Acta Geotechnica (2024) 19:3299–3319
https://doi.org/10.1007/s11440-023-02061-6

RESEARCH PAPER

An efficient Bayesian multi-model framework to analyze reliability of rock structures with limited investigation data

Akshay Kumar[1] · Gaurav Tiwari[1]

Received: 15 July 2022 / Accepted: 1 September 2023 / Published online: 4 October 2023

Abstract
Availability of insufficient data is a frequent issue resulting in the inaccurate probabilistic characterization of properties and, finally the inaccurate reliability estimates of rock structures. This study presents a Bayesian multi-model inference methodology which couples multi-model inference with traditional Bayesian approach to characterize uncertainties in both—(1) probability models, and (2) model parameters of rock properties arising due to insufficient data, and to estimate the reliability of rock slopes and tunnels considering their effect. Further, this methodology was coupled with Sobol's sensitivity, metropolis–hastings Markov chain Monte Carlo sampling and moving least square-response surface method to improve the computational efficiency and applicability for problems with implicit performance functions (PFs). Methodology is demonstrated for a Himalayan rock slope (implicit PF) prone to stress-controlled failure in India. Analysis is also performed using recently developed limited data reliability methods, i.e., traditional Bayesian (considers uncertainty in model parameters only) and bootstrap-based re-sampling reliability methods (considers uncertainties in model types and parameters). Proposed methodology is concluded to be superior to other methods due to its capability of considering uncertainties in both model types and parameters, and to include the prior information in the analysis.

Keywords Bayesian inference · Data insufficiency · Multi-model Bayesian inference · Rock structures · Sobol's sensitivity

1 Introduction

Stability analysis of rock structures is a complex problem majorly due to uncertainties in rock properties emanating due to inherent and knowledge-based reasons. Reliability methods provide a suitable alternative to analyze the stability of these structures in the presence of these uncertainties. Several reliability methods like Monte-Carlo simulations (MCSs) [50], first/second-order reliability methods (F/SORMs) [19, 20] and point estimate methods (PEMs) [1] etc., have been developed over the years and are currently used for different in situ problems [2, 42, 52, 54, 60]. Although the mathematical and operational framework of different reliability methods vary, their inputs are the parameters representing the uncertainties in properties, i.e., (1) statistical parameters [i.e., mean and standard deviation (SD)], and (2) best fit probability model. Accurate estimation of these best fit distribution model and model parameters is seldom possible for the rock projects due to availability of insufficient data. This is due to the high costs and significant practical difficulties in lab and in situ testing of rock masses [21, 44, 62]. Frequentist reliability approaches like MCSs, F/SORMs, PEMs, etc., assumes that the model type and distribution parameters estimated from the site-specific test data are deterministic and accurate quantities which is not true in the presence of limited data [31–33, 38, 43]. This in turn raises the question on the accuracy of the analysis based on frequentist reliability approaches.

Traditional Bayesian approaches provide an improvement in the frequentist approaches by considering the model parameters as random variables to consider their uncertainties. This is performed by incorporating the existing or prior available information from various sources with limited site-specific information [7]. Statistical

✉ Gaurav Tiwari
gauravt@iitk.ac.in

Akshay Kumar
akshaykj@iitk.ac.in

[1] Department of Civil Engineering, Indian Institute of Technology (IIT) Kanpur, Kanpur, Uttar Pradesh, India

inference is employed to determine a single "best" probability model and identified sole model is used to make inference from the data. These methods, however, suffer a major limitation of ignoring the uncertainty associated with probability model selection. It is difficult to identify a unique best probability model in the presence of a limited data set, rather than a large data set. In rock mechanics, most of the studies employing these approaches are restricted to explicitly model the uncertainties in intact rock properties and empirical models only [3–5, 8, 9, 12, 17, 57, 58]. Very limited studies are available in literature employing these approaches for analyzing the stability of practical rock slopes and tunnels as shown in Table 1. Further, the available studies neglect the uncertainty related to the probability model and consider the uncertainties in the parameters of the unique "best" model. Therefore, a methodology is required which can quantify both the model type and model parameters uncertainties in rock properties and propagate it to the reliability estimates of rock structures.

Table 1 A summary of the studies on the application of Bayesian approach for the reliability analysis of rock structures

References	Year	Rock structure	Description
Feng and Jimenez [22]	2015	Tunnel	Predicted the squeezing and time-dependent convergence
Li et al. [34]	2016	Slope	Provided a stage-by-stage performance updating framework with new monitoring information for progressive excavation
Zhou et al. [66]	2017	Slope	Performed a probabilistic kinematic analysis for the discontinuity-controlled rock slope instabilities
Aladejare and Wang [6]	2018	Slope	Investigated the influence of rock property correlation, i.e., between cohesion and friction angle, on the stability
Feng et al. [23]	2019	Tunnel	Proposed an approach to improve time-dependent convergence predictions by updating them with new information provided through successive convergence measurements
Zhao et al. [65]	2021	Tunnel	Presented a Bayesian back analysis approach to determine the mechanical parameters of the rock mass
Chang et al. [16]	2022	Tunnel	Proposed a probabilistic model to predict tunnel convergence by combining empirical model and relevance vector machine

In this spirit, a Bayesian multi-model inference (BMMI) approach is developed to consider the effect of uncertainties related to probability model type and model parameters in input properties on the reliability estimates of rock slopes and tunnels in the presence of limited test data. To attain the objective, the multi-model inference approach [13], wherein a set of possible probability models for a rock property can be identified to incorporate the uncertainty associated with model type, is coupled with the traditional Bayesian approach. In addition, this methodology employs Sobol's global sensitivity and the moving least square-response surface method (MLS-RSM) to enhance its computational efficiency and applicability for analyzing the problems with both explicit/implicit performance functions (PFs). The developed approach was demonstrated for a Himalayan rock slope prone to stress-controlled failure (single implicit PF) in India. A comparative assessment is also made with the results of the other methods currently in use for estimating engineering reliability in the presence of limited data—(1) traditional Bayesian approach (considers only model parameter uncertainty), and (2) recently developed resampling reliability method coupling bootstrap re-sampling with frequentist approaches (considers both model type and parameters uncertainties) [31, 32, 43].

2 Details of the components of methodology

This section explains the components involved in the present methodology.

2.1 Bayesian inference

The Bayesian inference is used to quantify the uncertainty associated with the parameters of the probability model (M). Bayesian inference treats model parameters ($\boldsymbol{\theta}$) probabilistically (as a random variable) by combining prior knowledge in the formulation via prior distribution ($p(\boldsymbol{\theta};M)$) and site-specific observation data ($\boldsymbol{d}$) expressed in terms of the likelihood function ($p(\boldsymbol{d}|\boldsymbol{\theta},M)$. The posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{d},M)$, i.e., the updated probability density function of parameters $\boldsymbol{\theta}$ according to the Baye's rule can be written as given below [7].

$$p(\boldsymbol{\theta}|\boldsymbol{d},M)=\frac{p(\boldsymbol{d}|\boldsymbol{\theta},M)p(\boldsymbol{\theta};M)}{p(\boldsymbol{d};M)}\propto p(\boldsymbol{d}|\boldsymbol{\theta},M)p(\boldsymbol{\theta};M) \quad (1)$$

where $p(\boldsymbol{d};M)$ is known as the normalizing factor or evidence used to make the cumulative distribution function (CDF) of posterior distribution equals to one and can be estimated as given below.

$$p(\boldsymbol{d};M) = \int p(\boldsymbol{d}|\boldsymbol{\theta},M)p(\boldsymbol{\theta};M) \tag{2}$$

The likelihood function $p(\boldsymbol{d}|\boldsymbol{\theta},M)$ for a property with observation data $\boldsymbol{d} = \{d_1.d_2,\ldots,d_N\}$ which is being modeled using probability distribution model M having parameters $\boldsymbol{\theta}$ can be formulated as below [7].

$$\begin{aligned} p(\boldsymbol{d}|\boldsymbol{\theta},M) &= f(d_1|\boldsymbol{\theta}) \times f(d_2|\boldsymbol{\theta}) \times \cdots \times f(d_n|\boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} f(d_i|\boldsymbol{\theta}) \end{aligned} \tag{3}$$

where $f(d_i|\boldsymbol{\theta},M)$ is the PDF value corresponding to model probability distribution M evaluated at the i-th data point d_i, and $\boldsymbol{\theta}$ represents the parameters of the M.

Estimation of posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{d},M)$ involves multi-dimensional integration and analytical solutions are computationally complex and in-efficient, except for the conjugate prior distributions. To overcome this issue, a numerical process known as Markov chain Monte Carlo (MCMC) simulation can be employed to generate a sequence of random samples from the posterior distribution [6]. Details of MCMC sampling are presented in Sect. 2.3.

2.2 Multimodal selection and model uncertainty

This methodology employs the multi-model inference approach [13, 63] to incorporate the uncertainty associated with the model identification of properties. Kullback–Leibler (K–L) information theory-based akaike information criterion (AIC) was used for the model selection in this study. The probability model with minimum AIC value is considered as a best fit model to represent the data. For small datasets, an updated AIC, i.e., $\mathrm{AIC_c}$ has been developed as given below [26].

$$\mathrm{AIC_c} = -2\log\left(p\left(\boldsymbol{d}|\hat{\theta},M\right)\right) + 2K + \frac{2K(K+1)}{N-K-1} \tag{4}$$

where $p(\boldsymbol{d}|\hat{\theta},M)$ is the likelihood function given the maximum likelihood estimate of the parameters $\hat{\theta}$, K is the number of parameters of the candidate model M, and N is the sample size of the data set. $\mathrm{AIC_c}$ should be used when $\frac{N}{K} < \sim 40$ and $\mathrm{AIC_c}$ converges to AIC with the increasing value of N [13, 63]. Given the small datasets, $\mathrm{AIC_c}$ is utilized for multi-model selection in this study. Once the $\mathrm{AIC_c}$ value is known for each model in the candidate set, the AIC differences values ($\boldsymbol{\Delta}_\mathrm{A}$) can be calculated to interpret a ranking of candidate models as given below [13].

$$\Delta_\mathrm{A}^{(i)} = \mathrm{AIC_c}^{(i)} - \mathrm{AIC_c}^{\min} \tag{5}$$

where $\mathrm{AIC_c}^{\min}$ is the minimum of the $\mathrm{AIC_c}^{(i)}$ values, $i = 1,2,3,\ldots,N_d$ and N_d is the total number of candidate models in the set. This transformation forces the best model to have $\Delta_\mathrm{A}^{(i)} = 0$ and all the other models to have positive values. Further, the likelihood of the model M_i can be expressed as $\exp\left(-\frac{1}{2}\Delta_\mathrm{A}^{(i)}\right)$ and by normalizing these likelihoods the $\mathrm{AIC_c}$ based model probabilities $\mathfrak{p}_i$ can be estimated as given below.

$$\mathfrak{p}_i = p(M_i|d) = \frac{\exp\left(-\frac{1}{2}\Delta_\mathrm{A}^{(i)}\right)}{\sum_{m=1}^{N_d}\exp\left(-\frac{1}{2}\Delta_\mathrm{A}^{(m)}\right)} \tag{6}$$

Larger the $\mathfrak{p}_i$ or lesser the $\Delta_\mathrm{A}^{(i)}$, the more plausible is model being the best fit for the given dataset among the candidate models. Therefore, $\mathfrak{p}_i$ and $\Delta_\mathrm{A}^{(i)}$ can be utilized to rank the models among the candidate sets.

2.3 Metropolis–hastings (MH) MCMC sampling

MCMC sampling was used in this study due to its feasibility to generate samples from an arbitrary distribution particularly when the density function is difficult to express analytically [11, 45]. Samples are obtained by exploring the entire domain of uncertain parameters based on their prior distributions. The MCMC was coupled with the Bayesian inference to allow any general prior distribution selection where the posterior distribution becomes complex and difficult to express analytically. Among the available options, the metropolis–hastings (MH) algorithm was employed in this study due its simplicity and efficiency to implement the MCMC sampling for the posterior of parameter $\boldsymbol{\theta}$ [25, 39]. The procedure of the MH algorithm can be summarized as given below.

(i) At stage $t = 1$, assume an initial value of parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_1$, which can be chosen randomly from the prior distribution or may simply be assigned the mean value.

(ii) At any subsequent stage t ($t = 2,3,\ldots$), generate a new proposal value $\boldsymbol{\theta}^*$ from a proposal distribution $T(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{t-1})$, which is considered to be multivariate Gaussian distribution for simplicity with mean value $\boldsymbol{\theta}_{t-1}$ and standard deviation $\boldsymbol{s}$ [6, 56, 59]. Parameter $\boldsymbol{s}$ is also known as tunning parameter or width/scaling of proposal distribution.

