Analyzing *in situ* stress: challenges in quantifying stress domains

Muhammad Amir Javaid & John P. Harrison
Department of Civil & Mineral Engineering – University of Toronto, Toronto, Ontario, Canada
Hossein A. Kasani
Nuclear Waste Management Organization (NWMO), Toronto, Ontario, Canada
Diego Mas Ivars
1) SKB, Swedish Nuclear Fuel and Waste Management Co, Solna, Sweden
2) Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

BRIDGING INFRASTRUCTURE & RESOURCES
GeoSaskatoon 2023 OCT 1–4

ABSTRACT
Accurately characterizing the state of *in situ* stress in rock is important for the design of all underground engineering projects, but is crucial for safety-critical projects such as deep geological repositories for nuclear waste. However, designers continue to be confronted with the challenging task of characterizing the significant variability and uncertainty found in the *in situ* stress state across the project volume, and particularly identifying separate stress domains. One routine approach is to partition, or group, stress data on the basis of depth below ground. In this paper we discuss a customary approach to identifying stress domains, and illustrate challenges in its application. We go on to present a novel approach that uses Bayesian linear segmented regression of Cartesian stress components to statistically characterize the variability and uncertainty in the depth of stress domain boundaries. Synthetic data are used to demonstrate the suitability and efficacy of this method, and it is then applied to stress measurements in crystalline rock obtained at the Forsmark site in Sweden. We conclude that use of a Bayesian approach is beneficial as it is able to formally augment stress measurement data with other valuable geological information.

RÉSUMÉ
Caractériser précisément l'état de contrainte *in situ* est important pour la conception d'infrastructures souterraines, d'autant plus pour des projets critiques au regard de leur sécurité, tels que le stockage géologique de déchets nucléaires. Cependant, la variabilité et l'incertitude importante associées aux contraintes *in situ* reste difficile à caractériser à l'échelle d'un site, nécessitant généralement l'identification de domaines de contrainte distincts. Une approche standard consiste à séparer, ou grouper, les données de contrainte selon la profondeur dans le sous-sol. Dans cet article, nous discutons d'une méthode classique permettant d'identifier des domaines de contrainte et illustrons les défis liés à son application. Nous présentons ensuite une nouvelle approche utilisant la régression segmentée linéaire bayésienne des composantes de contrainte cartésiennes pour caractériser statistiquement la variabilité et l'incertitude de la profondeur des limites des domaines de contrainte. Des données synthétiques sont utilisées pour démontrer la pertinence et l'efficacité de cette méthode, qui est ensuite appliquée aux mesures de contraintes obtenues en milieu cristallin sur le site de Forsmark en Suède. Nous concluons que l'utilisation d'une approche bayésienne est bénéfique car elle permet d'associer les données de mesure des contraintes avec d'autres informations géologiques précieuses.

1 INTRODUCTION

Characterization of the *in situ* stress state in rock is crucial for safety-critical projects such as deep geological repositories for safely accessible containment of nuclear waste. Extensive campaigns are often undertaken to obtain estimates of the *in situ* stress at various locations in the 3D space that comprises the entire project volume. However, such data lead to the challenging task of characterizing the significant variability and uncertainty in the *in situ* stress state, made more difficult due to the lack of robust and universally agreed methods for characterization of *in situ* stress.

Partitioning stress data into distinct depth-wise horizons is a commonly adopted approach, although there is no well-established heuristics for this. Further, the variability and uncertainty associated with estimates of the interface depths of these horizons (stress domains) is often ignored.

In this paper we highlight a few of the challenges in identifying the depth stress domains and present a method that can potentially characterize the variability and uncertainty in depth estimates of these stress domain boundaries. We examine a novel application of Bayesian linear segmented regression firstly using synthetic stress data and then applying the method to over 100 overcoring stress measurements obtained at the Forsmark site in Sweden. The results are presented and discussed, and challenges in objectively identifying depth stress domains and characterizing the variability and uncertainty in their boundaries are highlighted.

2 BACKGROUND

The *in situ* stress state at the Forsmark site has been partitioned (Martin, 2007) into four depth-wise domains of 0–150 m, 150–300 m, 300–400 m and 400–1000 m on the basis of two parameters: one third of the first invariant, $I_1/3$, and the ratio of major and intermediate principal stress, σ_1/σ_2. Smoothed parameter values were obtained by taking the moving median of six measurements.

