DOI: https://doi.org/10.68732/436d65ea7773b392 ISBN: 978-9989-2053-6-1

An adaptive data sampling strategy for in-situ stress characterization

A. Hamidi & J. P. Harrison
University of Toronto, Toronto, Canada
asieh.hamidi@utoronto.ca

D. Mas Ivars
Swedish Nuclear Fuel and Waste Management Company (SKB), Solna, Sweden
Royal Institute of Technology (KTH), Stockholm, Sweden

H. A. Kasani
Nuclear Waste Management Organization (NWMO), Toronto, Canada

A. Sánchez Juncal
Swedish Nuclear Fuel and Waste Management Company (SKB), Solna, Sweden

Abstract

Deep geological repositories (DGRs) for nuclear waste require reliable in-situ stress characterization to ensure appropriate safety during the phases of construction, operation, and post-closure. However, determining optimal data requirements and sampling strategies is challenging due to geological heterogeneity, measurement costs, and varying uncertainty tolerances across these phases. This study introduces a novel adaptive framework that integrates change point detection (e.g., due to regional stress field variations, geological complexities, etc.) and Bayesian hierarchical modeling to optimize stress characterization strategies. The framework applies a multi-tiered methodology, and provides explicit decision criteria for method selection, quantifies data adequacy through statistical power analysis, and identifies optimal measurement campaigns for different repository design phases, from preliminary site characterization (minimum viable datasets) to detailed design (high-confidence predictions). This methodology helps address the fundamental question "how much data are enough?" by providing quantitative, site-specific guidance for stress characterization programs, ultimately enabling more efficient resource allocation and improved confidence for performance assessments in a nuclear waste repository site.

Keywords

In-situ stress data, Deep geological repository, Bayesian hierarchical model, Data adequacy.

1 Introduction

Deep geological repositories (DGRs) are internationally recognized as the most viable long-term solution for high-level nuclear waste (HLW) disposal (Unokiwedi et al. 2025; Wang et al. 2018). From a rock engineering perspective, reliable in-situ stress characterization within the DGR host rock is a prerequisite for safe design, construction, and long-term performance assessment (Amadei & Stephansson 1997; Hudson & Harrison 2000). Stress measurement campaigns are costly, and geological heterogeneity means that data requirements are neither uniform across depth nor constant

across repository design phases, from preliminary site characterization through detailed design and final safety assessment (IAEA 2011; Zoback 2010).

Despite the recognized importance of stress characterization, the fundamental question, "*how much data are enough?*" lacks a quantitative, adaptive answer in current practice. Existing frameworks often treat the rock mass as statistically homogeneous or rely on predetermined measurement programs that are not explicitly linked to quantitative uncertainty reduction (Baecher & Christian 2005; Phoon & Kulhawy 1999). Bayesian hierarchical approaches offer a principled pathway for uncertainty quantification and information sharing across geological domains (Gelman et al. 2013), yet their application to DGR stress characterization has so far been limited (Bao & Burghardt 2022). This paper outlines an adaptive Bayesian-statistical framework that integrates change-point detection, distributional analysis, and hierarchical Bayesian modeling to optimize stress characterization strategies. The framework is validated using overcoring database from the Forsmark site in Sweden (Sjöberg et al. 2025), and provides phase-specific decision criteria for measurement campaign design.

2 Forsmark Overcoring Database

Forsmark, central Sweden, is the site of a proposed DGR and has been extensively investigated by SKB (Martin 2007; Stephansson & Ångman 1986). The raw overcoring database comprises 119 records from ten boreholes (Sjöberg et al. 2025); one borehole was excluded due to incomplete stress tensor records across all its measurements, a standard data quality filter. The dataset analyzed in this paper comprises 114 complete stress tensor measurements from nine boreholes, spanning depths of 10.4–501.8 m. The maximum principal stress σ_1 ranges from 1.8 to 63.4 MPa, and horizontal stresses σ_{Hmax} and σ_{Hmin} range from 1.8–62.6 MPa and 0–42.6 MPa respectively, with vertical stress σ_V up to 30.2 MPa, indicating substantial depth-dependent and spatial variability.

3 Adaptive Framework

To address the fundamental question of "*how much data are enough?*" for in-situ stress characterization in DGRs, a four-step adaptive Bayesian-statistical framework was developed. The framework is designed to account for geological heterogeneity across depth, varying data availability between segments, and the evolving uncertainty tolerance across repository design phases, conditions that make uniform or predetermined measurement programs statistically unjustified (Table 1).

