra, Veslefrikk, Brage, Aurora, Troll West, Troll East and Smeaheia. All sites are situated in a relaxed sedimentary basin, and stress data are sampled far above the basement boundary. It can then be argued that gravitational loading dominates. However, they have quite different stress histories when it comes to burial diagenesis and tectonic history, including erosion, uplift and isostasy after episodes of glacial loading (see DV1.2). While stress data from all these potentially similar sites may be used to demonstrate the integration of historical data via BHM, here we restrict to the sites located on the same geological structure Horda Platform (HP) as the target storage sites, i.e., Aurora, Smeaheia, Brage, Oseberg HP (the part of Oseberg within HP), Troll East and Troll West. From a geological point of view, sites on the same geological structure (i.e., within HP) are expected to be more similar in their in-situ stress distributions

relative to sites on other structures in the same basin (i.e., Viking Graben and East Shetland Platform). The common assumption in the industry is hydrostatic pore pressure in most of the HP area, also at depth > 3 km, the assumption in the industry is hydrostatic, although with less confidence due to lack of deep LOT data (see discussion DV1.2). A local buildup of pore pressure and horizontal stresses may be relevant in certain formations, especially in low permeable tertiary units rich in smectite which can be high locally. Such effects have been discussed and analyzed in DV1.3. Pore pressure distributions at the excluded sites on other structures have excess pore pressure, especially below 3km, significantly impacting the in-situ horizontal stress (excess pore pressure from diagenesis with impact on minimum horizontal stress from poroelasticity). NPD database does not provide measurement data of the pore pressure other than in permeable formations (reservoir) typically a distance away from the LOT test. Hence, to reduce the data uncertainty from pore pressure, this analysis focuses on the HP where hydrostatic pressure is the governing assumption. It should be noted that data selection in any data analysis is a nontrivial task, and may involve an iterative process with justification to determine a reasonable set of data to be included in the analysis. More details regarding the wide study area and associated NPD database can be found in the SHARP report of DV1.2.

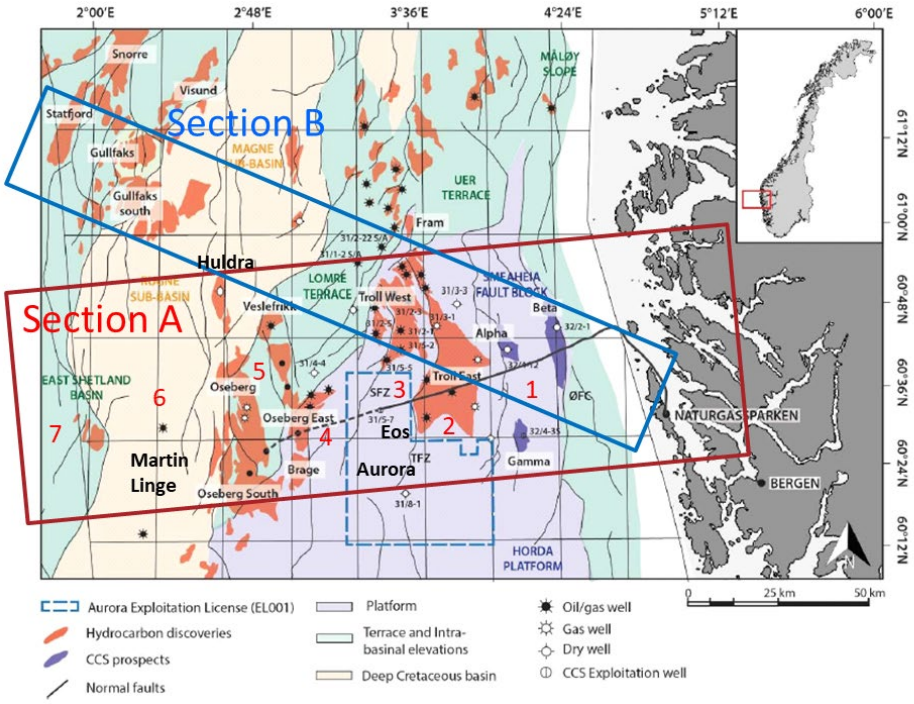

Figure 3-1 *Study areas A and B. The oil and gas fields and CCS sites are indicated. East Shetland Basin and Lomre Terrace (green), Viking Graben (yellow) and Horda Platform (violet). Structures/fault segments are indicated with numbers 1-7 (map from DV1.2).*

Figure 3-2 displays the data distribution of σ_h^c versus depth below seafloor (< 3000 m) for the six selected HP sites, where σ_h^c is the corrected minimum horizontal stress by subtracting the *water column* at seabed from σ_h. In this figure, the predominant LOT data are from the public NPD database, and the three high-quality XLOTs conducted at the two storage sites (two for Aurora and one for Smeaheia) are published by Equinor (Thompson et al., 2022b; Wu et al., 2022) and are also included here to augment the data at the storage sites, albeit still limited in

numbers. Note that while XLOT data are generally more accurate than LOT data, accounting for their respective measurement errors in the statistical analysis is nontrivial and beyond the scope of this report. Figure 3-2 shows that, although XLOT data has been included, the Aurora and Smeaheia storage sites still have only 5 and 3 stress measurements, respectively, while their nearby 4 sites on the same HP structure have as many as 77 data points. Indeed, with this limited stress data coupled with measurement errors and spatial variability, it is difficult to obtain meaningful stress profile predictions without external information. Hence, it is justifiable to integrate stress information from the nearby sites to constrain the stress predictions at the storage sites which are subject to a lack of data.