(iii) Calculate the ratio of the posterior density for the candidate ($\boldsymbol{\theta}^*$) and current ($\boldsymbol{\theta}_{t-1}$) values as shown below.

$$\boldsymbol{r} = \min\left(\frac{p(\boldsymbol{\theta}^*|\boldsymbol{d},M)T(\boldsymbol{\theta}^*,\boldsymbol{\theta}_{t-1})}{p(\boldsymbol{\theta}_{t-1}|\boldsymbol{d},M)T(\boldsymbol{\theta}_{t-1},\boldsymbol{\theta}^*)},1\right) \tag{7}$$

For the symmetric proposal distributions, i.e., $T(\boldsymbol{\theta}^*,\boldsymbol{\theta}_{t-1}) = T(\boldsymbol{\theta}_{t-1},\boldsymbol{\theta}^*)$, the ratio $\boldsymbol{r}$ is reduced to the expression as shown below.

$$\boldsymbol{r} = \min\left(\frac{p(\boldsymbol{\theta}^*|\boldsymbol{d},M)}{p(\boldsymbol{\theta}_{t-1}|\boldsymbol{d},M)},1\right) \tag{8}$$

(iv) Draw a random number $\boldsymbol{u}$ from the uniform distribution $\boldsymbol{U}(0,1)$, i.e., $\boldsymbol{u} \sim \boldsymbol{U}(0,1)$.

(v) Accept the candidate state $\boldsymbol{\theta}_t = \boldsymbol{\theta}^*$, if $\boldsymbol{r} \geq \boldsymbol{u}$. Otherwise, the current value is taken as the next value $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$.

(vi) Repeat step (ii)–(v) till the target number of samples (i.e., $\boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \ldots$) are obtained.

The choice of initial value and proposal distribution have strong influence on the convergence of the Markov chain toward stationary condition. The initial samples may be discarded as burn-in samples, as they might not be completely valid depending upon the assumption of initial value. The optimal selection (neither too high nor too low) of tunning parameter or scaling of proposal distribution can be made graphically through the trace plot (i.e., Markov chain sample versus sample number) and autocorrelation plot (i.e., autocorrelation function (ACF) against increasing lag values) of the simulated Markov chain samples. For a stationary Markov chain, trace plot should look like a hairy caterpillar (not show apparent anomalies) and autocorrelation plot should show a quick, exponential decrease between the correlation of sample. Beside visual inspection, the performance of chain can be assessed numerically through acceptance rate (i.e., percentage of accepted samples). An acceptance rate in between 20 and 40% is considered sufficiently efficient [24]. A MATLAB code was written to generate random samples from the posteriors.

2.4 Bootstrap re-sampling reliability approach

Bootstrap re-sampling approach is usually employed to quantify the effect of statistical uncertainties associated with both model type and parameters arising due to small size of sample and to estimate its effect on the response of structures by coupling this approach with frequentist reliability approach, i.e., MCSs [31, 32, 43]. In this method, a large number of re-constituted samples from the original data sample are generated which are assumed to be in close resemblance to the original sample. These surrogate samples are generated through random sampling with replacement from the original sample. Figure 1 shows an example of a bootstrap re-constituted sample for the variable $\boldsymbol{X} = \{X_1, X_2, \ldots, X_N\}$ with N sample data points. For each re-constituted sample every data point in the original sample has an equal probability of being chosen. Size of re-constituted samples is kept same as that of original sample that avoids any biasness resulting into sample statistics [29]. From these re-constituted samples, the bootstrap statistics [i.e., bootstrap mean μ_{B_α} and bootstrap standard deviation (SD) $\sigma_{\mathrm{B}_\alpha}$] for a sample statistical parameter α (i.e., sample mean and sample SD) quantifying the statistical uncertainty can be obtained as following.

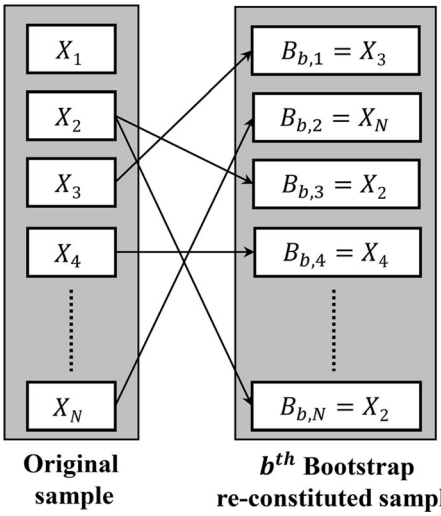

Fig. 1 Procedure to generate re-constituted samples using bootstrap approach from the original sample

$$\mu_{\mathrm{B}_\alpha} = \frac{1}{B}\sum_{b=1}^{N_\mathrm{B}} \alpha_b; \sigma_{B_\alpha} = \sqrt{\frac{1}{N_\mathrm{B}-1}\sum_{b=1}^{N_\mathrm{B}} \left(\alpha_b - \mu_{\mathrm{B}_\alpha}\right)^2} \tag{9}$$

where α_b is the sample statistical parameter estimated for the bth bootstrap re-constituted sample and N_B is the total number of re-constituted samples. For each re-constituted sample, traditional reliability analysis was performed via MCSs, and statistical parameters of P_f were estimated using standard statistical methods.

2.5 Sensitivity analysis

Sensitive properties were identified with significant effect on the PF using sensitivity analysis. This aided the improvement in the computational efficiency of the methodology by considering the uncertainties in sensitive properties only. Sobol's global sensitivity analysis (GSA) [48] was employed in this study due to its efficiency to explore the whole space of input variables. This is not possible in the local sensitivity analysis. Total order/effects (S_{T_i}) quantifying the relative contributions of input properties on the output variability can be estimated as given below.

$$S_{T_i} = 1 - \frac{V[E(Y|\boldsymbol{X}_{\sim i})}{V(Y)} \tag{10}$$

where E and V are the expectation and variance; X_i is the ith input parameter; $\boldsymbol{X}_{\sim i}$ represents the components of input vector $\boldsymbol{X}$ except X_i. Saltelli's MCSs based numerical method was employed to estimate S_{T_i} in which quasi-random samples for $\boldsymbol{X}$ were generated and arranged in the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as shown below.

$$\boldsymbol{A} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & \cdots & x_n^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_1^{(k)} & x_2^{(k)} & \cdots & x_i^{(k)} & \cdots & x_n^{(k)} \end{bmatrix} \tag{11}$$

$$\boldsymbol{B} = \begin{bmatrix} x_{n+1}^{(1)} & x_{n+2}^{(1)} & \cdots & x_{n+i}^{(1)} & \cdots & x_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n+1}^{(k)} & x_{n+2}^{(k)} & \cdots & x_{n+i}^{(k)} & \cdots & x_{2n}^{(k)} \end{bmatrix} \tag{12}$$

where k is the base sample and n is the input vector dimension. $\boldsymbol{C_i}$ is the matrix containing elements of $\boldsymbol{B}$ except the ith column, which is taken from $\boldsymbol{A}$.

$$\boldsymbol{C_i} = \begin{bmatrix} x_{n+1}^{(1)} & x_{n+2}^{(1)} & \cdots & x_i^{(1)} & \cdots & x_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n+1}^{(k)} & x_{n+2}^{(k)} & \cdots & x_i^{(k)} & \cdots & x_{2n}^{(k)} \end{bmatrix} \tag{13}$$

The output column vectors $\boldsymbol{Y_A}, \boldsymbol{Y_B}$ and $\boldsymbol{Y_{C_i}}$ are constructed by evaluating the PF, i.e., $\boldsymbol{Y} = G(\boldsymbol{X})$, for the realizations of $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C_i}$, respectively. S_{T_i} for the input X_i then can be calculated by employing Janon estimators [28] as given below.

$$S_{T_i} = 1 - \frac{\left(\frac{1}{k}\right)\sum_{j=1}^{k} y_B^{(j)} y_{C_i}^{(j)} - \left(\left(\frac{1}{k}\right)\sum_{j=1}^{k}\left[\frac{y_B^{(j)}+y_{C_i}^{(j)}}{2}\right]\right)^2}{\left(\frac{1}{k}\right)\sum_{j=1}^{k}\left[\frac{\left(y_B^{(j)}\right)^2+\left(y_{C_i}^{(j)}\right)^2}{2}\right] - \left(\left(\frac{1}{k}\right)\sum_{j=1}^{k}\left[\frac{y_B^{(j)}+y_{C_i}^{(j)}}{2}\right]\right)^2} \tag{14}$$

where $y_A^{(j)}, y_B^{(j)}$ and $y_{C_i}^{(j)}$ are the jth element of column vectors Y_A, Y_B and Y_{C_i}, respectively.

2.6 Moving least square-response surface method (MLS-RSM)

MLS-RSM was used to obtain an explicit surrogate relationship between input–output for the problems lacking explicit PFs, thus eliminating the requirement of repeated numerical simulations. This issue will be shown in the later sections. MLS-RSM can be mathematically expressed as given below [37].

$$\hat{G}(\boldsymbol{X}) = \boldsymbol{p}(\boldsymbol{X})\boldsymbol{a}(\boldsymbol{X}) \tag{15}$$

where $\boldsymbol{p}(\boldsymbol{X}) = \left[1 x_1 x_2 \ldots x_d x_1^2 x_2^2 \ldots x_d^2\right]_{1\times m}$ is a quadratic polynomial basis of function ($m = 2d + 1$). $\boldsymbol{a}(\boldsymbol{X})$ is a set of unknown coefficients, which is dependent on the $\boldsymbol{X}$ and can be determined, as given below.

$$\boldsymbol{a}(\boldsymbol{X}) = \boldsymbol{A}_{m\times m}^{-1}\boldsymbol{B}_{m\times h}\boldsymbol{Y}_{h\times 1} \tag{16}$$

where $\boldsymbol{Y} = [G(\boldsymbol{X}_1)G(\boldsymbol{X}_2)\ldots G(\boldsymbol{X}_h)]^{\mathbf{T}}$ is the matrix of known PF values obtained from the opted solution technique. Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be written as given below.

$$\boldsymbol{A}(\boldsymbol{X}) = \boldsymbol{P}_{m\times h}^{TT}\boldsymbol{W}_{h\times h}\boldsymbol{P}_{h\times m}; \boldsymbol{B}(\boldsymbol{X}) = \boldsymbol{P}_{m\times h}^{\mathrm{T}}\boldsymbol{W}_{h\times h} \tag{17}$$

where

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{p}(\boldsymbol{X}_1) \\ \boldsymbol{p}(\boldsymbol{X}_2) \\ \ldots \\ \boldsymbol{p}(\boldsymbol{X}_h) \end{bmatrix}_{h\times m}; \boldsymbol{W} = \begin{bmatrix} w_1(\boldsymbol{X}) & 0 & 0 \\ 0 & \ldots & 0 \\ 0 & 0 & w_h(\boldsymbol{X}) \end{bmatrix}_{h\times h} \tag{18}$$

where $w_i(\boldsymbol{X})$ is the spline weighting function (C^1 continuous) with compact support as shown below.

$$w_i(\boldsymbol{X}) = \begin{cases} 1 - 6r_i^2 + 8r_i^3 - 3r_i^4; & r \le 1 \\ 0; & r > 1 \end{cases} \tag{19}$$

where $r = \boldsymbol{X} - \boldsymbol{X}_{i2}/l_i$, l_i is the influence domain size chosen as twice the distance between $(1 + 2d)$th sample point and design point $\boldsymbol{X}$, d is the number of random variables and h is the number of sampling points. Latin Hypercube Sampling (LHS) based design of experiments technique was used for generating sampling points (random input vectors realizations) from input parameter distributions [40]. Nash–Sutcliffe efficiency (NSE) index was adopted to assess the accuracy of the RSM [41]. NSE was evaluated by estimating the PF values using original solving technique and RSM, i.e., $G_i^{\text{original}}(\boldsymbol{X})$ and $\hat{G}_i^{\text{RSM}}(\boldsymbol{X})$ at p random off-sample points of input properties generated via the LHS, as given below.

$$\text{NSE} = 1 - \left[\frac{\sum_{i=1}^{p}\left(G_i^{\text{original}}(\boldsymbol{X}) - \hat{G}_i^{\text{RSM}}(\boldsymbol{X})\right)^2}{\sum_{i=1}^{p}\left(G_i^{\text{original}}(\boldsymbol{X}) - G_i^{\text{mean}}(\boldsymbol{X})\right)^2}\right] \tag{20}$$

where $G_i^{\text{mean}}(\boldsymbol{X})$ is the mean value of $G_i^{\text{original}}(\boldsymbol{X})$. The RSM is rated *very good* for NSE value 0.75–1.0.

3 Methodology

This section explains the steps involved in the proposed BMMI methodology. As mentioned earlier, the main idea of the BMMI methodology is to couple the multi-model inference with traditional Bayesian and probabilistic tools. Figure 2 shows the flowchart explaining the implementation steps for the methodology.