Step 1 partitions the σ_1 depth profile into statistically homogeneous segments using three complementary change-point algorithms. Breakpoint consensus is when at least 30% of configurations agree within a ±30 m tolerance, with a minimum 20 m segment separation enforced. Each segment is classified as unimodal or multimodal by comparing one- versus two-component Gaussian Mixture Models (GMM) via the Bayesian Information Criterion, with $\Delta_{BIC} > 10$ indicating multimodality. Step 2 fits a Bayesian distributional model within each segment, characterizing the posterior of the segment mean μ and standard deviation ς under weakly informative priors to allow the data to dominate. Model adequacy is assessed by posterior predictive coverage, targeting posterior predictive coverage *PPC* $\geq$0.93.

Step 3 addresses data-sparse segments through hierarchical partial pooling, which borrows statistical strength from data-rich segments in proportion to each segment's data scarcity. The three strategies of no pooling, full pooling and partial pooling were formally compared, with partial pooling being adopted as optimal; a credible interval (CI) reduction metric quantifies the gain. Step 4 derives minimum sample size requirements analytically from the 95% CI half-width criterion for three precision targets, and extends this to a four-phase repository lifecycle framework with pre-defined CI targets at each design stage. A four-tier decision tree then provides an explicit method-selection recommendation per depth domain, ranging from a data-insufficiency flag at Tier 0 through to GMM-based ensemble approaches at Tier 3. The components of step 3 provide the quantitative, site-specific, and phase-adaptive decision criteria needed to optimize stress characterization campaigns at any stage of repository development.

4 Results

4.1 Segmentation and Distributional Signatures

Five consensus breakpoints at 73, 151, 238, 300, and 375 m partition the dataset into six depth segments (Fig.1), each with a statistically distinct stress regime. The ±30 m consensus tolerance

represents the positional uncertainty on each boundary, and this should be interpreted as a transition zone rather than a sharp interface. Mean values of σ_1 increase from 13.1 MPa (Segment 1, 10–72 m) to 51.4 MPa (Segment 6, 377–502 m), with Coefficient of Variation (CoV) ranging from 0.163 to 0.615, near-surface segments carry substantially greater relative uncertainty than deeper ones. All six segments are unimodal ($\Delta_{BIC} \leq 0$for Segments 1, 2, 3, 6; $\Delta_{BIC} \leq 2$ for Segments 4, 5), validating the use of Gaussian distributional models.

Table 1. Adaptive framework.