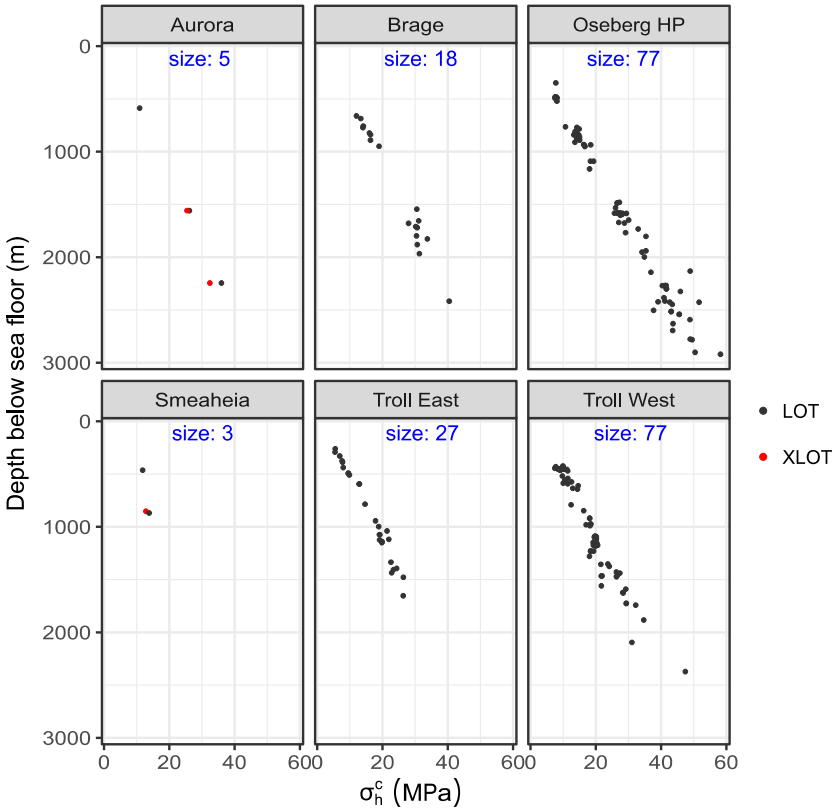

Figure 3-2 Stress data distribution from the 6 sites within the Horda Platform arranged by sites. σ_h^c is the corrected minimum horizontal stress by subtracting the water column at seabed from the total minimum horizontal stress σ_h.

3.3 Bayesian regression models

3.3.1 Basics

For data that can be structured into subgroups (e.g., site herein), there are three general statistical modelling approaches, namely, *complete pooling*, *no pooling* and *partial pooling*, details about which may be found in generic textbooks such as Gelman et al. (2013), with summary overviews for geotechnical applications by Bozorgzadeh and Bathurst (2022) and Feng et al. (2023). Here, only a brief introduction is provided.

The complete pooling model assumes a single set of common parameters for the whole dataset that comprises multiple subgroups, which is the customary modelling approach when attempting to integrate stress data from nearby areas as a means of augmenting the limited data at the target stie. This model ignores possible variations between groups. No pooling refers to fitting a model independently to each group of data. The no pooling model is often impractical for geotechnical data analysis since we usually do not have sufficient data per most groups (e.g., one site, one location, or one unit at one location) to allow for meaningful separate model fitting. As mentioned earlier, only three and five minimum horizontal stress measurements are available at the Smeaheia and Aurora storage sites, respectively, and in some cases stress data are not available under screening stage of CO2 storage sites (i.e Lisa, DK, see DV1.2 and DV3.2).

The complete pooling model assumes that group-specific model parameters are identical, while the no-pooling model assumes that they are independent. These two implicit assumptions are too strong considering the nature of geo-material and stress data. These assumptions can lead to under-exploitation of the data and, thus, suboptimal model performance. For instance, we would expect that the stress profiles at nearby sites within a structure like the HP area have some degree of similarity, rather than completely identical or different.

Partial pooling modelling (also called hierarchical modelling) strikes the middle ground between complete pooling and no pooling modelling in that it allows for group-varying model parameters which are assumed to be similar rather than necessarily identical or independent. This between-group similarity assumption, formally known as *exchangeability*, is implemented by assuming a higher-level population distribution with unknown hyperparameters for the group-specific parameters, thus allowing borrowing information across groups when estimating parameters (Gelman et al., 2013). Partial pooling can be thought of as a continuous generalisation of the two extreme cases of the complete pooling and no pooling modelling approaches: it reduces to the complete pooling and no pooling cases when the between-group variance goes to zero and infinity, respectively. Note that in this report, the partial pooling model and hierarchical model are referred to interchangeably.

3.3.2 Stress prediction equations

There exist a variety of 1D in-situ stress equations for predicting the magnitude of the minimum horizontal stress as a function of depth (see Zang and Stephansson, 2009 and Zhang et al., 2021 for an overview). The most widely used 1D stress prediction equation in geomechanical analysis is the so-called elastic-tectonic model assuming a constitutive relation of linear poroelasticity (Thiercelin and Plumb, 1994); this model has formulations for both isotropic and transversely isotropic rocks, with the simpler former formulation written as

$$\sigma_h = \frac{\nu}{1-\nu}\left(\sigma_v - \alpha P_p\right) + \alpha P_p + \frac{E}{1-\nu^2}\varepsilon_h + \frac{E\nu}{1-\nu^2}\varepsilon_H \tag{3.1}$$

where ν and E are respectively the Poisson's ratio and Young's modulus of rocks, P_p is the pore pressure, α is the Biot's coefficient, and ε_h and ε_H are the depth-uniform tectonic strains in the directions of the minimum and maximum horizontal stresses, respectively.