As mentioned earlier, analysis performed by the proposed methodology was also compared with the recently developed methodologies employed for the reliability analysis in the presence of limited data. Two such methodologies, i.e., traditional Bayesian methodology and bootstrap re-sampling reliability were used for the analysis in the next section. Hence, the steps involved in these

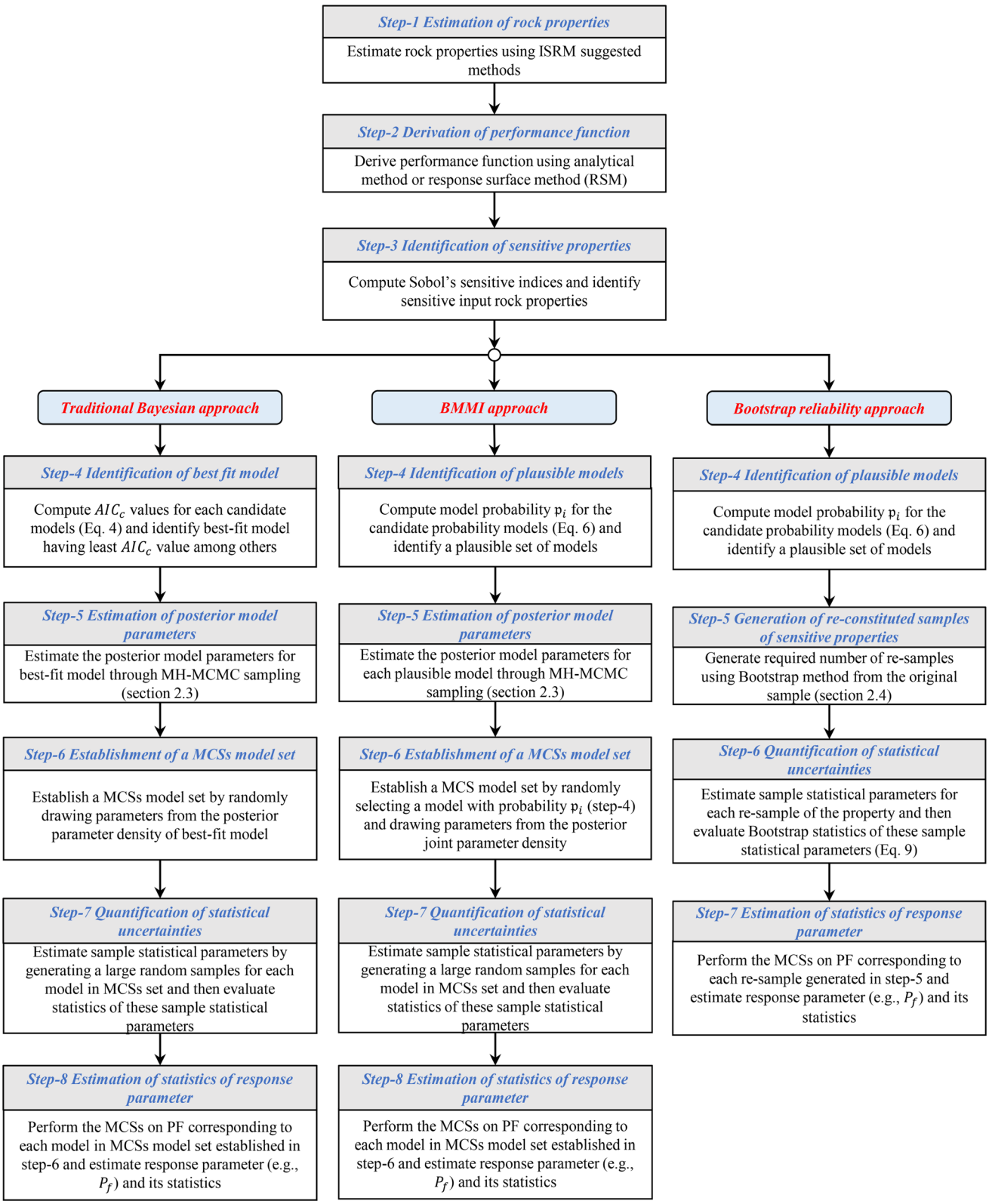

Fig. 2 Flowchart to implement the proposed Bayesian multi-model inference (BMMI) approach along with the traditional Bayesian and bootstrap re-sampling reliability approach

methodologies are also briefly summarized in Fig. 2. A MATLAB code was written to implement all the above steps sequentially in analyzing the stability of rock slopes.

4 Application example

Application example used in this study is a rock slope supporting the piers of world's highest Chenab railway bridge in the Jammu and Kashmir, India. The major reason to select the case study for this study is the well-studied geological and geotechnical properties of the rock mass at the site. Slope under consideration is a large slope with dimensions 293 × 196 m. Rock mass at the site was heavily jointed dolomite (unit weight = 25 kN/m^3) intersected by three major joint sets along with some random joint sets. Slope was adjudged to be prone to stress-controlled failure due to very close joint spacing and large dimensions. More details on the geology, location and geotechnical of the slope can be found in the literature [52]. Analysis was performed using different methodologies via the steps shown in flowchart in Fig. 2.

4.1 Analysis using Bayesian multi-model inference (BMMI) methodology

4.1.1 Step 1: estimation of rock properties

Intact rock and rock joint properties were estimated for this site using the standard tests conducted as per ISRM suggested guidelines [27]. Table 2 shows the statistics of the properties relevant to this study [52]. The original sample of rock properties contains 22 data points only, which are statistically small and insufficient.

4.1.2 Step 2: derivation of performance function (PF)

PF for the slope stability is usually expressed in terms of factor of safety (FOS). Slope under consideration is prone to stress-controlled failure and hence, the analytical formulation of PF was not possible. Hence, an explicit surrogate relationship between input rock properties and output response parameter (i.e., FOS) was derived using MLS-RSM. A total of $h = 200$ sampling points of input properties were generated using LHS based on their statistics (Table 2). FOSs were estimated for these realizations using Shear Strength Reduction (SSR) [18] in Phase^2 [46] by assuming rock as elastic perfectly plastic Hoek–Brown material. A typical finite element model of the slope prepared in Phase^2 is shown in Fig. 3. Vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ were determined using the realizations of input properties and corresponding FOSs (Sect. 2.6). NSE of the constructed MLS-RSM was 0.9748 for p = 50 off-sample points and hence, the performance of the RSM was rated to be as *very good*.

It is important to note that the authors have used the moving least square-response surface method (MLS-RSM) to construct an explicit expression between the input rock properties and FOS. MLS-RSM uses a locally weighted regression approach to fit a surface to the input–output data. In this method, the coefficients of the monomials change for every observation, which makes it difficult to write an explicit expression unlike the polynomial RSMs where the coefficients of the monomials remain constant for every observation [30, 37].

Table 2 Statistical parameters and best fit probability distribution (PD) model of rock properties from the original sample

Property	Mean	SD	Best fit PD
Hoek–Brown parameter m_i	12.7880	4.4288	Weibull
Uniaxial compressive strength (UCS) (MPa)	115.3636	48.7653	Loglogistic
Elastic modulus (E_i) (GPa)	64.5000	19.4881	Lognormal
Geological strength index (GSI)	41.1364	6.4535	Inverse Gaussian

4.1.3 Step 3: identification of sensitive properties

Sensitive properties were identified using Sobol's GSA and only the identified sensitive properties were considered for BMMI. Sobol's analysis was performed using the PF derived in the previous step. A total of $k = 10^5$ quasi-random samples were used for the Sobol's analysis (Sect. 2.5). Results are summarized in Fig. 4. S_{T_i} of UCS and GSI were estimated to be 30–64% more than those of E_i and m_i indicating their high sensitivities. Hence, the Bayesian analysis was performed by considering statistical uncertainties in UCS and GSI only while considering E_i and m_i as random variables.

4.1.4 Step 4: identification of plausible models

Initially, the candidate probability models for UCS and GSI were decided with the requirement that their values (i.e., $\boldsymbol{x}$) are always non-negative and real (i.e., $\boldsymbol{x} \in [0, \infty)$) (Table 3). This is due to non-negative nature of rock properties under consideration. Table 3 shows the estimated AIC$_c$ and probability $\mathfrak{p}_i$ values for these models using Eqs. (4)–(6). Figure 5 also shows the histogram of data for UCS and GSI along with the candidate models pdfs. For UCS, the candidate probability models have approximately similar AIC$_c$ or probability $\mathfrak{p}_i$ values, except exponential model. Further, the values of $\Delta_A^{(i)}$ were estimated to be minimal for these models indicating that any of

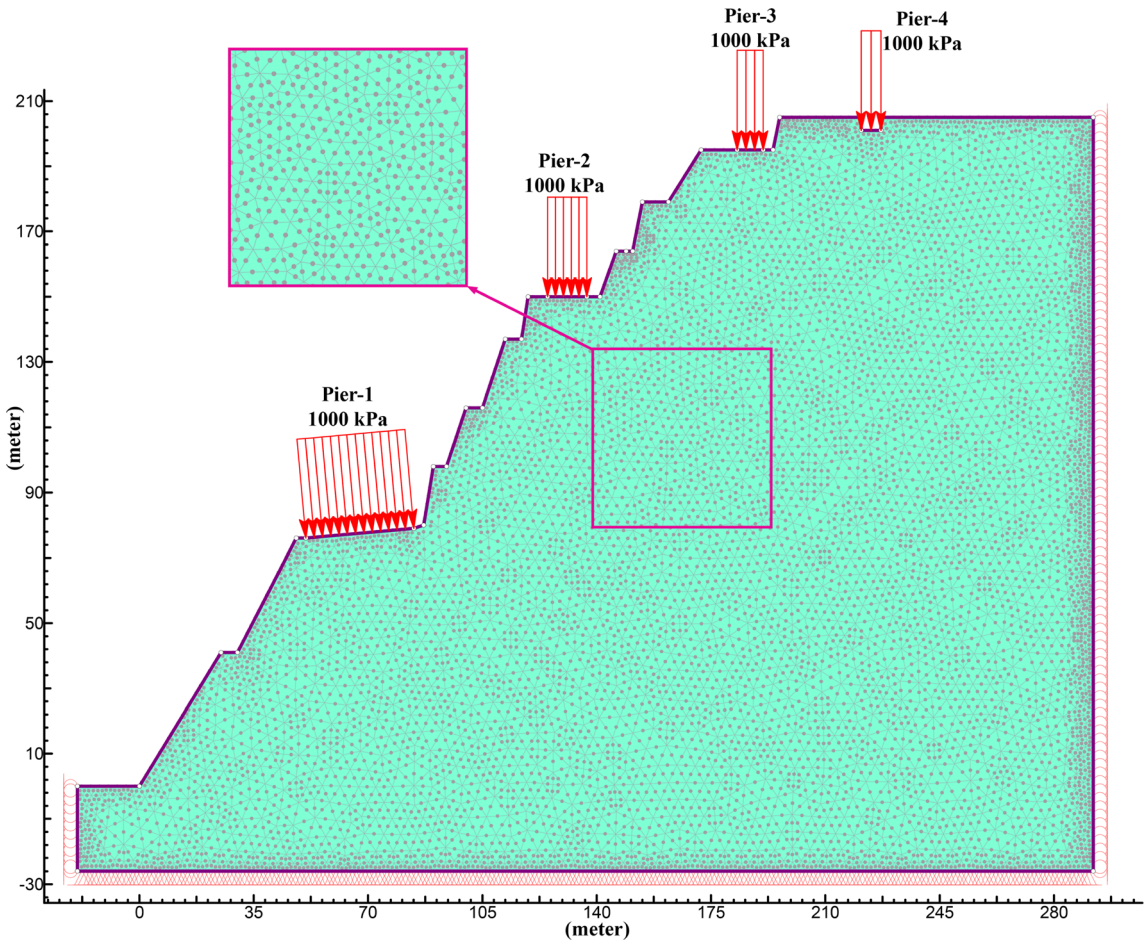

Fig. 3 Typical numerical model prepared in Phase[2] for the stability analysis of rock slope

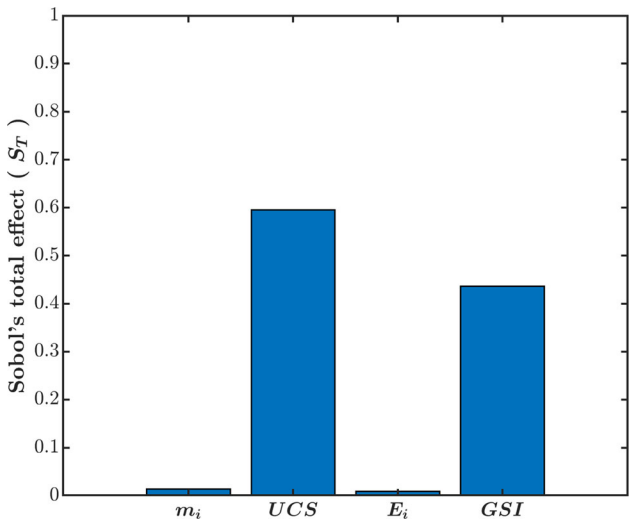

Fig. 4 Sobol's indices indicating the sensitivities of rock properties toward the FOS of the rock slope