Steps	Methodology & criteria	Equations & Parameters
Change-point detection & domain segmentation Identify depth-related domains	Three algorithms applied to the σ_1 depth profile: - **PELT**: exact dynamic programming (penalized cost minimization) - **Binary Segmentation**: recursive splitting - **Bottom-Up**: agglomerative segment merging **Consensus**: ≥30% of configurations agree within ±30m; minimum 20 m separation enforced. **Sensitivity confirmed**: 20–50% threshold and ±20–40 m tolerance represents identical breakpoints. Segments classified via GMM BIC comparison (1-component vs 2-component Gaussian Mixture): Δ_{BIC}> 10 is multimodal, otherwise unimodal.	Eq. 1. $\sum_{i=1}^{m+1}[C(y_{s_{i-1}+1:s_i})+\beta]$ C(·) = within-segment sum of squared deviations s_i = change-point positions m = number of change-points β = complexity penalty Eq. 1a. $BIC = -2\cdot\ln(\hat{L})+k\cdot\ln(n)$ k = no. model parameters n = segment sample size Δ_{BIC}> 10 → multimodal
Bayesian Distributional Modeling Statistically characterize stress in all domains	**Goal**: characterize posterior of mean μ and std ς within each statistically uniform segment treated as homogeneous (by ΔBIC ≤ 2) Prior for μ centered on segment sample mean $\bar{x}_s$(weakly informative, allows data to dominate) **Implementation**: PyMC, 4 chains, 1500 draws,700 tuning steps per chain (NUTS sampler) **Model adequacy**: posterior predictive coverage PPC = fraction of observed σ_1 within 95% posterior predictive interval Target PPC ≥ 0.93; wide CI reflects data scarcity.	Priors, Eq. 2. $\mu \sim N(\bar{x}_s, 20)$ $\varsigma \sim HalfNormal(15)$ $\sigma_1^{(i)} \sim N(\mu, \varsigma)$ $\bar{x}_s$ = segment sample mean (weakly informative center) Adequacy check: PPC = $\lvert\{\sigma_1^{(i)} \in 95\%PPD\}\rvert$ / N Target: PPC ≥ 0.93
Hierarchical Bayesian Pooling Share data among segments	Segments with n ≤ 9 → **pooling needed** Three pooling strategies formally compared: **No pooling**: fully independent segment models **Full pooling**: single collapsed model (ignores between-segment differences) **Partial pooling** (optimal): hierarchical; middle ground between no- and full-pooling **Strength-borrowing**: sparse segments are pulled toward the global mean proportional to their data scarcity (automatic regularization) **Hyperpriors**: weakly informative, N(25,20) spans full σ_1 range, HalfNormal(15) ensures σ > 0.	Hyperpriors, Eq. 3. $\mu_{global} \sim N(25, 20)$ $\varsigma_{global} \sim HalfNormal(15)$ Segment means, Eq. 3a. $\mu_s \sim N(\mu_{global}, \varsigma_{global})$ s = 1, … , 6 Likelihood, Eq. 3b. $\varsigma_{within} \sim HalfNormal(15)$ $\sigma_1^{(i)} \sim N(\mu_s, \varsigma_{within})$ CI reduction metric Eq. 3c. 1 − pooled CI½ / unpooled CI½
Statistical Power and Data Adequacy + Decision Tree Confirm validity of sample size predictions	Minimum sample size N_{min} derived from: **95% CI half-width criterion** Three engineering **precision targets**:Δ = ±2 MPa, ±3 MPa, ±5 MPa Decision tree: four-tier method assignment:**Tier 0** (N < 8): insufficient data; **Tier 1** (unimodal,Coefficient of Variation(CoV)< 0.20, N < 25); **Tier 2** (unimodal, 0.20 ≤ CoV < 0.40); **Tier 3** (multimodal or CoV ≥ 0.40). **Output**: method label + data sufficiency flag	Eq. 4. $N_{min} = (z_{\alpha/2}\cdot\hat{\varsigma}/\Delta)^2$ $z_{\alpha/2}$ = 1.96 (95% conf.) $\hat{\varsigma}$= segment observed std Δ = target half-width (±2, ±3, or ±5 MPa) Phase CI targets (MPa): Ph1 ≤ 15 | Ph2 ≤ 10 Ph3 ≤ 7 | Ph4 ≤ 5 Decision tree inputs: {N, CoV, distribution type}

4.2 Bayesian Distributional Model Performance

Posterior predictive coverage ranged from 0.93 to 1.0 across all segments (Table 2), confirming well-calibrated models against the conventional benchmark that ≥ 95% of observed values fall within the 95% posterior predictive interval (Gelman et al. 2013). Coverage of exactly 1.0 in smaller segments (*n*=6–13) reflects wide predictive intervals relative to sparse observations, a consequence of high within-segment variability rather than model mis-specification. CI widths directly reflect data

availability: Segment 2 (n=26, CoV=0.271) achieves 4.25 MPa, while Segment 5 (n=6, CoV=0.487) yields 21.49 MPa, quantitatively identifying where additional measurements are most needed.

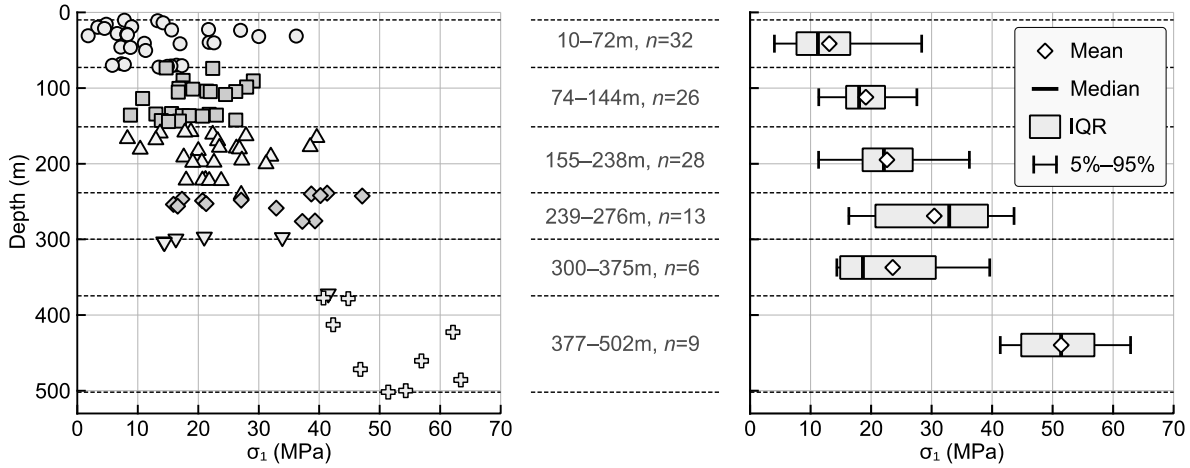


Fig.1 Multi-algorithm change-point detection and final 6-segment depth profile.