One shortcoming of this approach is that the results are sensitive to both the number and continuity of the measurements used in the smoothing. To demonstrate this we first replicate Martin (2007) analysis using sample size of 6, and then perform the analysis using sample sizes of 10, 15, and 20 (Figure 1). As expected, the smoothing becomes more pronounced with increasing number of samples. On the basis of these profiles it could be argued that there exist two stress domain boundaries at depths of approximately 80 m and 170 m, but the results are not unequivocal. A more pronounced boundary is seen at about 300 m depth, but we believe that this could be a result of the lack of stress measurements between 300 to 400 m. Another, and more serious, drawback to this analysis is the use of the first invariant of stress. This scalar measure ignores the principal stress orientations (Gao & Harrison, 2018a) and can thus wrongly identify quite different stress states (in terms of principal stress orientations) as being equivalent. These two disadvantages indicate that an improved technique is required for identifying stress domains. Furthermore, it is highly unlikely that such crisp or hard stress domain boundaries will exist in any geological environment: it is almost always the case such domain boundaries exhibit considerable uncertainty due to the various geological processes.

Recently, Bayesian linear regression of Cartesian stress components has been proposed as a technique for obtaining mean stress estimates from posterior distributions (Javaid et al., 2022a, 2022b). These authors demonstrated the efficacy of the technique by analysing over 100 overcoring stress measurements obtained at the Forsmark site. Here, we introduce a Bayesian linear segmented regression method that can improve the predicted mean stresses together with quantifying the variability and uncertainty associated with depths (boundaries) of the depth stress domains.

3 METHODOLOGY

In this section the multivariate (MV) stress model, the generalized MV Bayesian stress model and previously proposed Bayesian linear regression model for stress are first explained, and then the method of analysis used in the current studies is introduced.

3.1 Bayesian linear regression model

A multivariate model for quantifying the variability of *in situ* stress has recently been proposed (Gao & Harrison 2016, 2017, 2018a, 2018b). This model is faithful to the tensorial nature of stress, but has limited application in practical rock engineering due to it having a frequentist basis and thus

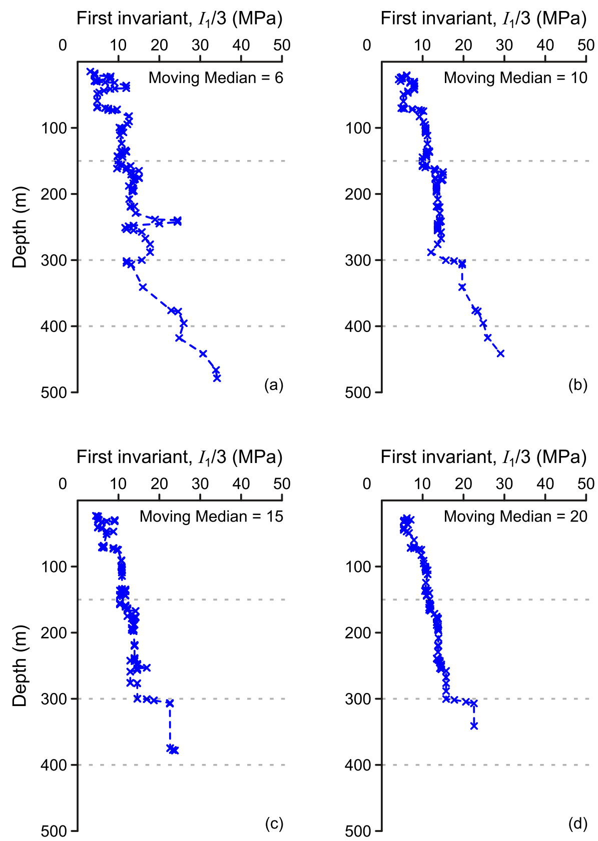

Figure 1. Moving median analysis of Forsmark data using four different sample sizes. Depth boundaries of stress domains from Martin (2007) are shown in dashed lines.

requiring a large number of *in situ* stress measurements. To overcome this shortcoming, Feng & Harrison (2019) and Feng et al. (2020, 2021) proposed a generalized MV Bayesian stress model for use in the case of limited data.