Table 3 Candidate probability distribution models and their corresponding AIC_c, AIC_c difference $\Delta_A^{(i)}$ and AIC_c based probability $\mathfrak{p}_i$ values for the UCS and GSI

Candidate model	UCS			GSI		
	AIC_c	$\Delta_A^{(i)}$	$\mathfrak{p}_i$	AIC_c	$\Delta_A^{(i)}$	$\mathfrak{p}_i$
Rayleigh	235.604	0.294	0.172	180.773	**33.823**	**0.000**
Exponential	255.116	**19.806**	**0.000**	209.743	**62.793**	**0.000**
Lognormal	237.393	2.083	0.070	147.000	0.050	0.223
Weibull	236.092	0.782	0.135	150.483	3.533	0.039
Gamma	235.399	0.089	0.191	147.212	0.262	0.200
Inverse Gaussian	238.373	3.063	0.043	146.950	0.000	0.228
Nakagami	235.420	0.110	0.189	147.562	0.612	0.168
Loglogistic	235.310	0.000	0.200	147.899	0.949	0.142

Text in bold represents the candidate models with $\Delta_A^{(i)} > 10$ or approximately zero $\mathfrak{p}_i$ values

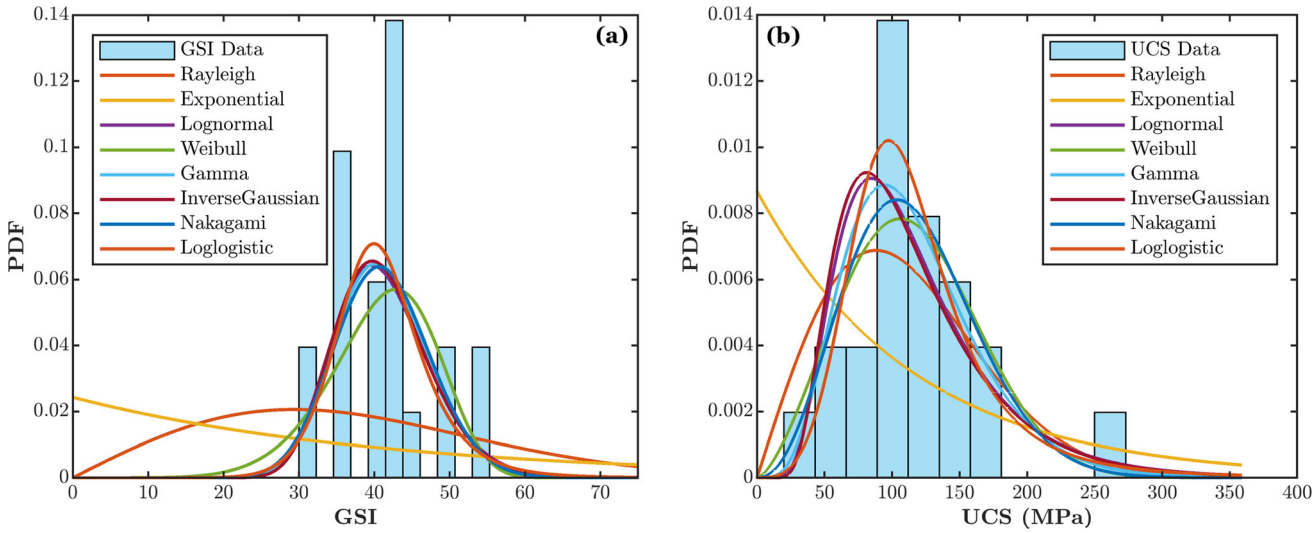

Fig. 5 Histogram plot of the original sample along with the candidate probability models for the **a** GSI and **b** UCS

these models can be considered to fit the data satisfactorily. Similar observations were made for GSI regarding AIC_c, $\mathfrak{p}_i$ and $\Delta_A^{(i)}$ values of candidate models, except Rayleigh and exponential models. This reinforces the argument that the identification of a best fit model is impractical from the small sized sample. The models with $\Delta_A^{(i)} < 10$ [13] have much higher probabilities to be considered as the best fit model to represent the data satisfactorily. Due to this reason, all the models mentioned above (Table 3) except exponential model for UCS, and Rayleigh and exponential models for GSI were considered as plausible models for further analyses.

Further, an analysis has been performed to emphasize the effect of sample size on the determination of best fit model. For this analysis, $\mathfrak{p}_i$ values for candidate models were determined for the samples of sizes (N_s) varying between 22 (size of original data) to 10^6 for GSI. Samples of different sizes were obtained by generating random realizations from the model having least AIC_c for original dataset (best fit for original data). Best fit model form original data of GSI was found to be inverse Gaussian (Table 2). Figure 6 shows the analysis results. It can be observed that the $\mathfrak{p}_i$ values for candidate models were of approximately similar magnitudes for small sized samples ($N_s \leq 300$ approximately). As the sample size increases, lognormal and inverse Gaussian were found to have approximately similar $\mathfrak{p}_i$ values for $N_s \leq 8 \times 10^4$. This coincidence of $\mathfrak{p}_i$ values for these distributions up to a large N_s value could be due to approximately similar AIC_c values from original sample. As N_s increases beyond this, $\mathfrak{p}_i$ values for inverse Gaussian and lognormal distributions monotonically increased and decreased, respectively. The $\mathfrak{p}_i$ value for inverse Gaussian distribution became unity for $N_s = 10^6$. It can be concluded that N_s should be significantly higher ($N_s \geq 8 \times 10^4$ for this case) to assess the best fit model with acceptable certainty and much higher with complete certainty ($N_s = 10^6$) which is practically impossible.

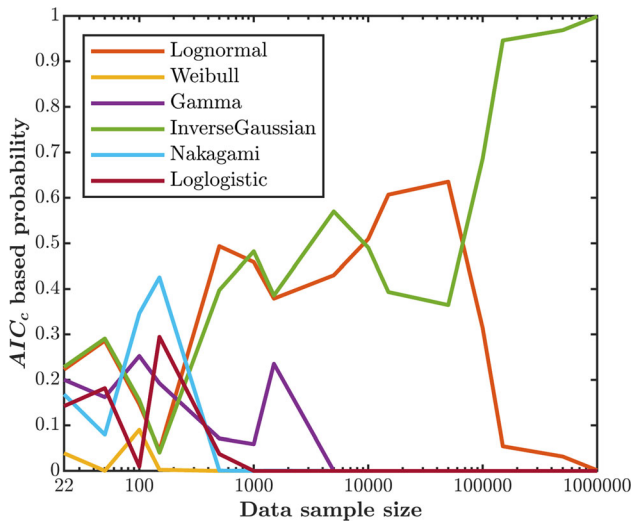

Fig. 6 Variation of AIC_c based probability $\mathfrak{p}_i$ with the size of data sample of the plausible models for GSI

4.1.5 Step 5: estimation of posterior model parameters

Once the plausible models were known, the posterior model parameters for these models (i.e., $p(\boldsymbol{\theta}|\boldsymbol{d}, M)$) were estimated through Bayesian inference utilizing MH-MCMC sampling (Sect. 2.3). The prior distribution was taken as the uniform distribution, i.e., $p(\boldsymbol{\theta}; M)$ [6, 57, 58]. The bounds of mean and standard deviation (SD) for UCS and GSI were adopted from the literature [5]. Aladejare

Table 4 Prior information/knowledge for the sedimentary rock properties [5]

Statistical parameter	UCS (MPa)		GSI	
	Min	Max	Min	Max
Mean	4.4	264	22.60	60.40
SD	0.0176	289.344	3.86	16.308

and Wang [5] summarized the typical ranges of mean and SD of properties for sedimentary rocks based on extensive literature review ($\sim$ 135 research articles). Table 4 summarizes the prior values for mean and SD of UCS and GSI. To perform MCMC analysis, the range of model parameters corresponding to each plausible model, i.e., $\boldsymbol{\theta}$, were estimated from the bounds of mean and SD using standard relations [7]. Table 5 summarizes the estimated prior range of model parameters. MCMC analysis was then performed by generating a total of 5×10^4 random samples. Initial 5×10^4 samples were discarded by considering them as burn-in samples identified via visual inspection of trace plots. The convergence of simulated chains was assessed through the trace and autocorrelation plots. Trace plots did not show any anomalies and autocorrelation plot showed an exponential decrease between the sample correlation. Further, the acceptance rate of all Markov chains simulated was within 20–40% for each model parameter. Figure 7 shows the typical trace and autocorrelation plots of the Markov chain samples of lognormal distribution parameters for the GSI. It was observed that the simulated Markov chains are stationary for GSI. Similar observations were made for all the plausible models of other properties.

Figure 8 also shows the joint probability density functions (jpdf) and the marginal pdfs of model parameters for plausible models of GSI quantifying the uncertainties in the parameters. The red points along the jpdf show the parameter values estimated from the original data. It can be concluded that significant uncertainties exist in the model parameters in the presence of limited data and should be considered in the analysis.

4.1.6 Step 6: establishment of a MCSs model set

In this step, the MCSs model set of UCS and GSI were constructed based on the $\mathfrak{p}_i$ values of candidate models. The $\mathfrak{p}_i$ values were considered as the weighting factor which is the ratio of the number of times a candidate model pdf was generated to the total generated pdfs (M). For this study, a total (M) of 10^4 models for the UCS and GSI were generated. For example, a value of $\mathfrak{p}_i = 0.172$ for a model (for e.g., Rayleigh model for UCS) implies that this model was generated for 1720 ($\mathfrak{p}_i \times M = 0.172 \times 10^4 = 1720$) times out of 10^4 models. Associated model parameters for the model were chosen at random from the posterior distribution of model parameters (Fig. 8 for GSI) evaluated in the previous step. Figure 9 shows the results of total generated models (i.e., MCSs model set) of the plausible models for UCS and GSI. These MCSs model sets were used to quantify the uncertainties in the statistics of properties (i.e., mean and SD) and response parameter (probability of failure, P_f) of the slope from small sample size in the next steps.

Table 5 Range of plausible model parameters to define the uniform prior distribution for the UCS and GSI

Candidate model	Model parameter	UCS		GSI	
		min	max	min	Max
Rayleigh	Parameter 1 (λ)	3.510	210.6415	–	–
Lognormal	Parameter 1 (a)	0	5.575949	2.9083	4.0989
	Parameter 2 (b)	6.67E−05	2.893485	0.0638	0.6474
Weibull	Parameter 1 (μ)	1.27E−146	264.0044	24.1935	66.502
	Parameter 2 (ω)	0.0106	34,295.34	1.4251	19.8231
Gamma	Parameter 1 (a)	0.0002	2.25E+08	1.92	244.8495
	Parameter 2 (b)	1.17E−06	19,027.26	0.2466	11.7706
Inverse Gaussian	Parameter 1 (μ)	4.4	264	22.6	60.4
	Parameter 2 (σ)	0.0010	5.94E+10	43.3927	14,788.91
Nakagami	Parameter 1 (μ)	0.5	56,250,000	0.8808	61.5876
	Parameter 2 (σ)	19.3603	153,416	525.6596	3914.176
Loglogistic	Parameter 1 (μ)	1.4816	5.575949	3.1179	4.1009
	Parameter 2 (λ)	3.68E−05	36.25539	0.0352	0.3978