This elastic-tectonic model is a generalisation of the elastic uniaxial strain model considering the effect of only gravitational loading on horizontal stresses (Anderson et al., 1973), and reduces to such if the tectonic component is assumed to be zero. The advantage of this type of physics-based stress prediction model is that it explicitly accounts for individual stressing components (e.g., overburden stress, pore pressure and tectonic loading) as well as rocks' properties in determining the horizontal principal stresses, and when calibrated properly with relevant measured data, it has the flexibility to give reasonable stress predictions in various geological settings.

Unfortunately, the NPD stress database does not come with measurements of pore pressure and rocks' poroelastic properties, so this report adopts the simple stress prediction equation as a function of depth directly. Such an empirical stress prediction approach is also common in practice, especially when the stress scale of interest is large (e.g., Andrews et al., 2016; Andrews and de Lesquen, 2019; Breckels and van Eekelen, 1982; Thompson et al., 2022a). Here, we use a linear function as the stress prediction equation to demonstrate BHM for integrating stress information from different sites, which is written in the form as

$$\sigma_{\mathrm{h}} = \beta_0 + \beta_1 \mathrm{d} + \gamma\left(P_{\mathrm{p}} - P_{\mathrm{pN}}\right) \tag{3.2}$$

where d is the depth (below sea floor), P_{p} is the measured pore pressure, P_{pN} is the hydrostatic pore pressure, γ is the pore pressure/stress coupling coefficient accounting for the effect of non-hydrostatic pore pressure on the total minimum horizontal stress (Breckels and van Eekelen, 1982; Hillis, 2001), β_0 is the unknown intercept parameter which can be considered as the tectonic stress component in the minimum horizontal stress direction, and β_1 is the unknown slope parameter. Because the pore pressure on HP is assumed to be hydrostatic, the last term of Eq. (3.2) becomes zero.

It should be noted that the linear function in Eq. (3.2) could also be replaced with a power function or a polynomial function of depth to account for the possible nonlinear depth variation of the underlying stress states. Again, there are many available 1D stress prediction equations based on physics and/or empiricism. Since this report focuses on demonstrating how BHM allows logical borrowing of stress information from nearby sites, the task of identifying the best stress prediction equation among available candidate models is beyond the scope of this report. Nonetheless, the formal Bayesian model comparison approach presented in Section 3.4.3 can be applied to this task if it is of interest. It is worth noting that the choice of the stress prediction equation is usually subject to the availability of various data types, geological understanding, and project needs.

3.3.3 Complete pooling model

In the complete pooling model, the parameters are assumed to be identical across different sites, i.e., a set of global parameters that do not vary between sites. Hence, the complete pooling regression model based on the prediction equation of Eq. (3.2) can be written for data points $i = 1, \ldots, N$ as

$$\sigma^{\mathrm{c}}_{\mathrm{h}(i)} = \mathbb{E}\left[\sigma^{\mathrm{c}}_{\mathrm{h}(i)}\right] + e_i = \beta_0 + \beta_1 d_i + e_i \tag{3.3}$$

$$e_i \sim \mathrm{Normal}(0, \tau^2) \tag{3.4}$$

where $\mathbb{E}[\cdot]$ denotes expected (mean) value, $\{\beta_0, \beta_1\}$ are the two regression coefficient parameters; e_i is the residual error reflecting the discrepancy between the measured stress value and the predicted value by the adopted stress prediction equation, and is assumed to follow a zero-mean normal distribution with common variance τ^2.

Equations (3.3) and (3.4) constitute the likelihood function for observations of minimum horizontal stresses. The model parameters $\{\beta_0, \beta_1, \tau\}$ require specification of prior distributions in the Bayesian framework. Throughout this report, weakly informative priors are adopted for the parameters of all models to reflect the absence of prior knowledge about their specific values. Here, the following weakly informative priors are assigned:

$$\beta_0 \sim \text{Normal}(0, 4^2) \tag{3.5}$$

$$\beta_1 \sim \text{Normal}(0, 0.1^2) \tag{3.6}$$

$$\tau \sim \text{Half-Normal}(0, 4^2) \tag{3.7}$$

where ς is assumed a half-normal prior distribution to enforce its strictly positive range. The above priors express weak prior knowledge that the values of the three parameters $\{\beta_0, \beta_1, \tau\}$ will most likely (with roughly 0.95 probability) fall between (-8, 8), (-0.2, 0.2) and (0, 8), respectively. The weakly soft constraints on the parameters implied by these priors are reasonable considering the realistic ranges of the data; the same reasoning applies to the priors chosen for the Bayesian partial pooling (hierarchical) model in the report.

With priors and likelihood function specified, the posterior distribution can be formulated based on Bayes' rule as the product of the priors and likelihood up to a normalising constant. In most cases, posterior distributions of parameters are analytically intractable and numerically approximated using the simulation method known as Markov chain Monte Carlo.

3.3.4 Partial pooling model

The partial pooling (i.e., hierarchical) model explicitly accounts for between-group variations by assuming a population with unknown parameters for the (exchangeable) group-specific parameters, thus allowing for the borrowing of information amongst them. The partial pooling model chosen in this report assumes exchangeable intercepts β_0 and slopes β_1 with ς being set as identical across sites.