Table 2. Bayesian distributional model results.

Seg	N	μ posterior (MPa)	95% CI (MPa)	CI width (MPa)	PPC coverage	Seg	N	μ posterior (MPa)	95% CI (MPa)	CI width (MPa)	PPC coverage
1	32	13.08	(10.23, 15.98)	5.75	0.938	4	13	30.48	(23.38, 37.20)	13.81	1.000
2	26	19.18	(17.03, 21.28)	4.25	1.000	5	6	23.38	(12.21, 33.70)	21.49	1.000
3	28	22.61	(19.77, 25.39)	5.62	0.929	6	9	51.45	(45.07, 58.19)	13.12	1.000

4.3 Hierarchical Strength Borrowing

Partial pooling reduced the 95% CI half-width by 25% for Segment 5 (8.4 to 6.3 MPa) and by 26% for Segment 4 (5.75 to 4.28 MPa), the two most data-limited non-sparse segments (Fig. 2). Segment 2, μ = 19.1 MPa, deviates the most from the global mean (22.4 MPa), and partial pooling slightly increases the CI, a well-understood and expected regularization effect in hierarchical models (Gelman et al. 2013). These results confirm that hierarchical pooling is most beneficial precisely where data are scarce, which is the operationally relevant condition during early repository characterization phases.

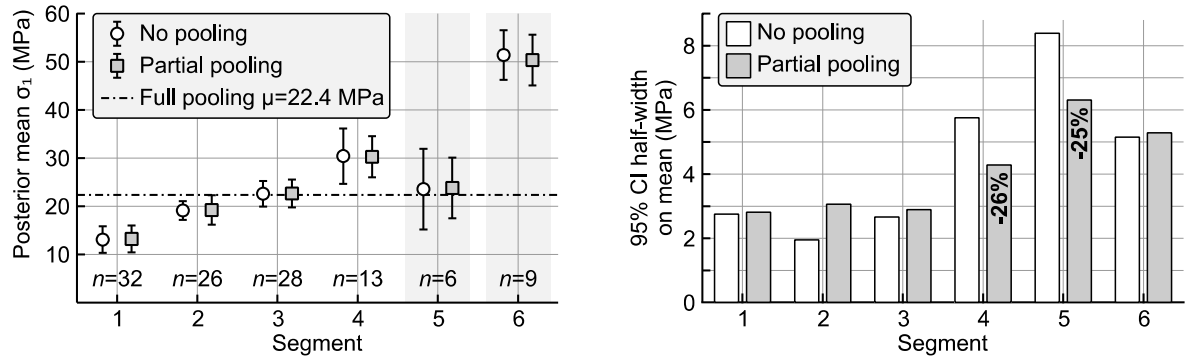


Fig. 2 Hierarchical Bayesian partial pooling: (a) no pooling vs partial pooling; (b) Uncertainty reduction from pooling.

4.4 Sample Size Requirements and Phase Adequacy

Minimum viable sample sizes vary substantially across segments (Table 3), driven entirely by within-segment variability. Segment 2 (ς = 5.18 MPa) requires only 5 measurements for ±5 MPa precision, while Segment 5 (ς = 11.47 MPa) requires 21 for the same tolerance and 57 for the stricter ±3 MPa engineering target.

To evaluate whether the current dataset meets engineering tolerance requirements at each stage of repository development, four characterization phases were defined based on cumulative measurement availability, as shown in Fig. 3 a: Phase 1, Preliminary Site Characterization (n = 17), Phase 2, Advanced Characterization (n = 38), Phase 3, Detailed Design (n = 65), and Phase 4, Final Validation (n = 114). These phases correspond to analytically defined sampling thresholds aligned with repository planning stages, not to specific excavation or in-situ verification events. The CI width achieved at each

phase (Fig. 3 b) is compared against pre-defined engineering tolerance targets (15, 10, 7, and 5 MPa respectively). As Table 3 shows, segment 5 (n = 6) falls below the minimum viable dataset even at the coarsest precision level, quantitatively identifying that future measurement campaigns should prioritize this segment.