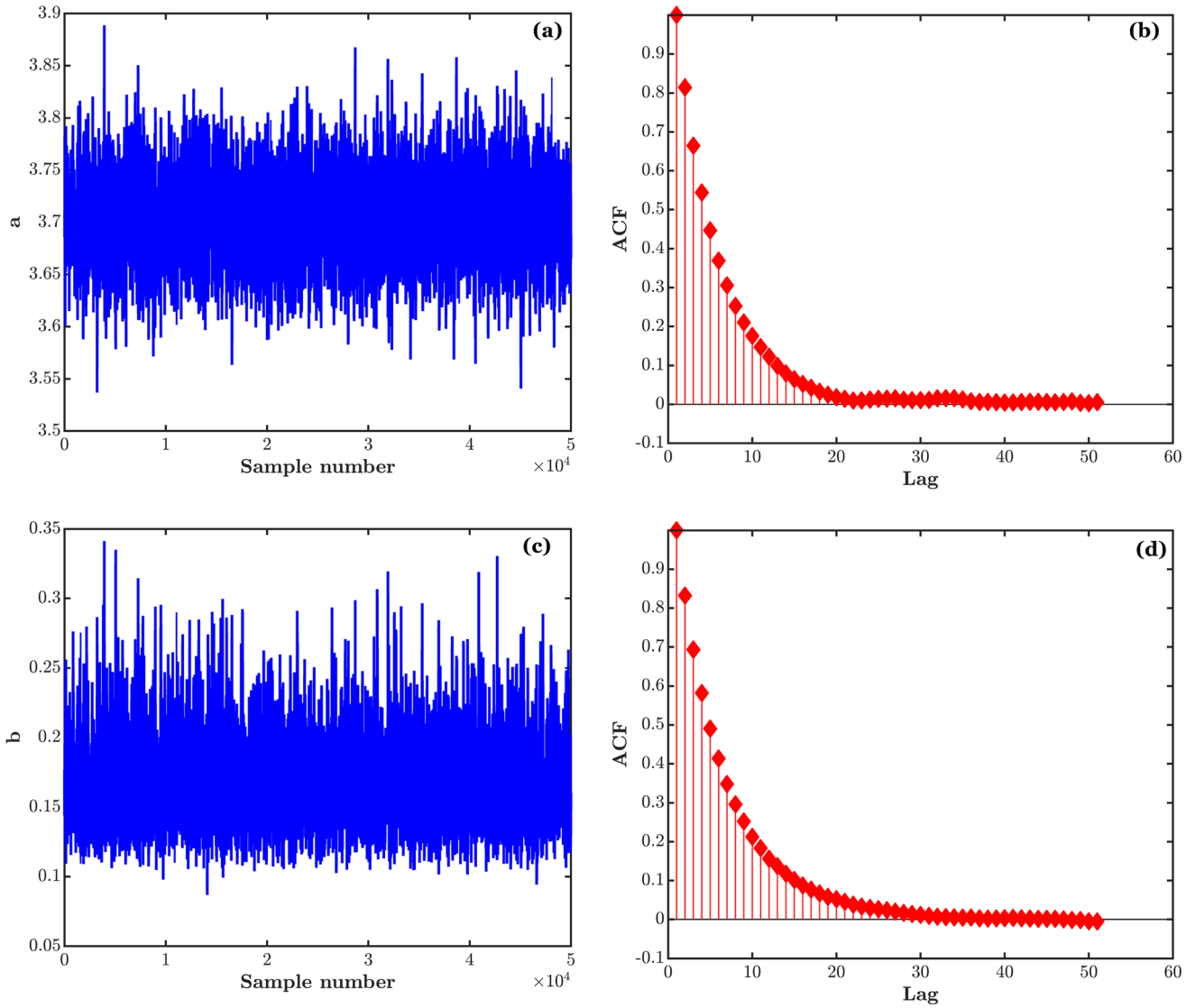

Fig. 7 Convergence analysis of the Markov chains **a** trace plot for parameter 1 'a' **b** autocorrelation plot for parameter 1 'a' **c** trace plot for parameter 2 'b' and **d** autocorrelation plot for parameter 2 'b' of the lognormal probability model for GSI

4.1.7 Step 7: quantification of statistical uncertainties

Sample statistics, i.e., mean and SD, of the property were estimated for total generated models in the previous step via standard relations between model parameters of pdfs and moments. This results in a total of 10^4 sample statistics values from which their mean and SD were evaluated. Table 6 and Fig. 10 summarize the analysis results. It can be observed that the means of the sample statistics were coinciding with the sample statistics estimated for the original sample (Table 2). Further, the SDs of the sample statistics indicate the statistical uncertainties in them due to small size of samples.

4.1.8 Step 8: estimation of statistics of response parameter

In this step, a model from the MCSs set constructed in the previous step was selected and traditional reliability analysis was performed resulting the values of P_f corresponding to the model in the MCSs set. This step is repeated for all models sequentially in the MCSs sets ($M = 10^4$) resulting into a total of 10^4 P_f values from which the statistics P_f were estimated. Traditional reliability analysis for the model in set was performed by carrying out MCSs on the PF (prepared in step 2). MCSs was performed by generating 5×10^4 random samples based on the statistics of selected model from the MCSs set of UCS and GSI and the best fit models of m_i and E_i (Table 2). P_f was considered as the area under the pdf of FOS with value less than 1. Figure 11 and Table 7 show the analysis results

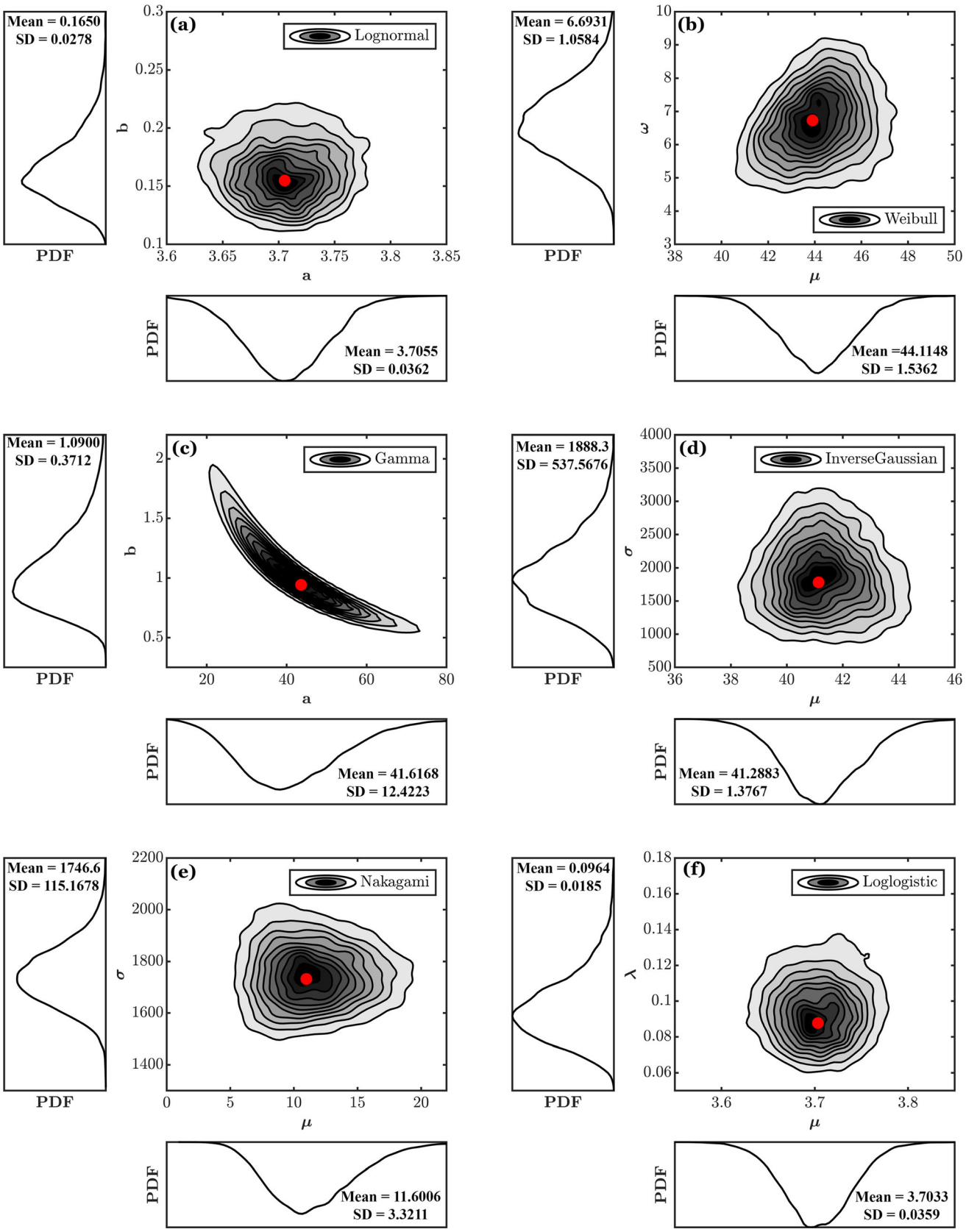

Fig. 8 Marginal and joint pdf of the posterior model parameters for the plausible models **a** Lognormal **b** Weibull **c** Gamma **d** Inverse Gaussian **e** Nakagami and **f** Loglogistic of GSI

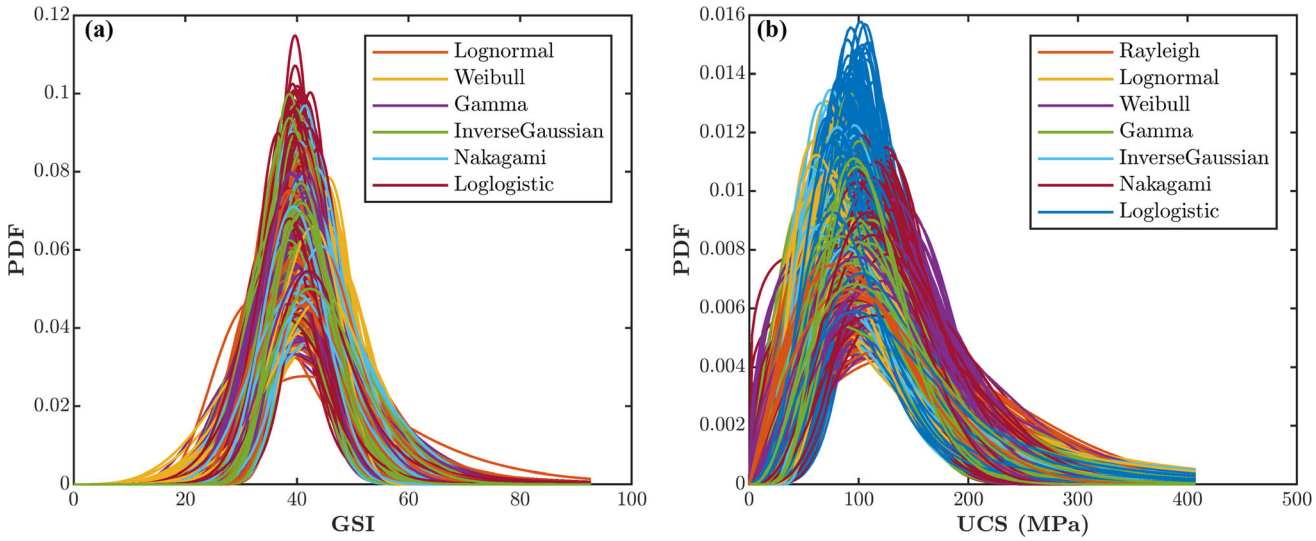

Fig. 9 MCSs model sets with a total of 10^4 models established for the **a** GSI and **b** UCS for the BMMI approach

Table 6 Statistics of the sample statistics for UCS and GSI estimated via different reliability methodologies

Parameter	Reliability methodology	UCS (MPa)		GSI	
		Mean	SD	Mean	SD
Sample mean	BMMI	115.9484	11.8974	41.1439	1.4561
	Traditional Bayesian	117.4880	11.5548	41.2311	1.3389
	Bootstrap reliability	115.1048	10.2045	41.1226	1.3124
Sample SD	BMMI	44.8957	20.0365	6.6459	1.1739
	Traditional Bayesian	52.2661	9.6956	6.2886	1.0025
	Bootstrap reliability	46.2528	10.6103	6.2222	0.9109

including empirical CDFs P_f, respectively. The SD signifies the effect of statistical uncertainties on the P_f due to small size samples. The expected performance level of the slope in accordance with probability descriptions provided by the USACE [55] was mapped in the range of *good to unsatisfactory*.

An important point is that the uncertainties in the sample statistics and P_f could be affected by the $\Delta_A^{(i)}$ values of plausible models of a property. Data of a property with multiple plausible probability models having very small $\Delta_A^{(i)}$ (close to zero) values may have higher effect of model type uncertainty on the sample statistics and P_f. Reason is the significant mixing of plausible models (approximately equal contribution from multiple plausible models) in the established MCSs model set which may eventually lead to higher uncertainties in the sample statistics and P_f.

4.2 Analysis using traditional Bayesian methodology

Initial three steps involved in the traditional Bayesian methodology are same as that of the BMMI methodology. The candidate probability models for UCS and GSI were first chosen similar to BMMI. AIC_c values were estimated for these models [Eq. (4)] corresponding to the original sample. Best fit models for GSI and UCS (with least AIC_c values) were estimated to be inverse Gaussian and loglogistic, respectively (Table 3).

Posterior model parameters of the best fit models for the UCS and GSI were estimated through Bayesian inference utilizing MH-MCMC sampling by generating 5×10^4 random samples. The major difference, compared to the BMMI, is that the analysis was performed only for best fit models instead of all plausible models. Analysis details are similar as explained in BMMI (step 5). Figure 8d shows the jpdf and marginal pdfs of inverse Gaussian model parameters (best fit for GSI) estimated from the MH-MCMC sampling quantifying the uncertainties associated with them. Further, the MCSs model sets having $M = 10^4$ models based on the best fit models were constructed for the UCS and GSI. The major difference is that all models in the MCSs set were corresponding to best fit models instead of the mixing of plausible models. Figure 12 shows the constructed MCSs model sets corresponding to best fit models for UCS and GSI.