Suppose a given stress dataset can be grouped into J sites, with each site j ($j = 1,2, \dots, J$) having N_j data points. The partial pooling model can be formulated for data points $i = 1,2, \dots, n_j$ at sites $j = 1,2, \dots, J$ as

$$\sigma^{c}_{h(ij)} = \mathbb{E}\left[\sigma^{c}_{h(ij)}\right] + e_{ij} = \beta_{0(j)} + \beta_{1(j)} d_{ij} + e_{ij} \tag{3.8}$$

$$e_{ij} \sim \text{Normal}(0, \tau^2) \tag{3.9}$$

where $\beta_{0(j)}$ and $\beta_{1(j)}$ are the model parameters specific to site j (not common parameters across sites). Their respective population distributions to encode the exchangeability assumption to allow for between-site similarity/heterogeneity are specified as a common normal prior distribution with unknown hyperparameters (i.e., parameters of priors) as

$$\beta_{0(j)} \sim \text{Normal}\left(\mu_{\beta_0}, \tau^2_{\beta_0}\right) \quad (3.10)$$

$$\beta_{1(j)} \sim \text{Normal}\left(\mu_{\beta_1}, \tau^2_{\beta_1}\right) \quad (3.11)$$

where μ_{β_0} and $\tau^2_{\beta_0}$ are the mean and variance of the assumed population of site-varying β_0s, and μ_{β_1} and $\tau^2_{\beta_1}$ are the mean and variance of the assumed population of site-varying β_1s. The four hyperparameters are assigned the following weakly informative hyperpriors (i.e., priors on hyperparameters) as

$$\mu_{\beta_0} \sim \text{Normal}(0, 4^2) \quad (3.12)$$

$$\tau_{\beta_0} \sim \text{Half-Normal}(0, 1^2) \quad (3.13)$$

$$\mu_{\beta_1} \sim \text{Normal}(0, 0.1^2) \quad (3.14)$$

$$\tau_{\beta_1} \sim \text{Half-Normal}(0, 0.025^2) \quad (3.15)$$

The weakly informative prior for the standard deviation τ of the residual errors is assigned the same as the complete pooling model as

$$\tau \sim \text{Half-Normal}(0,\ 4^2) \quad (3.16)$$

3.4 Results

In this report, all Bayesian models are fitted using the probabilistic programming language Stan (Stan Development Team, 2023) which uses the No-U-Turn Sampler (Carpenter et al., 2017; Hoffman & Gelman, 2014) for Markov chain Monte Carlo simulation. Three Markov chains, each consisting of 3000 simulated posterior draws, are run in parallel to assess their convergence and approximate parameters' posterior distributions. Simulated posterior distributions are usually summarised by their means and spreads (e.g., credible intervals (CIs) and standard deviations (SDs)) that are taken as the best estimates and associated uncertainties of the parameters, respectively.

3.4.1 Posterior estimation

Figure 3-3 illustrates the posterior estimates of the two regression coefficients β_0 and β_1 (i.e., intercept and slope parameters) from the customary complete pooling model and the proposed partial pooling (hierarchical) model. It is obvious that the complete pooling estimates of both β_0 and β_1 exhibit remarkably less uncertainties as indicated by the narrower 95% CIs compared to the partial pooling estimates. However, as will be shown later, these parameter estimates are overly confident due to the disregard for the between-site stress heterogeneity (i.e., a common set of model parameters for all sites) when integrating stress data from different sites, hence leading to sub-optimal stress predictions.

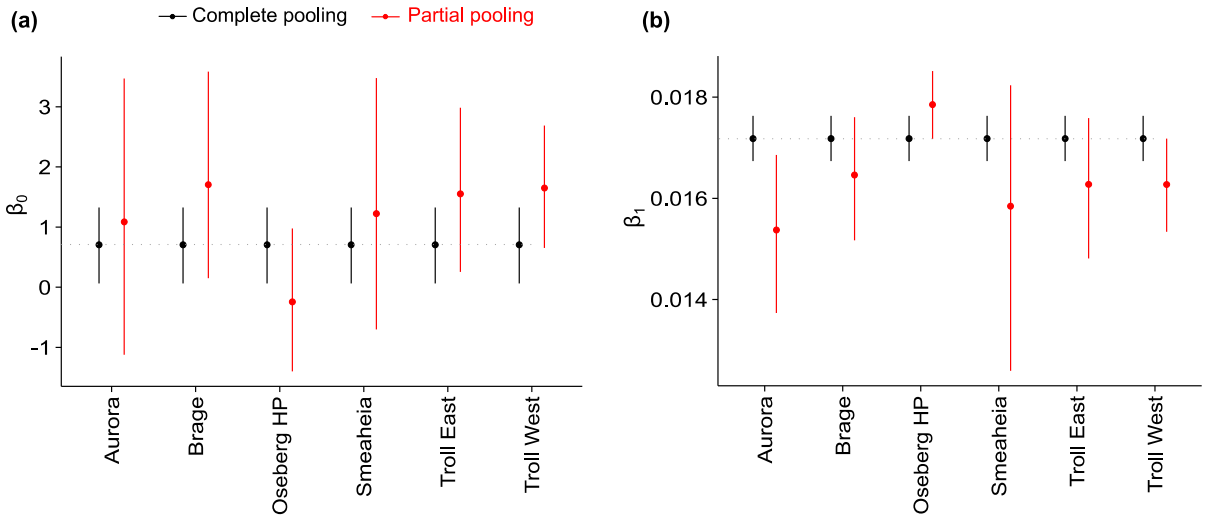

Figure 3-3 Posterior means and 95% credible intervals (CIs) of model parameters from the complete pooling and partial pooling (i.e., hierarchical) models

Figure 3-4 shows the posterior means and 95% CIs of the fitted minimum horizontal stress σ_h^c against the measured values obtained from the complete and partial pooling models. It can be seen that both models give a reasonably good fit to the measured stress data, particularly to smaller stress values at shallower depths. The relatively poorer fit to the larger stress values may be attributed to the deviation from the assumed hydrostatic pore pressure at greater depths. This figure also shows that the complete pooling model gives smaller uncertainties (narrower 95% CIs) in the fitted values propagated from its overconfident parameter uncertainties, the partial pooling model yields fitted σ_h^c values (i.e., posterior means) that are notably closer to the measured values for the interested Aurora storage site.