Table 3. Minimum sample size requirements and current availability.

Seg.	N available	ς (σ_1 std MPa)	CoV	Min N (±5 MPa)	Min N (±3 MPa)	Min N (±2 MPa)
1	32	8.05	0.615	10	28	62
2	26	5.18	0.271	5	12	26
3	28	7.32	0.324	9	23	52
4	13	11.01	0.362	19	52	117
5	6	11.47	0.487	21	57	127
6	9	8.36	0.163	11	30	67

Extending this analysis to the full repository life-cycle, proportional sampling across all six segments was simulated at four characterization phases, with CI targets calibrated relative to the overall σ_1 standard deviation (12.7 MPa) to reflect realistic engineering tolerances at each stage (IAEA 2011; NEA 2008). All four phases meet their targets, with CI width decreasing progressively from 13.0 MPa at Phase 1 (n = 17, preliminary site characterization) to 4.7 MPa at Phase 4 (n = 114, final validation), answering the question of how much data are enough at each design stage (Fig. 3 a-b).The per-segment minimum sample size curves (Fig. 4) further show that requirements vary from 5 to 57 measurements at ±3 MPa precision, with Segments 4 and 5 remaining below their minimum thresholds at current data levels, confirming where future measurement efforts should be concentrated.

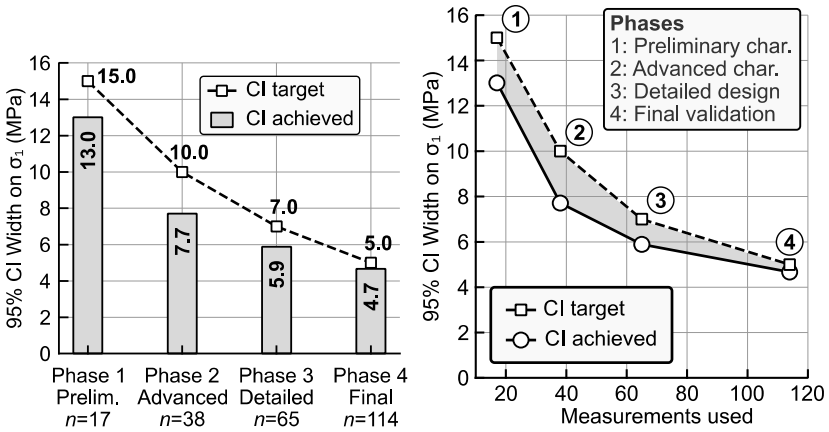


Fig. 3 Data requirements: (a) credible intervals vs. phase; (b) credible intervals vs. number of data.

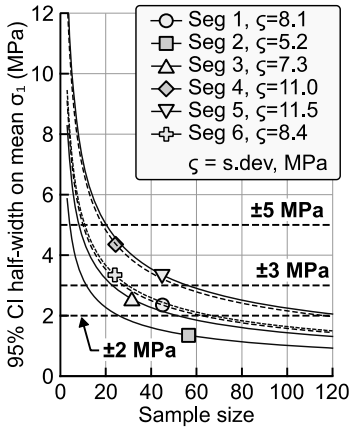


Fig. 4 Data adequacy at different repository life phase CI thresholds.

5 Discussion

The results demonstrate that depth-dependent segmentation is statistically appropriate for the Forsmark dataset, as uniform measurement density across all depths is both statistically unjustified and operationally inefficient given the order-of-magnitude variation in data requirements across segments. Measurement campaigns should therefore prioritize segment coverage before increasing local density.

The decision tree (Fig. 5) synthesizes all preceding results into explicit method-selection criteria per segment, from data-insufficient segments requiring additional measurements to well-sampled segments adequately characterized by single Gaussian models (Gelman et al. 2013). The hierarchical pooling provides an immediate practical tool for data-sparse segments: a 25% CI reduction is achieved for Segment 5 without any new measurements, indicating that this is a valuable interim strategy during early characterization phases. A limitation of this study is that validation is performed at a single site using overcoring data only. The specific breakpoint depths and segment statistics are site-specific; however, the framework itself, change-point detection, distributional modeling, hierarchical pooling, and power-based data adequacy, is site-independent and transferable to any stress measurement dataset.