Statistical uncertainties in the UCS and GSI were quantified by estimating the sample statistical parameters corresponding to each model in the MCSs set via standard

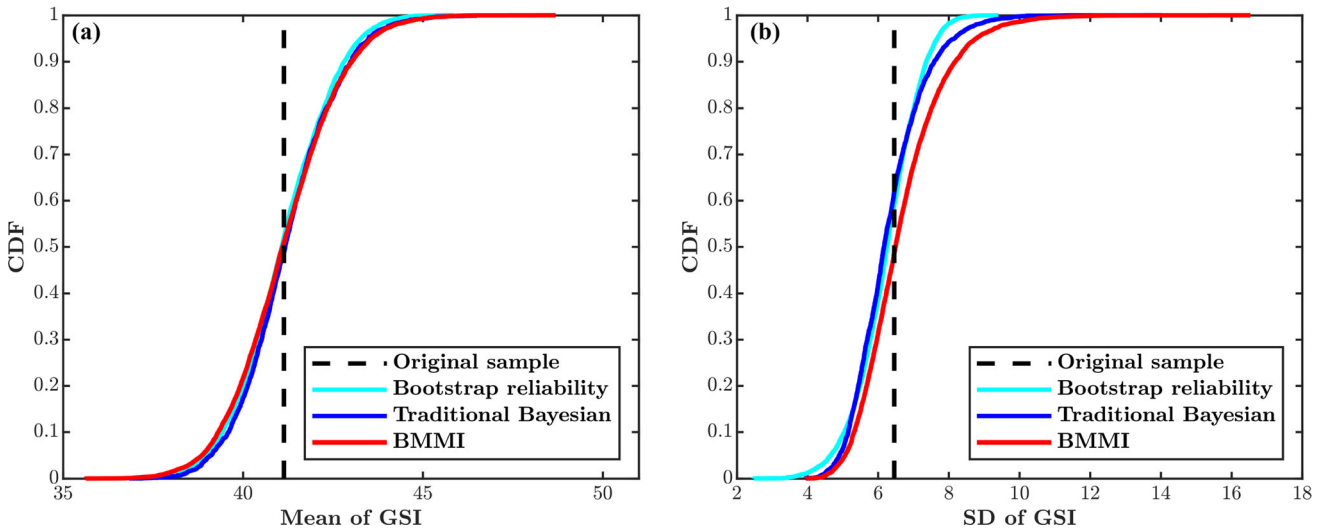

Fig. 10 Empirical CDFs of **a** mean of GSI and **b** SD of GSI estimated via different reliability methodologies

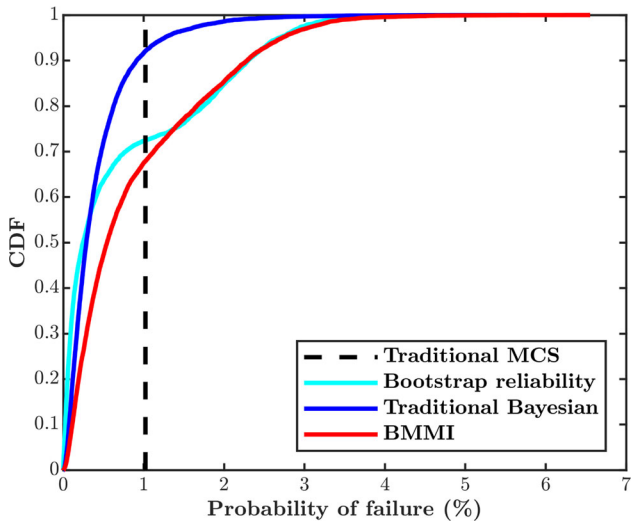

Fig. 11 Empirical CDF of the probability of failure P_f (%) for the rock slope via different reliability methodologies

Table 7 Statistics of probability of failure (P_f) for the rock slope estimated via various reliability methodologies

Reliability methodology	Mean of P_f	SD of P_f	Confidence interval [2.5–97.5%]	Expected performance
BMMI	0.9046	0.8875	[0.0480–3.1050]	Good–unsatisfactory
Traditional Bayesian	0.4246	0.4489	[0.0330–1.6780]	Good–poor
Bootstrap reliability	0.7447	0.9439	[0.0060–2.9830]	Good–unsatisfactory

relations between model parameters of pdfs and moments. Table 6 and Fig. 10 provide the results obtained from the traditional Bayesian approach. It can be observed that the means of the sample statistics were coinciding with sample statistics estimated from the original sample (Table 2). Statistical uncertainties due to the small size of sample is indicated by the SDs of sample statistics. Finally, the traditional reliability analysis was performed corresponding to each model ($M = 10^4$) in the MCSs model sets for UCS and GSI constructed in the previous step. Analysis details are similar as explained for the BMMI (step 8). Figure 11 and Table 7 show the analysis results including empirical CDF of P_f. The effect of statistical uncertainties due to small size samples is indicated by the SD of P_f. From this approach, the expected performance level of the slope was mapped in the range of *good to poor*.

4.3 Analysis using bootstrap re-sampling reliability methodology

Initial procedures in the bootstrap reliability methodology are same to that of BMMI methodology. In this step, a total of $N_B = 10^4$ number of bootstrap re-constituted samples were generated for sensitive properties (i.e., UCS and GSI). Re-constituted samples were generated from their original samples as explained in Sect. 2.4.

Quantification of statistical uncertainties in the statistics of sensitive properties was done by estimating their bootstrap statistics. Sample statistics were estimated for individual re-constituted sample of GSI and UCS. From these values, the bootstrap statistics (mean and SD) of the sample statistics were evaluated using Eq. (9). Table 6 and Fig. 10 summarize the results. It was observed that bootstrap means of the sample statistics were coinciding with the

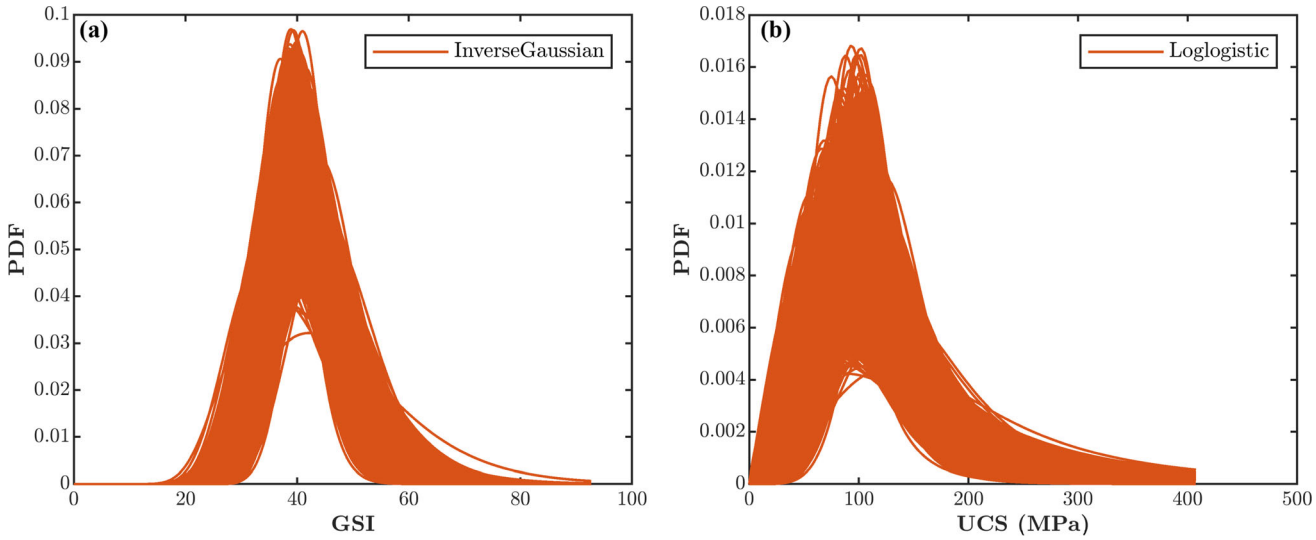

Fig. 12 MCSs model sets with a total of 10^4 models established for the **a** GSI and **b** UCS for the traditional Bayesian reliability methodology

ones estimated for the original sample (Table 2). Further, significant bootstrap SDs were observed for sample statistics of both properties signifying the statistical uncertainties in sensitive properties invoked by the small size of samples. Finally, the traditional reliability analysis was performed for each re-constituted sample of the UCS and GSI generated resulting in a value of P_f corresponding to individual re-constituted sample. Firstly, the best fit model with least AIC_c value was selected. Then, the traditional reliability analysis was performed ($N_B = 10^4$ times) via MCSs on the PF by generating 5×10^4 random samples. Random samples were generated based on the statistics of re-constituted samples of UCS and GSI and the best fit models of m_i and E_i. Figure 11 and Table 7 show the analysis results. The SD signifies the effect of statistical uncertainties on P_f due to small size samples. From this approach the expected performance level of the slope was mapped in the range of *good to unsatisfactory*.

5 Discussions

It is well-known that the rock projects often suffer with the availability of insufficient data of rock properties. Limited data restricts the capability of rock designers to accurately perform the probabilistic characterization of input properties. It is very difficult to state a precise threshold number of data points to classify the terminology "minimum number" precisely. Ruffolo and Shakoor [47] observed that for a 95% confidence interval and a maximum of 20% acceptable strength deviation from the mean of UCS of rocks, 10 UCS samples are needed to be tested. Some studies state this threshold number to be 30 [38, 61]. These guidelines are very crude as they lack statistical proof and are based on limited data. Tang et al. [51], based on their detailed statistical study, concluded that this "minimum number" could be 54–458 for COV ranging from 0.3 to 0.1 for the marginals and 25–2577 for correlation coefficient varying from − 0.9 to − 0.1 for the copula of geotechnical properties, respectively.

An analysis was performed for the present case study to determine the quantity of data required to obtain the best fit model and convergence of model parameters for input properties precisely. Figures 6 and 13 show the analysis results. The observable differences between AIC values of candidate models could initially be observed for a sample size of ∼ 500 and the clear identification of the best-fit model could be made for ∼ 100,000 samples. Amount of data required to obtain the convergence for model parameters was significantly lower (∼ 10^3) than that to determine the best fit model accurately (10^6). Overall, it is highly impractical to perform this quantity of lab/in situ tests to obtain the precise statistics of geotechnical properties invoking statistical uncertainties which are required to be considered in the analysis as suggested in the proposed methodology which could be used for very limited data of inputs. However, it is still practically impossible to perform this amount of lab and in situ testing required to determine even the parameters of the model (∼ 10^3) accurately. This is due to practical difficulties (such as sample collection, sample disturbance, site preparation and data interpretation), high costs, time consumption, etc., involved in the rock testing. Hence, it is inaccurate to assume that the best fit model and its parameters estimated from the limited site-specific test data are "true estimates" of the statistics of rock mass properties under consideration. This is even true for the high budget rock projects where the quantity of in situ and laboratory testing is often

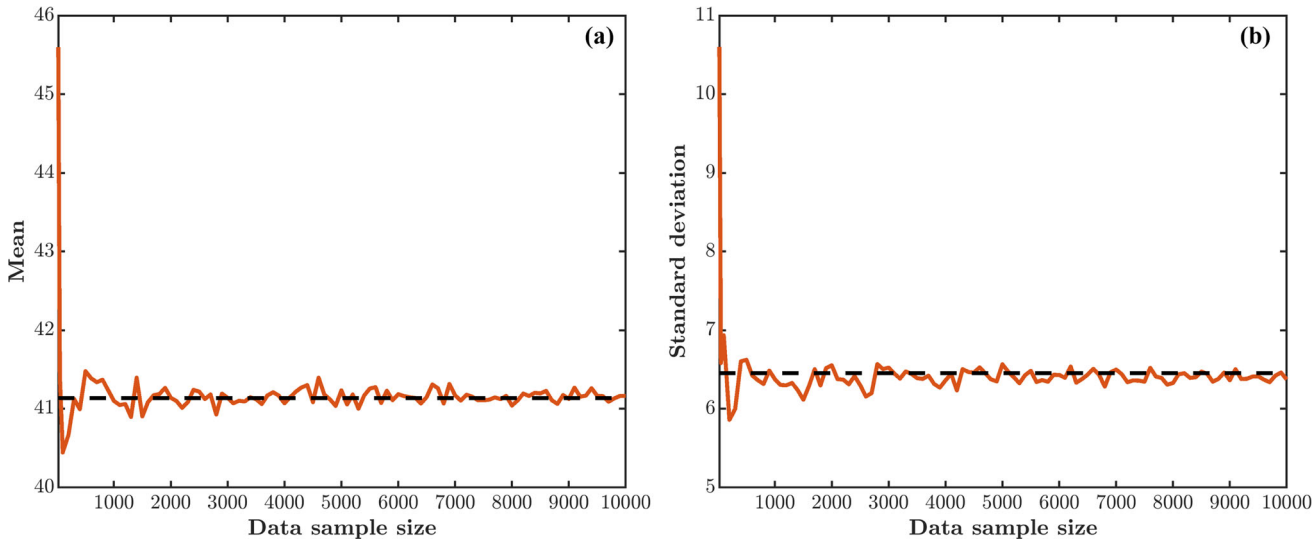

Fig. 13 Variation of the **a** mean of GSI and **b** SD of GSI with the size of data sample

limited. For example, the number of in situ plate load tests conducted for Chenab Bridge in India and Kazunogawa hydro power cavern in Japan to determine the rock mass deformability were approximately 20–30 [14, 53]. Condition is worser for small budget rock projects, where a very limited amount is spent on rock investigation. Under these conditions, analysis performed by the traditional reliability methods are highly inaccurate which assumes that the best fit model and its parameters estimated from the limited test data are "true estimates" of the population parameters. Previous section demonstrated the proposed BMMI methodology for the reliability analysis of a rock slope with limited data of input properties.