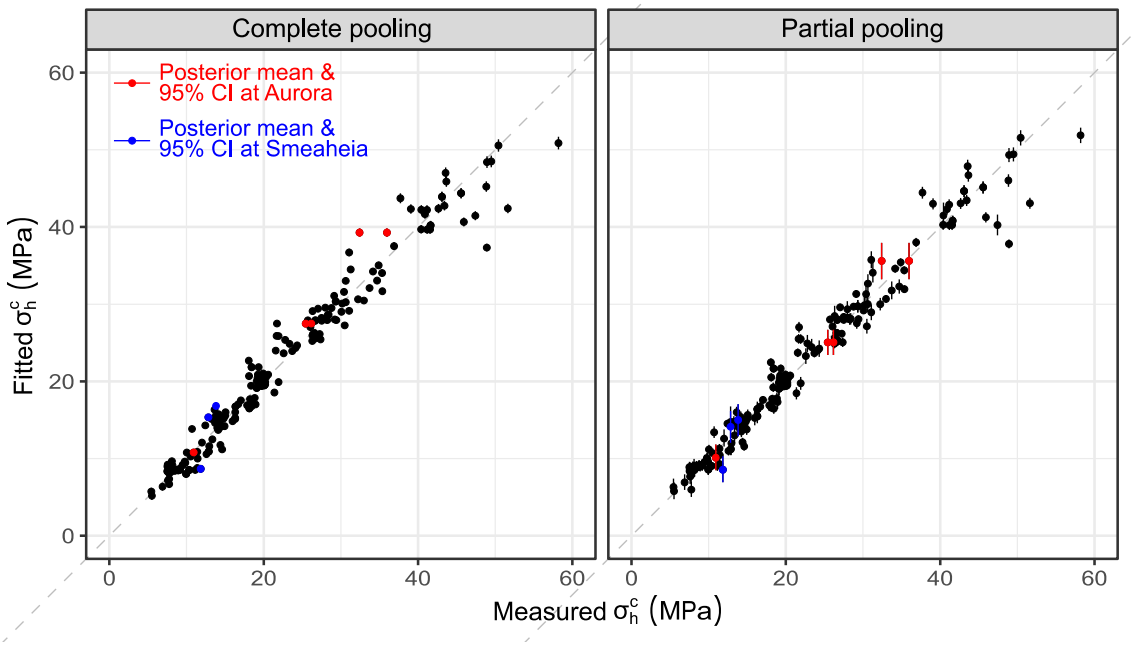

Figure 3-4 Posterior means and 95% credible intervals (CIs) of fitted σ_h^c versus measured σ_h^c from the complete pooling and partial pooling (i.e., hierarchical) models

3.4.2 Posterior prediction

Figure 3-5 displays the posterior predictions (after observing site-specific data) of the minimum horizontal stress profiles along with their 90% predictive intervals (PIs) obtained from the complete and partial pooling models. It can be seen that for the Aurora site, the complete pooling model generally over-predicts all measured stress data, particularly for the two data points at great depths which are not even captured by the 90% PIs. On the other hand, the partial pooling model is able to adapt to all stress data including the two deep stress measurements. These two models in question also exhibit a substantial difference in the predicted stress profiles at greater depths at the Smeaheia site where there is only limited data, while the two sets of predictions at the other four sites with abundant data are practically the same. Such observation indicates that the complete pooling model cannot reasonably adapt to the limited local data as it considers all sites to be statistically identical and thus integrates a large number of data from other sites directly without any discounting, while the partial pooling model borrows more appropriate information from other sites by accounting for the between-site similarity (also heterogeneity) in the data.

It should be noted that the linear trend in Figure 3-5 is only valid near the caprocks where the data points exist. The high stresses at the sea floor may not reflect the field condition and should not be used. In the further study, the regression function could also be replaced with a power function or a polynomial function of depth including the origin at the sea floor to account for the possible nonlinear depth variation of the underlying stress states.