The Bayesian stress model assumes that the six distinct components of a complete 3D stress tensor obtained via stress measurement, $\mathbf{Y}_{\text{data}}$, follow a multivariate normal distribution so that

$$\mathbf{Y}_{\text{data}} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Omega}), \qquad [1]$$

where

$$\mathbf{Y}_{\text{data}} = [\sigma_x \quad \tau_{xy} \quad \tau_{xz} \quad \sigma_y \quad \tau_{yz} \quad \sigma_z] \qquad [2]$$

with the prior distributions

$$\boldsymbol{\mu} \sim \text{MVN}(\boldsymbol{\mu}_0, \boldsymbol{\Omega}_0) \text{ and } \boldsymbol{\Omega}^{-1} \sim \text{Wishart}(\mathbf{S}, \nu). \qquad [3]$$

As Equation 3 shows, the prior distributions $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ have their own parameters, $\boldsymbol{\mu}_0$, $\boldsymbol{\Omega}_0$, $\mathbf{S}$, and ν; these are the mean stress vector, the covariance matrix, the Wishart distribution scale matrix and the number of degrees of freedom, respectively.

Following the development of this generalized MV Bayesian stress model, Javaid et al. (2022a, 2022b) proposed a Bayesian linear regression model for estimating the mean stress vectors as:

$$\mathbf{Y}_{ij} \sim \text{Normal}(\boldsymbol{\mu}_{ij}, \boldsymbol{\omega}_j), \tag{4}$$

where

$$\boldsymbol{\mu}_{ij} = \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}, \tag{5}$$

with

$$\begin{aligned}\boldsymbol{\beta}_{0j} &= [\beta_{0[1]}\ \ \beta_{0[2]}\ \ \beta_{0[3]}\ \ \beta_{0[4]}\ \ \beta_{0[5]}\ \ \beta_{0[6]}]^T \\ &= [\beta_{0[\sigma_x]}\ \ \beta_{0[\tau_{xy}]}\ \ \beta_{0[\tau_{xz}]}\ \ \beta_{0[\sigma_y]}\ \ \beta_{0[\tau_{yz}]}\ \ \beta_{0[\sigma_z]}]^T,\end{aligned} \tag{6}$$

and

$$\boldsymbol{\beta}_j = [\beta_{[\sigma_x]}\ \ \beta_{[\tau_{xy}]}\ \ \beta_{[\tau_{xz}]}\ \ \beta_{[\sigma_y]}\ \ \beta_{[\tau_{yz}]}\ \ \beta_{[\sigma_z]}]_j^T. \tag{7}$$

As Equations 5–7 show, the Cartesian stress components are regarded as independent response variables, with the explanatory variable being depth below ground surface. Therefore, $\boldsymbol{\omega}_j$ in Equation 4 are the estimates of standard deviations for all individual response variables. The terms in Equation 6 represent the value of the stress components at the ground surface, with those in Equation 7 representing the rate of increase of these with respect to depth. We assume $\sigma_z = \tau_{yz} = \tau_{zx} = 0$ at the ground surface, with the result that $\beta_{0[\sigma_z]} = \beta_{0[\tau_{yz}]} = \beta_{0[\tau_{zx}]} = 0$. The number of individually regressed variables and regression parameters for a complete 3D stress tensor are thus six and nine respectively.

3.2 Bayesian linear segmented regression

The Bayesian linear segmented regression allows us to obtain distributions of the breakpoints between adjacent segments. Here, the breakpoints represent stress domain boundaries and as already noted there is considerable uncertainty in the depths of these. By statistical convention, the linear segmented or piecewise regression model is written using a dummy or indicator variable function (e.g. Young 2017). However, the general linear segmented regression model can be explained more conveniently as in the following algorithm given as Equation 8.

$$y = \begin{cases} \beta_0 + \beta_1 x + \varepsilon & x \le \Psi \\ \beta_0 + \beta_1 \Psi + \beta_2 (x - \Psi) + \varepsilon & \text{otherwise} \end{cases} \tag{8}$$

Here, Ψ is the depth of the stress domain boundary and ε is the error in predictive estimates obtained from regression. The model presented in Equation 8 can be expanded to include more breakpoints by including the desired number of Ψ terms. In the Bayesian context (Brilleman et al. 2017, Gelman & Hill 2007), regression coefficients (β_0, β_1) and all the breakpoints (Ψ_i) follow their own distributions rather taking fixed point estimates of centrality or some other statistical characteristic position.