5.1 Comparative analyses

BMMI methodology differs from the traditional Bayesian approach as it can consider the uncertainties in both model type and associated parameters. It was observed that the uncertainty associated with model type has significant effect on the total uncertainty of input properties along with that of response parameter (i.e., P_f) for the rock slope. While the mean values of sample statistics were matching well with each other (0.21–16.42% difference) for the case study, the major difference was observed in the SDs of the statistics of properties (2.88–51.61% difference) as shown in Table 6. SD was significantly lower for the traditional Bayesian approach compared to BMMI signifying the underestimation of uncertainties in the statistics of properties majorly due to ignorance of model type uncertainty in traditional Bayesian approach. These underestimated uncertainties propagated during the estimation of statistics of P_f also. Statistics of P_f were estimated to be approximately 50% lower (Table 7) for traditional Bayesian as compared to BMMI methodology emphasizing the importance of considering the uncertainty in model type along with uncertainties of model parameters.

In contrast, bootstrap-based methodology can consider the uncertainties in both model types and parameters like BMMI methodology. For this criterion (considering uncertainties in both model and parameters), both bootstrap-based method and BMMI are superior to traditional Bayesian method. While the mean values of sample statistics were matching with each other (0.05–3.02% difference) for the case study, the major difference was observed in the SDs of the statistics of properties (9.87–47.05% difference) as shown in Table 6. SDs were significantly lower for the bootstrap-based methodology as compared to BMMI. The reason could be the difference in the analysis procedure adopted in these methodologies. Bootstrap method estimates the sampling distributions of properties by resampling (with replacement) from the original sample (data at hand) and creating many bootstrap samples [49]. There is no provision of inclusion of prior information in this method and standard deviation may only fluctuate in a range defined by the data of original sample only. In contrast to this, BMMI includes the prior information in the estimation of statistics of input properties. This prior information is usually collected from literature. Prior information can have wide range (hence significant SD) in it since data are collected from the wide variety of sites around the world as observed for this case study. This may contribute to the higher SD in the estimated statistics of input properties. These uncertainties propagated during the estimation of statistics of P_f also. Statistics of P_f were underestimated by 6.35–17.68% by bootstrap reliability methodology as compared to the BMMI (Table 7) for the considered case study. Effect of

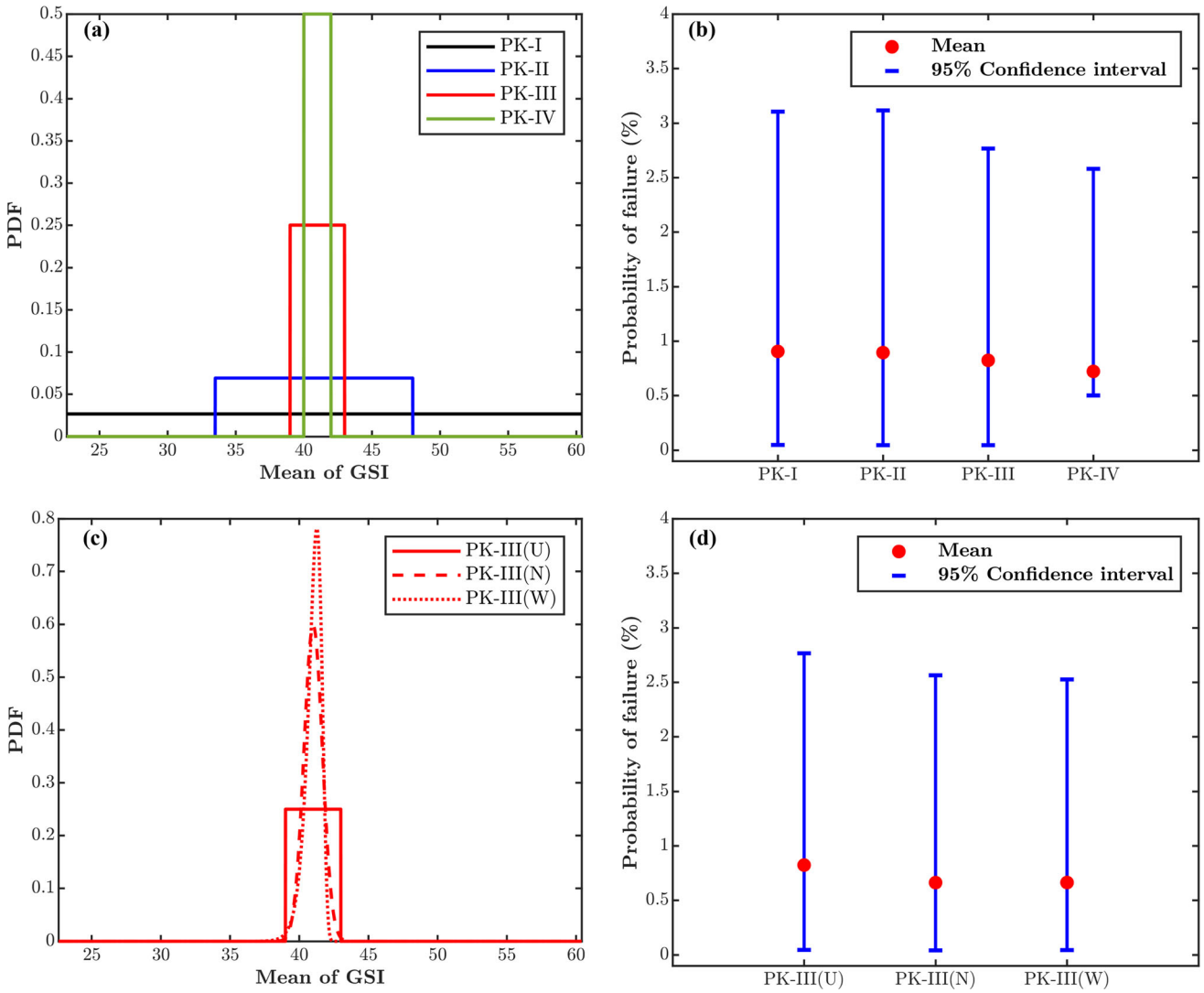

Fig. 14 **a** Prior knowledge–I (PK–I), PK–II, PK–III and PK–IV with uniform distribution for Mean of GSI **b** probability of failure P_f (%) corresponding to PK–I, PK–II, PK–III and PK–IV with uniform distribution from BMMI **c** PK–III with three prior distribution, Uniform (U), Normal (N) and Weibull (W) [i.e., PK–III(U), PK–III(N) and PK–III(W)] for Mean of GSI and **d** Probability of failure P_f (%) corresponding to PK–III(U), PK–III(N) and PK–III(W) from BMMI

prior knowledge on the statistics of input properties and P_f are discussed in next section in detail.

A comparison was also made with the recently proposed Bayesian Model Averaging (BMA) methodology. Details of the methodology can be seen in the literature [35, 64]. The BMA starts from identifying the plausible probability models for all the inputs. Then, a set of models comprising all possible combinations from plausible probability model are constructed and their corresponding fitting probabilities are estimated. Then, for each combination the Bayesian inference is performed and samples from the posterior distribution of model parameters are generated via MCMC sampling. Next, corresponding to each of these samples, the probability of failure (P_f) is estimated using MCSs by generating random realizations of inputs. Thus, for each combination in the set a probability distribution function (PDF) of P_f is estimated, which are then averaged as per their fitting probabilities. The averaged PDF of P_f considered to have the impact of both the model selection and model parameters uncertainties. An analysis was also performed using the BMA and its results were compared with those from the proposed methodology. While the mean and SD of P_f from both approaches matched well ($\sim$ 5%), the BMMI required 99.52% less computational efforts as compared to the BMA.

It could be observed that the proposed methodology is relatively complex and mathematical compared to the traditional deterministic and reliability methods. Further, the final decision-making in the proposed methodology is more difficult as the outputs are the intervals of the probability of

Table 8 Statistics of the sample statistics of UCS and GSI for four different prior knowledge ranges estimated through BMMI

Prior knowledge (PK)	Statistical parameter	UCS		GSI	
		Mean	SD	Mean	SD
PK-I	Sample mean	115.9484	11.8974	41.1439	1.4561
	Sample SD	44.8957	20.0365	6.6459	1.1739
PK-II	Sample mean	115.6923	11.1870	41.1797	1.4338
	Sample SD	44.7801	20.1381	6.6095	1.1274
PK-III	Sample mean	115.3818	8.4591	41.0545	1.0980
	Sample SD	43.2089	17.9698	6.4086	0.5488
PK-IV	Sample mean	115.5057	5.7374	41.0545	1.0980
	Sample SD	41.4797	16.2779	6.4086	0.5488

Table 9 Statistics of probability of failure (P_f) for four different prior knowledge ranges estimated through BMMI

Prior knowledge (PK)	Mean of P_f	SD of P_f	Confidence interval [2.5–97.5%]
PK-I	0.9046	0.8875	[0.0480–3.1050]
PK-II	0.8941	0.8838	[0.0460–3.1170]
PK-III	0.8225	0.8087	[0.0460–2.7660]
PK-IV	0.7211	0.7554	[0.0500–2.5800]

Table 10 Statistics of the sample statistics of UCS and GSI for three types of prior distributions estimated through BMMI

Prior knowledge (PK)	Statistical parameter	UCS		GSI	
		Mean	SD	Mean	SD
PK-III (U)	Sample mean	115.3818	8.4591	41.0545	1.0980
	Sample SD	43.2089	17.9698	6.4086	0.5488
PK-III (N)	Sample mean	116.0917	6.3687	41.0541	0.8387
	Sample SD	40.8307	16.2188	6.3870	0.3512
PK-III (W)	Sample mean	116.3206	6.2115	41.0853	0.7651
	Sample SD	41.0003	16.2822	6.3954	0.3320

failure (i.e., P_f) in the proposed methodology as compared to the precise values of performance functions (deterministic) and P_f (traditional reliability method). However, the traditional methods often include the originally unavailable information in the analysis, like assigning a precise value to the inputs (generally mean), neglecting all other data (in deterministic analysis) and/or assigning a precise PDF using limited data (in traditional reliability). In contrast, the proposed methodology accepts and considers (with no subjective judgements) the lack of available input data and propagates this imprecision to the outputs by estimating the intervals of P_f. As per Dubois [10], "It is better for engineers to know that you do not know than make a wrong decision because you delusively think you know. It allows one to postpone such a wrong decision in order to start a new measurement campaign, for instance."

Table 11 Statistics of probability of failure (P_f), for three types of prior distributions estimated through BMMI

Prior knowledge (PK)	Mean of P_f	SD of P_f	Confidence interval [2.5–97.5%]
PK-III (U)	0.8225	0.8087	[0.0460–2.7660]
PK-III (N)	0.6612	0.7643	[0.0420–2.5640]
PK-III (W)	0.6621	0.7626	[0.0440–2.5260]

5.2 Impact of prior knowledge on the statistics of properties and response parameter

Informativeness and confidence of prior knowledge is regarded as a key factor in the uncertainty characterization via Bayesian approach [15, 56]. To assess the importance of prior information in BMMI methodology, an analysis was performed to evaluate the effect of prior information on the statistics of response parameter. Effect of prior information was estimated for (1) range of statistics, and (2) prior-distribution of properties. To evaluate the effect of prior range of statistics of properties, ranges were narrowed down by approximately 60%, 90% and 95%, respectively, for prior knowledge II (PK-II), prior knowledge III (PK-III) and prior knowledge IV (PK-IV) cases as compared to original ranges represented by prior knowledge I case (PK-I). Reducing range implies the higher (or more accurate) information. Based on the information levels, these cases could be arranged as: PK-IV > PK-III > PK-II > PK-I. Prior distributions of statistics of properties were assumed to be same (i.e., uniform) for all the cases. Figure 14a shows the corresponding prior distribution for the mean of GSI. Results are summarized in Tables 8 and 9; and Fig. 14b. With the increasing level of information, the uncertainty in the sample statistical parameters was continuously reducing as indicated by decrease in SDs of sample statistical parameters (0.51–53.25%). The uncertainty in P_f was also continuously reducing as indicated by the reducing length of confidence intervals. Minor change

was observed in the confidence interval length for PK-II case (0.46%), however, significant changes were observed for PK-III and PK-IV cases (11.02–17.24%) as compared to PK-I case. In the case of rock mechanics, the level of prior information could be higher for a site if the data are available from the nearby sites. In this scenario, the uncertainty in the estimated P_f would be lower as compared to the case, where the statistics of properties are directly adapted from the literature which is based on the world-wide collected data. Hence, it would be better if the prior information could be collected from the nearby sites to reduce the uncertainties in the prior information and hence, the resulting uncertainties in the response estimates for rock structures.