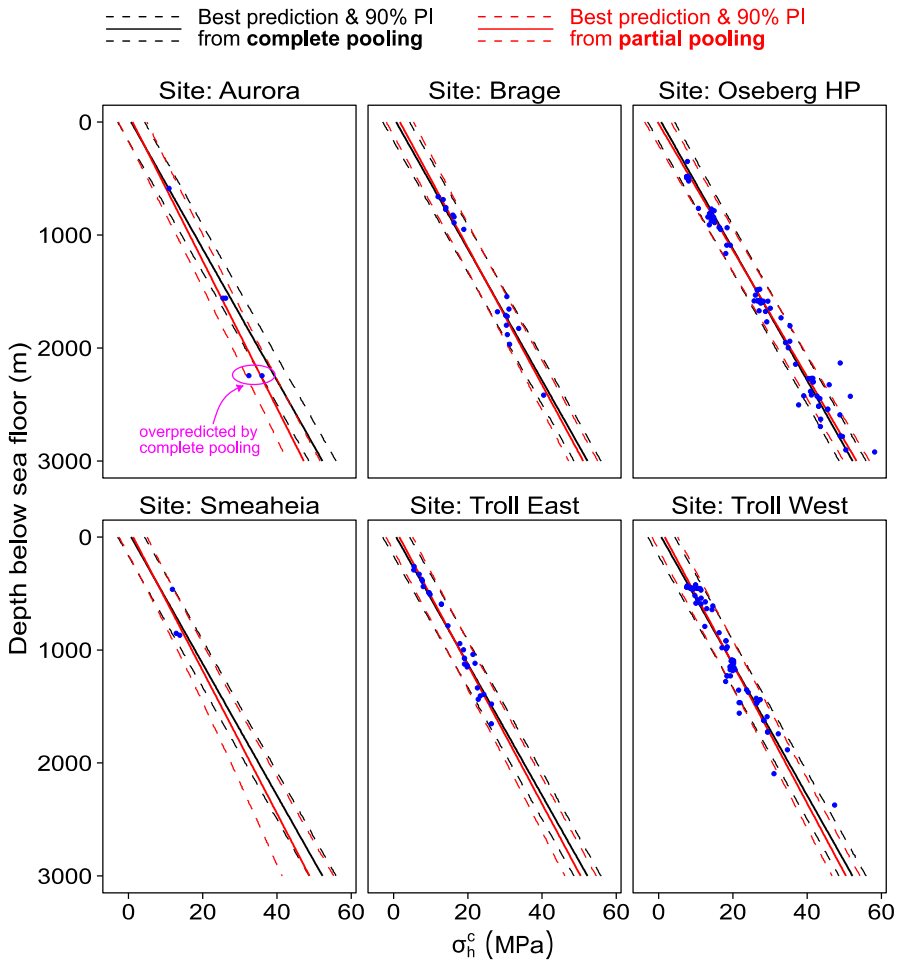

Figure 3-5 Predicted profiles of the minimum horizontal stress from the Bayesian complete pooling and partial pooling (hierarchical) models. (The solid black line represents the best predictions of stress states along depth, and two dotted lines represent the predictive uncertainty range of stresses with a probability of 0.9)

3.4.3 Model comparison based on out-of-sample prediction

This section provides a formal comparison of the Bayesian complete pooling and partial pooling models in question. Statistical models are usually compared based on their out-of-sample predictive accuracy, which can be estimated by two general approaches, namely, cross-validation and information criteria (e.g., AIC and DIC) being approximations to different versions of the former approach. For probabilistic models, a common measure of predictive accuracy is the expected log predictive density (elpd), and larger elpd indicates higher expected predictive accuracy. In this study, we use the widely advocated and most accurate leave-one-out cross-validation (LOO-CV) approach to estimate the elpd for all Bayesian regression models.

Here, we briefly introduce the LOO-CV method for Bayesian model comparison, and refer readers for more details to the standard statistics texts (Gelman et al., 2013, 2014; Vehtari et al., 2017) and relevant geotechnical and geoscience literature (Bozorgzadeh & Bathurst, 2019; Feng et al., 2023). Suppose we have a dataset $\{y_1, \dots, y_n\}$ that is modelled using a Bayesian model conditioned on parameters θ. LOO-CV estimates the elpd and its standard error (SE) by iteratively fitting the model to the held dataset y_{-i} after leaving out one data point y_i:

$$\mathrm{elpd}_{\mathrm{LOO}} = \sum_{i=1}^{n} \log p(y_i|y_{-i}) = \sum_{i=1}^{n} \log \int p(y_i|\theta)p(\theta|y_{-i})\,\mathrm{d}\theta \tag{3.17}$$

$$\mathrm{SE}(\mathrm{elpd}_{\mathrm{LOO}}) = \sqrt{n\mathrm{V}_{i=1}^{n}[\log p(y_i|y_{-i})]} \tag{3.18}$$

where $\log p(y_i|y_{-i})$ and $p(y_i|\theta)$ are the pointwise log predictive density and the likelihood for the left-out data point y_i, respectively, $p(\theta|y_{-i})$ is the posterior distribution of θ from the model fitted to y_{-i} and $\mathrm{V}(\cdot)$ denotes the variance operator. Note that the standard error comes from considering the observed data of size n as a sample from the underlying data population. In this study, it is the measured minimum horizontal stresses that are left out for comparison to their predicted values.

Table 3-1 summarises the elpd estimates for both statistical models and their differences (i.e., Δelpd) relative to the largest elpd estimate of −449.6 from the proposed partial pooling model. It is shown that the partial pooling model has a significantly better predictive performance than the baseline complete pooling model, as manifested by their elpd difference of −8.0 being nearly 2 times larger than its standard error of 4.2. Hence, from a viewpoint of out-of-sample predictive accuracy (generalisation performance) about stresses, the proposed partial pooling model via BHM should be preferred over the customary complete pooling model.

Table 3-1 LOO-CV model comparison

Model	Parameter configuration	elpd(SE)	Δelpd(SE)
Partial pooling	Exchangeable β_0 and β_1, and complete pooling ς across sites	−449.6(22.1)	0
Complete pooling	Complete pooling all model parameters across sites	−457.6(20.6)	−8.0(4.2)

elpd: expected log predictive density

Δelpd: difference in elpd between each model and the model with largest elpd

SE: standard error of estimated elpd

4 Conclusion

This study introduces a comprehensive workflow for probabilistic fault stability assessment for a CO2 storage, focusing on quantifying uncertainties in assessment, ranking the importance of various parameters affecting fault stability, and Bayesian updating for reducing uncertainties in critical inputs. The study also demonstrates the methods using cases in the Horda platform in North Sea, including Vette fault zone western boundary fault of Smeaheia area.

The key conclusions from the workflow for probabilistic fault stability assessment are:

- Probability of failure can be used as a more reliable stability measure compared to traditional deterministic measures (e.g., allowable injection pressure, mobilized shear strength), particularly when conventional criteria yield conflicting assessment.
- Demonstration case illustrates the effectiveness of Sobol sensitivity analysis in identifying the ranking of inputs and understanding parameter interaction.
- This study clearly demonstrates the value of Pf as a quantitative measure for stability assessment and resource prioritisation in early field development by clarifying the parameter to be prioritized for further investigation or monitoring.