3.3 Our analyses

To test the suitability and efficacy of the Bayesian linear segmented regression method, we have first applied the method to synthetically generated stress data that contains a distinct stress domain boundary at a depth of 100 m. The synthetically generated stress data are shown in Figure 2. These data were generated using tensorial method (Gao & Harrison 2017) by specifying a covariance matrix with very little dispersion and drawing random samples from the MVN distribution of Equation 1. The gradient discontinuity shown in these data could be explained by several plausible geological processes due to locked-in or residual stresses (Zang & Stephansson 2010) arising from removal of ice loading after glacial retreat. The figure shows only σ_x, σ_y and τ_{xy}, because the components σ_z, τ_{yz} and τ_{zx} are considered to have uniform gradients across the

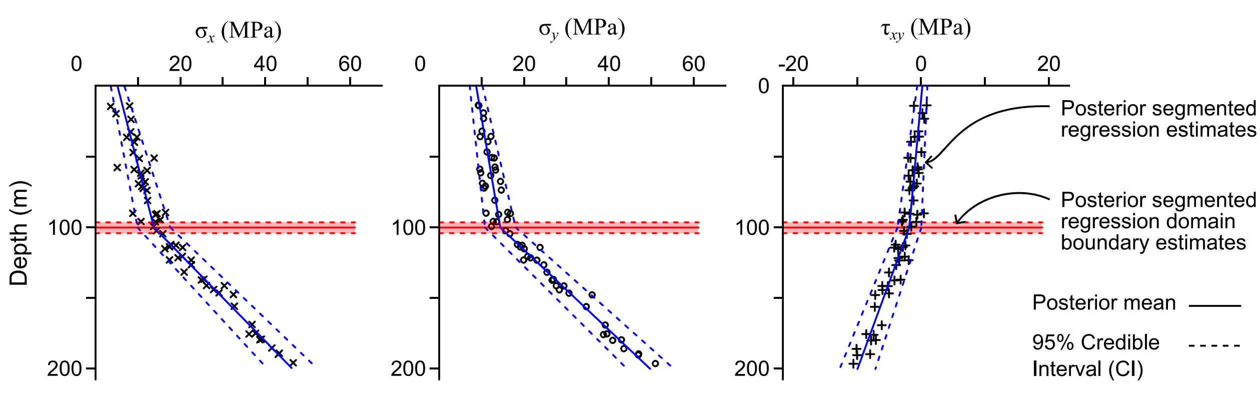

Figure 2. Posterior estimates of mean stresses and stress domain boundary from Bayesian linear segmented regression on synthetic stress data

depth of interest and thus no gradient discontinuities are anticipated.

Bayesian linear segmented regression estimates the posterior distributions of the regression coefficients using the model presented in Equations 4–7. We assume that the regression coefficients and the depth of the stress domain boundary all follow normal distributions. We have used uninformative priors on the regression coefficients, and an informative prior on the stress domain boundary with a mean depth of 100 m and a standard deviation of 10 m. The priors are thus

$$\boldsymbol{\beta}_{0j} \sim \text{Normal}(0, 100), \qquad \boldsymbol{\beta}_{1j} \sim \text{Normal}(0, 100),$$
$$\boldsymbol{\beta}_{2j} \sim \text{Normal}(0, 100), \qquad \Psi \sim \text{Normal}(100, 10). \qquad [9]$$

Discussion on Figure 2 is presented in the following section.

A further analysis was performed using 114 overcoring stress measurements obtained at the Forsmark site in Sweden. As noted earlier, the suggestion has been made that there are stress domain boundaries at depths of 150 m and 300 m (Martin 2007).

The Bayesian linear segmented regression uses the uninformative priors on the regression coefficients provided in Equation 9 together with informative priors of

$$\Psi_1 \sim \text{Normal}(150, 20), \text{ and}$$
$$\Psi_2 \sim \text{Normal}(300, 20). \qquad [10]$$

on the two depth stress domain boundaries.