6 Conclusions

This paper introduces an adaptive Bayesian-statistical framework that integrates change-point detection, distributional analysis, and hierarchical Bayesian modeling to optimize stress characterization strategies, and presents an example application of it using the Forsmark stress database. Multi-algorithm change-point detection identified six statistically justified depth segments, and Bayesian distributional models yielded well-calibrated segment-level posteriors (PPC coverage 0.93–1.0), with CI widths reflecting data scarcity and variability. Hierarchical partial pooling reduces the CI half-width by up to 25% for data-sparse segments, showing that this is an applicable strategy for uncertainty reduction strategy without additional measurements. The minimum sample size requirements were found to vary from 5 to 57 measurements per segment for ±3 MPa precision, demonstrating that uniform sampling strategies are statistically inappropriate and that spatial coverage across all depth segments is more critical than local measurement density. Finally, the framework is seen to be able to provide explicit quantitative decision criteria from preliminary characterization through final validation.

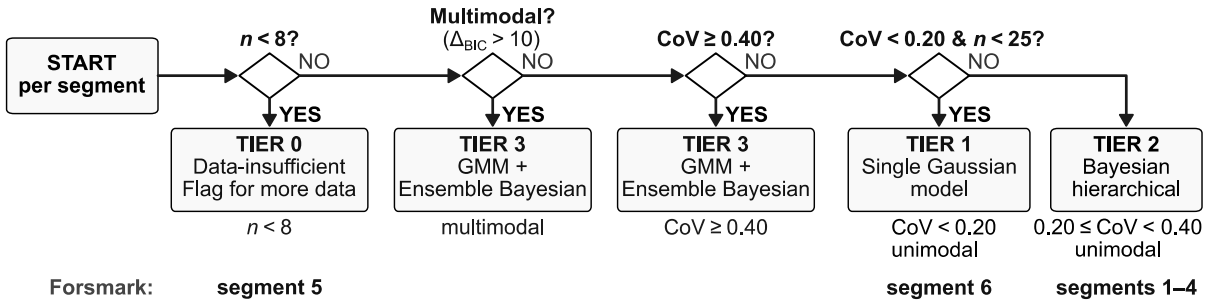


Fig. 5 Decision tree framework.

Acknowledgement

This research was jointly funded by NWMO and SKB. The authors gratefully acknowledge their financial support.

References

Amadei B, Stephansson O (1997) *Rock Stress and Its Measurement*. Chapman & Hall.

Baecher G.B, Christian J.T (2005) *Reliability and statistics in geotechnical engineering*. John Wiley & Sons.

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

Gelman A, Carlin J.B, Stern H.S, Dunson D, Vehtari A, Rubin D.B (2013) *Bayesian Data Analysis*. CRC Press.

Hudson J.A, Harrison J.P (2000) *Engineering Rock Mechanics: an introduction to the principles*. Elsevier.

IAEA (2011) *Geological Disposal Facilities for Radioactive Waste*. Intl. Atomic Energy Agency: SSG-14.

Martin C.D (2007) *Quantifying in situ stress magnitudes and orientations for Forsmark* (SKB Report R-07-26). Swedish Nuclear Fuel and Waste Management Company.

NEA OECD (2008) *Moving Forward with Geological Disposal of Radioactive Waste*, A Collective Statement by the NEA Radioactive Waste Management Committee (RWMC). Nuclear Energy Agency: Report 6433.

Phoon K.K, Kulhawy F.H (1999) Characterization of geotechnical variability. *Can. Geotech. J.*, 36, 612–624.

Sjöberg J, Figueiredo B, Hakala M (2025) *Reliability ranking of overcoring stress measurements at Forsmark.* SKB Report R-21-08. ISSN 1402-3091, Svensk Kärnbränslehantering.

Stephansson O, Ångman P (1986) Hydraulic fracturing stress measurements at Forsmark and Stidsvig, Sweden. *Bulletin of the Geological Society of Finland*, *58*(1), 307-333.

Unokiwedi O.P, Gao J, Bethune T, Awolayo A.N (2025) Deep geological repositories—A review of design concepts, near-field evolution, and their implications for nuclear waste containment. *J. Environmental Radioactivity*, *289*, 107750.

Wang J, Chen L, Su R, Zhao X (2018) The Beishan underground research laboratory for geological disposal of high-level radioactive waste in China: planning, site selection, site characterization and in situ tests. *J. Rock Mech. Geotech. Eng.*, *10*(3), 411-435.

Zoback M.D (2010) *Reservoir Geomechanics*. Cambridge University Press.