Main findings from the study on stress estimates using Bayesian approach are as follows:

- Stress data from nearby sites in a relatively homogeneous geological setting (e.g., sedimentary basin where gravitational forces dominate, hydrostatic pore pressure, limited impact of tectonics) indeed provide useful constraining information for sites where stress measurements are scarce Note: This exercise was for demonstration of methodology in the relative homogenous HP area, however, the lithology, burial compaction, uplift and glacial loading all impacts on spatial variation in stresses in the HP area. In such a sense, the depths and lithology of tests will impact given trendline, and these variations of tests depth and lithology varies across the various fields and influences the trends to some extent from site to site (See DV3.2, DV1.2)
- The Bayesian modelling approach enables probabilistic quantification of stress prediction uncertainty, which is suitable for probabilistic assessment of geomechanics-related risks (e.g., fault stability, caprock integrity and induced seismicity)
- The customary complete pooling approach to integrate stress data from multiple sites fails to account for the possible between-site heterogeneity in the data, and hence tend

to yield overconfident stress predictions, especially for target CO_2 storage sites where local stress data are limited.

- The proposed Bayesian hierarchical model is able to borrow appropriate stress across nearby sites by explicitly accounting for between-site similarity/heterogeneity, thereby leading to more accurate predictions of the minimum horizontal stress profile especially for the storage sites of limited data.
- The Bayesian model has the potential to incorporate other available sources of information in the forms of prior distributions to further constrain the stress states at a site, e.g., borehole breakouts, drilling-induced fractures, faulting regimes and frictional faulting theory.
- Such established statistical method demonstrated here may be useful in screening studies in early phase of CO2 storage site developments and in detailed risking studies to better understand the uncertainty range from well proven statistical methods. For detailed evaluations of stress, local stress measurements from XLOT in the relevant formations in combinations with logs, geology and stress history in combination with numerical study will be the more robust method for precise estimates (DV1.2).

5 Further study

For further study, it's essential to recognize that while this research provides valuable insights, its results are site-specific, primarily demonstrating the methodology rather than a comprehensive site evaluation. Future work for the probabilistic assessment and associated sensitivity analysis should focus on applying this approach to other sites, such as the Gamma area in Smeaheia and the Aurora field, to understand regional variations and validate the methodology. Moreover, refining the method for determining the range of input data will enhance accuracy. Additionally, exploring different applications like caprock integrity assessment and reservoir simulation could broaden the scope and impact of this research approach.

6 References

Anderson, R. A., Ingram, D. S., & Zanier, A. M. (1973). Determining fracture pressure gradients from well logs. Journal of Petroleum technology, 25(11), 1259-1268.

Andrews, J. S., & de Lesquen, C. (2019). Stress determination from logs. Why the simple uniaxial strain model is physically flawed but still gives relatively good matches to high quality stress measurements performed on several fields offshore Norway. In 53rd US Rock Mechanics/Geomechanics Symposium, New York, USA.

Andrews, J. S., Fintland, T. G., Helstrup, O. A., Horsrud, P., & Raaen, A. M. (2016). Use of unique database of good quality stress data to investigate theories of fracture initiation, fracture propagation and the stress state in the subsurface. In 50th US Rock Mechanics/Geomechanics Symposium, Houston, USA.

Bao, T., & Burghardt, J. (2022). A Bayesian approach for in-situ stress prediction and uncertainty quantification for subsurface engineering. Rock Mechanics and Rock Engineering, 55(8), 4531-4548.

Box, G.E.P., Draper, N.R. (1987). Empirical model-building and response surfaces, Empirical model-building and response surfaces. John Wiley & Sons, Oxford, England.

Bozorgzadeh, N., & Bathurst, R. J. (2019). Bayesian model checking, comparison and selection with emphasis on outlier detection for geotechnical reliability-based design. Computers and Geotechnics, 116, 103181.

Bozorgzadeh, N., & Bathurst, R. J. (2022). Hierarchical Bayesian approaches to statistical modelling of geotechnical data. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 16(3), 452-469.

Bozorgzadeh, N., Grande, L., Choi, J.C., Skurtveit, E. (2022). D5.1: Internal guideline for uncertainty quantification of rock mechanical properties (No. SHARP Storage – Project no 327342 D5.1).

Breckels, I. M., & Van Eekelen, H. A. M. (1982). Relationship between horizontal stress and depth in sedimentary basins. Journal of Petroleum Technology, 34(09), 2191-2199.

Breckels, I. M., & Van Eekelen, H. A. M. (1982). Relationship between horizontal stress and depth in sedimentary basins. Journal of Petroleum Technology, 34(09), 2191-2199.

Ching, J., Phoon, K.-K. (2013). Mobilized shear strength of spatially variable soils under simple stress states. Struct. Saf. 41, 20–28. https://doi.org/10.1016/j.strusafe.2012.10.001

Choi, J. C., Skurtveit, E., and Grande, L., 2019, Deep neural network-based prediction of leak-off pressure in offshore Norway accepted for Offshore Technology Conference 2019, Huston, USA

Choi, J.C., Skurtveit, E., Huynh, K.D.V., Grande, L. (2023). Uncertainty of stress path in fault stability assessment during CO2 injection: Comparing smeaheia 3D geomechanics model with analytical approaches. Int. J. Greenh. Gas Control 125, 103881. https://doi.org/10.1016/j.ijggc.2023.103881

Feng, Y., Bozorgzadeh, N., & Harrison, J. P. (2020). Bayesian analysis for uncertainty quantification of in situ stress data. International Journal of Rock Mechanics and Mining Sciences, 134, 104381.

Feng, Y., Gao, K., & Lacasse, S. (2023). Bayesian partial pooling to reduce uncertainty in overcoring rock stress estimation. Journal of Rock Mechanics and Geotechnical Engineering.

Feng, Y., Gao, K., Mignan, A., & Li, J. (2021). Improving local mean stress estimation using Bayesian hierarchical modelling. International Journal of Rock Mechanics and Mining Sciences, 148, 104924.

Gassnova (2021). Smeaheia CO2 Storage Prospect Reference Dataset. https://doi.org/10.11582/2021.00012

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis (3rd ed.). Boca Raton, FL, USA: CRC Press.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. Statistics and computing, 24(6), 997-1016.