4 RESULTS AND DISCUSSION

For the analysis of synthetic stress data, the Bayesian posterior estimates of mean stresses for σ_x, σ_y and τ_{xy} are

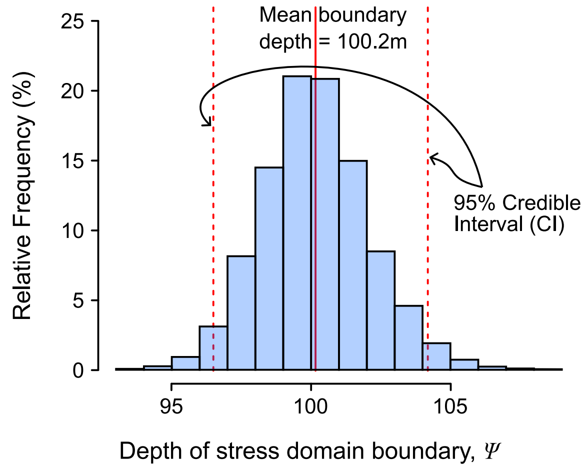

Figure 3. Plot of relative frequency for uncertain domain boundary in synthetic stress data

shown in Figure 2 with the relative frequency plot for the depth stress domain boundary being given in Figure 3. Figure 2 shows the posterior mean and 95% Credible Interval (CI) of (a) the mean stresses σ_x, σ_y and τ_{xy}, and (b) the depth stress domain boundary. As these data were generated using distributions with small variance, the 95% CI for both the mean stresses and the depth stress domain boundary are relatively narrow. The posterior mean of the depth stress boundary is 100.2 m, which is almost equal to the specified value of 100.0 m. The 95% CI for the depth of this boundary is even smaller than the standard deviation used in priors (Equation 9), i.e. <10 m, as shown in Figure 3.

The results from the analysis of the actual Forsmark overcoring stress measurements are provided in Figure 4 and Figure 5. These two plots indicate that the analysis has produced meaningful estimates of both the variation of the

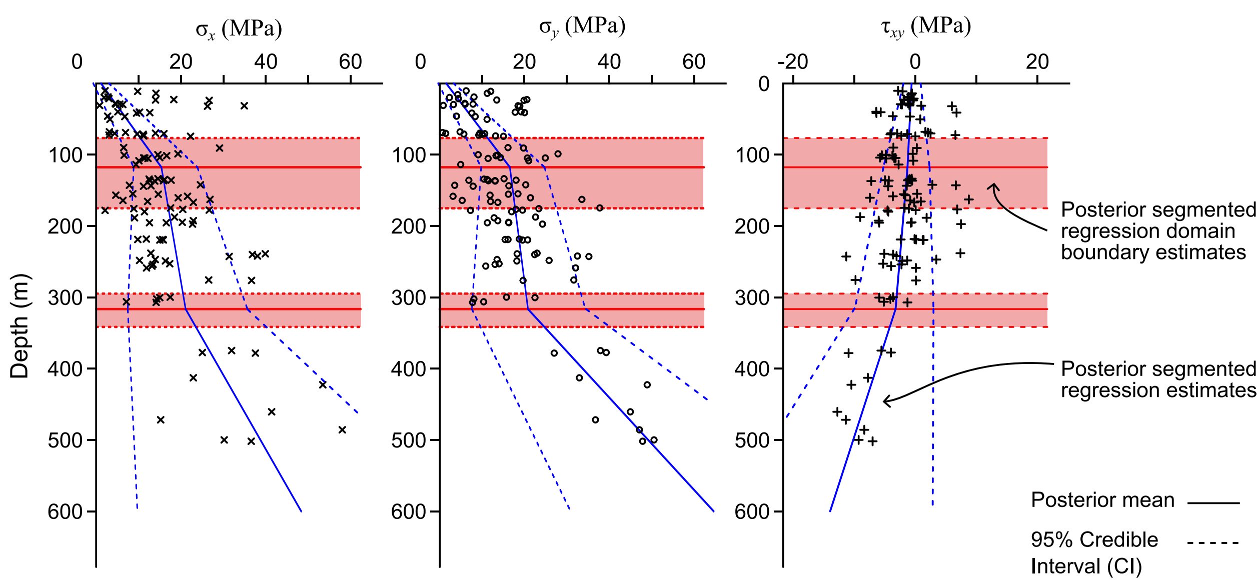

Figure 4. Posterior estimates of mean stresses and stress domain boundary from Bayesian linear segmented regression on Forsmark overcoring stress measurements

stress magnitude with depth and the locations of the stress boundaries. It is noteworthy that the large variability in the data has propagated into correspondingly large 95% CIs for all the estimates.

The posterior means of the upper (Ψ_1) and the lower (Ψ_2) depth stress boundaries are 117.6 m and 316.5 m respectively, which need to be compared to the earlier deterministic estimates of 150 m and 300 m (Martin 2007), respectively. Although these estimates appear to be incompatible, it should be noted from Figure 5 that the deterministic estimates fall within the 95% CI for both the upper and lower stress domain boundary ranges.

This analysis of the Forsmark overcoring stress data demonstrates how deterministic depth estimates of stress domain boundaries can be misleading. However, the Bayesian linear segmented regression generates distributions of likely boundary positions, and these could be used in rock engineering design analysis to fully account for uncertainty in their location. This is particularly important for the design of sensitive projects such as underground nuclear waste repositories.

5 CONCLUSIONS

Partitioning *in situ* stress data into domains based on depth below ground surface is a common procedure in rock engineering, but there is a lack of robust and universally agreed methods to carry out this task. This paper highlights some challenges inherent to the task of characterizing variability and uncertainty in the position of such stress domain boundaries.

Moving median analyses to identify domain boundaries appear to be sensitive to the size interval selected and thus these types of analyses involve some element of subjectivity. Moreover, a lack of stress measurements in any depth range can produce steps in moving median profiles that may falsely be identified as stress domain boundaries.

To overcome the deficiencies associated with moving median analyses we have demonstrated the application of Bayesian linear segmented regression to two datasets: the first comprising synthetically generated stress data and the second actual overcoring stress measurements from the Forsmark site in Sweden.

The synthetic stress data were generated to show two sharply distinct depth stress domains based on different gradients of three Cartesian stress components. The analysis correctly identified the domain boundary, returning values of posterior mean depth and posterior stress gradients very close to the values used in the data generation. Similarly, the 95% CIs of these parameters were very small, indicating confidence in the mean values.

Application of the method to the Forsmark overcoring stress data produced large 95% CIs for both the depth of the domain boundaries and the predicted mean stresses. This was expected, given the large scatter in the measured data. Regarding the depth of the domain boundaries, the analysis returned posterior mean estimates that differed notably from the mean values used in the priors. Nevertheless, the prior mean values – which were obtained using expert opinion – were contained within the 95% CIs of the boundary locations.

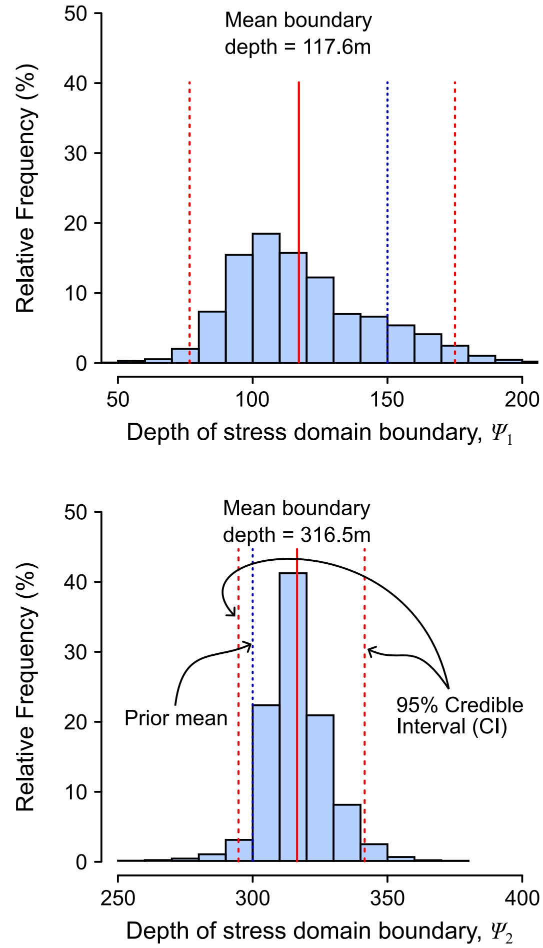

Figure 5. Plot of relative frequency for uncertain domain boundaries for analysis of Forsmark overcoring stress data

Our analyses indicate that Bayesian linear segmented regression is efficacious and overcomes some of the challenges posed by other methods. However, the method requires prior knowledge of both mean locations of boundaries and variability of stress across domains. Examining how site-specific geological knowledge can be applied to this is a subject of our ongoing research.