Grande, L., Mondol, N.H., Hopper, J., Roberts, D., Phillips, D. (2022). DV1.1b Stress Drivers and Outline of Proposed Numerical Modelling Campaign.

Grande, L., Forsberg, C. F., Mondol, N. H., Roberts, D., & Phillips, D. (2023). *DV1.2 Lithology Assessment & Constitutive Model* (DV1.2).

Grande, L., Skurtveit, E., Cuss, R., Singh, M., & Vosgerau, H. (2023). *DV3.2 Stress and burial history impact on present day state* (Sharp report DV3.2).

Hasofer, A.M., Lind, N.C. (Niels C., Division, U. of W.S.M. (1973). An Exact and Invariant First-order Reliability Format. Solid Mechanics Division, University of Waterloo.

Hillis, R. R. (2001). Coupled changes in pore pressure and stress in oil fields and sedimentary basins. Petroleum Geoscience, 7(4), 419-425.

Kloek, T., van Dijk, H.K. (1978). Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo. Econometrica 46, 1–19. https://doi.org/10.2307/1913641

Lacasse, S., Nadim, F. (2007). Probabilistic geotechnical analyses for offshore facilities. Georisk Assess. Manag. Risk Eng. Syst. Geohazards 1, 21–42. https://doi.org/10.1080/17499510701204224

McKay, M.D., Beckman, R.J., Conover, W.J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21, 239–245. https://doi.org/10.2307/1268522

Mesri, G., Shahien, M. (2003). Residual Shear Strength Mobilized in First-Time Slope Failures. J. Geotech. Geoenvironmental Eng. 129, 12–31. https://doi.org/10.1061/(ASCE)1090-0241(2003)129:1(12)

Michie, E.A.H., Mulrooney, M.J., Braathen, A. (2021). Fault interpretation uncertainties using seismic data, and the effects on fault seal analysis: a case study from the Horda Platform, with implications for CO_2 storage. Solid Earth 12, 1259–1286. https://doi.org/10.5194/se-12-1259-2021

Mulrooney, M.J., Osmond, J.L., Skurtveit, E., Faleide, J.I., Braathen, A. (2020). Structural analysis of the Smeaheia fault block, a potential CO2 storage site, northern Horda Platform, North Sea. Mar. Pet. Geol. 121, 104598. https://doi.org/10.1016/j.marpetgeo.2020.104598

Raaen, A.M., Horsrud, P., Kjørholt, H., Økland, D., 2006. Improved routine estimation of the minimum horizontal stress component from extended leak-off tests. International Journal of Rock Mechanics and Mining Sciences 43, 37–48. https://doi.org/10.1016/j.ijrmms.2005.04.005

Rahman, M.J., Choi, J.C., Fawad, M., Mondol, N.H. (2021). Probabilistic analysis of Vette fault stability in potential CO2 storage site Smeaheia, offshore Norway. Int. J. Greenh. Gas Control 108, 103315. https://doi.org/10.1016/j.ijggc.2021.103315

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J.H.A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., Maier, H.R. (2021). The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. Environ. Model. Softw. 137, 104954. https://doi.org/10.1016/j.envsoft.2020.104954

Saltelli, A., Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. Environ. Model. Softw. 25, 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012

Skurtveit, E., Choi, J.C., Osmond, J., Mulrooney, M., Braathen, A. (2018). 3D fault integrity screening for smeaheia CO2 injection site, in: 14th Greenhouse Gas Control Technologies Conference Melbourne. pp. 21–26.

Sobol′, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul., The Second IMACS Seminar on Monte Carlo Methods 55, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Sudret, B., Marelli, S., Wiart, J. (2017). Surrogate models for uncertainty quantification: An overview, in: 2017 11th European Conference on Antennas and Propagation (EUCAP). Presented at the 2017 11th European Conference on Antennas and Propagation (EUCAP), pp. 793–797. https://doi.org/10.23919/EuCAP.2017.7928679

Thompson, N., Andrews, J.S., Wu, L., Meneguolo, R. (2022a). Characterization of the in-situ stress on the Horda platform – A study from the Northern Lights Eos well. Int. J. Greenh. Gas Control 114, 103580. https://doi.org/10.1016/j.ijggc.2022.103580

Thompson, N., Andrews, J. S., Wu, L., & Meneguolo, R. (2022a). Characterization of the in-situ stress on the Horda platform–A study from the Northern Lights Eos well. International Journal of Greenhouse Gas Control, 114, 103580.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing, 27(5), 1413-1432.

Wiprut, D., Zoback, M. (2000). Constraining the stress tensor in the Visund field, Norwegian North Sea: Application to wellbore stability and sand production. Int. J. Rock Mech. Min. Sci. 37, 317–336. https://doi.org/10.1016/S1365-1609(99)00109-4

Wong, R.C.K., Ma, S.K.Y., Wong, R.H.C., Chau, K.T. (2007). Shear strength components of concrete under direct shearing. Cem. Concr. Res. 37, 1248–1256. https://doi.org/10.1016/j.cemconres.2007.02.021

Wu, L., Skurtveit, E., Thompson, N., Michie, E., Fossen, H., Braathen, A., Fisher, Q., Lidstone, A., & Bostrøm, B. (2022). *Containment Risk Assessment and Management of Co2 Storage on the Horda Platform* (SSRN Scholarly Paper 4272132). https://doi.org/10.2139/ssrn.4272132

Zang, A., & Stephansson, O. (2009). Stress field of the Earth's crust. Springer Science & Business Media.

Zhang, Y., Yin, S., & Zhang, J. (2021). In situ stress prediction in subsurface rocks: An overview and a new method. Geofluids, 2021, 1-11.