ACKNOWLEDGEMENTS

We wish to acknowledge the financial support of the Nuclear Waste Management Organization of Canada (NWMO), the Swedish Nuclear Fuel and Waste Management Co of Sweden (SKB), and NSERC of Canada.

REFERENCES

Brilleman, S.L., Howe, L.D., Wolfe, R. and Tilling, K. 2017 Bayesian piecewise linear mixed models with a random change point. *Epidemiology*, 28(6) 827–833 (https://doi.org/10.1097/EDE.0000000000000723)

Feng Y., Bozorgzadeh, N. and Harrison, J.P. 2020. Bayesian analysis for uncertainty quantification. of *in situ* stress data *Int. J. Rock Mech. Min. Sci.* 134 (2020) 104381 (*https://doi.org/10.1016/j.ijrmms.2020.104381*)

Feng, Y. and Harrison, J.P. 2019. Improving estimation of *in situ* stress using a Bayesian approach. *Proc. 7th Int. Symp. Geotech. Safety and Risk (Taipei, Taiwan) ISGSR 2019*, Research Publishing (Singapore)

Feng, Y., Harrison, J.P. and Bozorgzadeh, N. 2021. A Bayesian approach for uncertainty quantification in overcoring stress estimation. *Rock Mech. And Rock. Eng*. 54:627–645 (*https://doi.org/10.1007/s00603-020-02295-w*)

Gao, K. and Harrison, J.P. 2016. Mean and dispersion of stress tensors using Euclidean and Riemannian approaches. *Int. J. Rock Mech. Min. Sci.* 85 165–73 (*http://dx.doi.org/10.1016/j.ijrmms.2016.03.019*)

Gao, K. and Harrison, J.P. 2017. Generation of random stress tensors. *Int. J. Rock Mech. Min. Sci.* 94 18–26 (*http://dx.doi.org/10.1016/j.ijrmms.2016.12.011*)

Gao, K. and Harrison, J.P. 2018a. Multivariate distribution model for stress variability characterisation. *Int. J. Rock Mech. Min. Sci.* 102 144–54 (*https://doi.org/10.1016/j.ijrmms.2018.01.004*)

Gao, K. and Harrison, J.P. 2018b. Scalar-values measure of stress dispersion. *Int. J. Rock Mech. Min. Sci.* 106 234–242. (*https://doi.org/10.1016/j.ijrmms.2018.04.008*)

Gelman, A. and Hill, J. 2007. *Data analysis using regression and multilevel/hierarchical models*., pp. 40-42, Cambridge Press: Cambridge, UK.

Javaid, M.A., Harrison, J.P., Mas Ivars, D. and Kasani, H.A. 2022a. A Bayesian regression analysis of *in situ* stress using overcoring data. *EUROCK 2022: ISRM Symposium on Rock and Fracture Mechanics in Rock Engineering and Mining* Espoo, Finland. (*http://dx.doi.org/10.1088/1755-1315/1124/1/012075*)

Javaid, M.A., Harrison, J.P., Kasani, H.A. and Mas Ivars, D. 2022b. Incorporating an assessment of uncertainty into estimates of *in situ* stress in rock as a function of depth. *GeoCalgary 2022: Canadian Geotechnical Society Conference on Reflection on Resources*, Calgary, Alberta, Canada.

Martin, C.D. 2007. *Quantifying in situ stress magnitudes and orientations for Forsmark*. Forsmark stage 2.2. SKB report no. R-07-26, available online (*https://www.skb.com/publication/1540892*)

Young, D.S. 2017 *Handbook of regression methods*. Boca Raton: Taylor & Francis Group

Zang, A. and Stephansson, O. 2010 *Stress field of the earth's crust*. London: Springer