It should be noticed in the previous sections that the analysis was performed by considering the prior distribution of properties to be uniform. To assess the effect of prior distribution, an analysis was performed by changing the prior distributions of statistics of properties with constant ranges. PK-III case explained in the previous section was considered for the analysis with three types of distributions [i.e., Uniform (U), Normal (N) and Weibull (W)]. In other words, the analysis was considered for three different prior cases, i.e., PK-III (U), PK-III (N) and PK-III (W). Figure 14c shows the distributions for mean of GSI. Tables 10 and 11; and Fig. 14d summaries the analysis results. It was observed that the uncertainty in the sample statistical parameters was lower for Weibull and normal distributions compared to uniform distribution as indicated by decrease in SDs of sample statistical parameters (9.39–39.05%). Similar trend was also observed for the uncertainty in the P_f indicated by the reduced length of confidence interval of P_f for Weibull and normal distributions (7.28–8.75%) as compared to uniform distribution. This could be due to less informative nature of uniform distribution (also known as non-informative prior) as compared to the normal and Weibull distributions [36]. In other words, more confidence is shown by the normal and Weibull distributions to central values as compared to the uniform distribution.

6 Conclusion

This study presented a novel Bayesian multi-model inference (BMMI) approach to characterize the uncertainties associated with both the probability model type and parameters arising due to limited data of properties and to assess their effect on the reliability estimates of rock structures. For this methodology, model type uncertainty was first quantified by employing the multi-model inference approach and then the uncertainties in the model parameters were estimated via traditional Bayesian framework. The proposed framework uses the MCMC method with the MH algorithm to simulate the posterior distributions of model parameters. Response surface methodology and global sensitivity analysis were coupled with this methodology to enhance its robustness and to reduce computational efforts. The proposed methodology was demonstrated for a Himalayan rock slope prone to stress-controlled failure in detail. Further, analyses were also performed using methods, i.e., traditional Bayesian and bootstrap reliability, frequently employed to perform reliability analysis with limited data. The proposed methodology was found to be superior to other methods as it can consider the uncertainties in both model types and parameters and can include the prior information in the analysis. Traditional Bayesian and bootstrap methods underestimated the uncertainties in the statistics of input properties (0.21–51.61% and 0.05–47.05%, respectively) as compared to BMMI due to their inherent issue of neglecting uncertainties in model type and prior information in the analysis respectively. This leads to the underestimation of uncertainty in P_f (49.42–53.06% and 6.35–17.68%, respectively) by these methods as compared to BMMI. Overall, this method overcomes the limitations of traditional methods and can be used for a wide variety of problems with explicit/implicit and single/multiple PFs.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmadabadi M, Poisel R (2016) Probabilistic analysis of rock slopes involving correlated non-normal variables using point estimate methods. Rock Mech Rock Eng 49:909–925. https://doi.org/10.1007/s00603-015-0790-2
2. Aladejare AE, Akeju VO (2020) Design and sensitivity analysis of rock slope using Monte Carlo simulation. Geotech Geol Eng 38:573–585. https://doi.org/10.1007/s10706-019-01048-z
3. Aladejare AE, Idris MA (2020) Performance analysis of empirical models for predicting rock mass deformation modulus using regression and Bayesian methods. J Rock Mech Geotech Eng 12:1263–1271. https://doi.org/10.1016/j.jrmge.2020.03.007
4. Aladejare AE, Wang Y (2017) Sources of uncertainty in site characterization and their impact on geotechnical reliability-based design. ASCE ASME J Risk Uncertain Eng Syst A Civ Eng 3:04017024. https://doi.org/10.1061/ajrua6.0000922
5. Aladejare AE, Wang Y (2017) Evaluation of rock property variability. Georisk 11:22–41. https://doi.org/10.1080/17499518.2016.1207784
6. Aladejare AE, Wang Y (2018) Influence of rock property correlation on reliability analysis of rock slope stability: from property characterization to reliability analysis. Geosci Front 9:1639–1648. https://doi.org/10.1016/j.gsf.2017.10.003

7. Ang AHS, Tang WH (2007) Probability concepts in engineering: emphasis on applications to civil and environmental engineering, 2e instructor site. Wiley, Hoboken
8. Asem P, Gardoni P (2019) Bayesian estimation of the normal and shear stiffness for rock sockets in weak sedimentary rocks. Int J Rock Mech Min Sci 124:104129. https://doi.org/10.1016/j.ijrmms.2019.104129
9. Asem P, Gardoni P (2021) A generalized Bayesian approach for prediction of strength and elastic properties of rock. Eng Geol 289:106187. https://doi.org/10.1016/j.enggeo.2021.106187
10. Bárdossy G, Fodor J (2004) Evaluation of uncertainties and risks in geology: new mathematical approaches for their handling. Springer, Berlin
11. Beck JL, Au S-K (2002) Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. J Eng Mech 128:380–391
12. Bozorgzadeh N, Harrison JP (2019) Reliability-based design in rock engineering: application of Bayesian regression methods to rock strength data. J Rock Mech Geotech Eng 11:612–627. https://doi.org/10.1016/j.jrmge.2019.02.002
13. Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 33:261–304. https://doi.org/10.1177/0049124104268644
14. Cai M, Kaiser PK, Uno H et al (2004) Estimation of rock mass deformation modulus and strength of jointed hard rock masses using the GSI system. Int J Rock Mech Min Sci 41:3–19
15. Cao Z, Wang Y, Li D (2016) Quantification of prior knowledge in geotechnical site characterization. Eng Geol 203:107–116. https://doi.org/10.1016/j.enggeo.2015.08.018
16. Chang X, Wang H, Zhang Y et al (2022) Bayesian prediction of tunnel convergence combining empirical model and relevance vector machine. Measurement (London) 188:110621. https://doi.org/10.1016/j.measurement.2021.110621
17. Contreras LF, Brown ET, Ruest M (2018) Bayesian data analysis to quantify the uncertainty of intact rock strength. J Rock Mech Geotech Eng 10:11–31. https://doi.org/10.1016/j.jrmge.2017.07.008
18. Dawson EM, Roth WH, Drescher A (1999) Slope stability analysis by strength reduction. Geotechnique 49:835–840
19. Duzgun HSB, Bhasin RK (2009) Probabilistic stability evaluation of oppstadhornet rock slope, Norway. Rock Mech Rock Eng 42:729–749. https://doi.org/10.1007/s00603-008-0011-3
20. Düzgün HŞB, Paşamehmetoğlu AG, Yücemen MS (1995) Plane failure analysis of rock slopes: a reliability approach. Int J Surf Min Reclam Environ 9:1–6. https://doi.org/10.1080/09208119508964707
21. Duzgun HSB, Yucemen MS, Karpuz C (2002) A probabilistic model for the assessment of uncertainties in the shear strength of rock discontinuities. Int J Rock Mech Min Sci 39:743–754. https://doi.org/10.1016/S1365-1609(02)00050-3
22. Feng X, Jimenez R (2015) Predicting tunnel squeezing with incomplete data using Bayesian networks. Eng Geol 195:214–224. https://doi.org/10.1016/j.enggeo.2015.06.017
23. Feng X, Jimenez R, Zeng P, Senent S (2019) Prediction of time-dependent tunnel convergences using a Bayesian updating approach. Tunn Undergr Space Technol 94:103118. https://doi.org/10.1016/j.tust.2019.103118
24. Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman and Hall/CRC, Boca Raton
25. Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109
26. Hurvich CM, Tsai C-L (1995) Model selection for extended quasi-likelihood models in small samples. Biometrics 51:1077–1084
27. ISRM (1981) Rock characterization, testing and monitoring. ISRM suggested methods 211
28. Janon A, Klein T, Lagnoux A et al (2014) Asymptotic normality and efficiency of two Sobol index estimators. ESAIM Prob Stat 18:342–364. https://doi.org/10.1051/ps/2013040
29. Johnson RW (2001) An introduction to the bootstrap. Teach Stat 23:49–54
30. Krishnamurthy T (2003) Response surface approximation with augmented and compactly supported radial basis functions. In: 44th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference. https://doi.org/10.2514/6.2003-1748
31. Kumar A, Tiwari G (2022) Application of re-sampling stochastic framework for rock slopes support design with limited investigation data: slope case studies along an Indian highway. Environ Earth Sci 81:1–25
32. Kumar A, Tiwari G (2022) Jackknife based generalized resampling reliability approach for rock slopes and tunnels stability analyses with limited data: theory and applications. J Rock Mech Geotech Eng 14:714–730. https://doi.org/10.1016/j.jrmge.2021.11.003
33. Li DQ, Tang XS, Phoon KK (2015) Bootstrap method for characterizing the effect of uncertainty in shear strength parameters on slope reliability. Reliab Eng Syst Saf 140:99–106. https://doi.org/10.1016/j.ress.2015.03.034
34. Li XY, Zhang L, Jiang SH (2016) Updating performance of high rock slopes by combining incremental time-series monitoring data and three-dimensional numerical analysis. Int J Rock Mech Min Sci 83:252–261. https://doi.org/10.1016/j.ijrmms.2014.09.011
35. Li DQ, Wang L, Cao ZJ, Qi XH (2019) Reliability analysis of unsaturated slope stability considering SWCC model selection and parameter uncertainties. Eng Geol 260:105207. https://doi.org/10.1016/j.enggeo.2019.105207
36. Liu XF, Tang XS, Li DQ (2021) Efficient Bayesian characterization of cohesion and friction angle of soil using parametric bootstrap method. Bull Eng Geol Environ 80:1809–1828. https://doi.org/10.1007/s10064-020-01992-8
37. Lü Q, Xiao ZP, Ji J et al (2017) Moving least squares method for reliability assessment of rock tunnel excavation considering ground-support interaction. Comput Geotech 84:88–100. https://doi.org/10.1016/j.compgeo.2016.11.019
38. Luo Z, Atamturktur S, Juang CH (2013) Bootstrapping for characterizing the effect of uncertainty in sample statistics for braced excavations. J Geotech Geoenviron Eng 139:13–23. https://doi.org/10.1061/(asce)gt.1943-5606.0000734
39. Metropolis N, Rosenbluth AW, Rosenbluth MN et al (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092
40. Montgomery DC (2001) Design and analysis of experiments. Wiley, New York, pp 200–201
41. Moriasi DN, Arnold JG, Van Liew MW et al (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 50:885–900
42. Pandit B, Babu GLS (2018) Reliability-based robust design for reinforcement of jointed rock slope. Georisk 12:152–168. https://doi.org/10.1080/17499518.2017.1407800
43. Pandit B, Tiwari G, Latha GM, Babu GLS (2019) Probabilistic characterization of rock mass from limited laboratory tests and field data: associated reliability analysis and its interpretation. Rock Mech Rock Eng 52:2985–3001. https://doi.org/10.1007/s00603-019-01780-1
44. Ramamurthy T (2010) Engineering in rocks for slopes foundations and tunnels. PHI Learning Pvt. Ltd., New Delhi
45. Robert CP, Casella G (2004) The metropolis—hastings algorithm. In: Robert CP, Casella G (eds) Monte Carlo statistical methods. Springer, Berlin, pp 267–320

46. Rocscience, (2014) Phase2 version 8.020, finite element analysis for excavations and slopes. Rocscience Inc, Toronto
47. Ruffolo RM, Shakoor A (2009) Variability of unconfined compressive strength in relation to number of test samples. Eng Geol 108:16–23
48. Saltelli A, Ratto M, Andres T et al (2008) Global sensitivity analysis: the primer. Wiley, Hoboken
49. Singh K, Xie M (2010) Bootstrap: a statistical method. In: International encyclopedia of education, pp 46–51
50. Tamimi S, Amadei B, Frangopol DM (1989) Monte Carlo simulation of rock slope reliability. Comput Struct 33:1495–1505
51. Tang XS, Li DQ, Cao ZJ, Phoon KK (2017) Impact of sample size on geotechnical probabilistic model identification. Comput Geotech 87:229–240. https://doi.org/10.1016/j.compgeo.2017.02.019
52. Tiwari G, Latha GM (2019) Reliability analysis of jointed rock slope considering uncertainty in peak and residual strength parameters. Bull Eng Geol Environ 78:913–930. https://doi.org/10.1007/s10064-017-1141-1
53. Tiwari G, Latha GM (2020) Stability analysis and design of stabilization measures for Chenab railway bridge rock slopes. Bull Eng Geol Environ 79:603–627. https://doi.org/10.1007/s10064-019-01602-2
54. Tiwari G, Pandit B, Latha GM, Sivakumar Babu GL (2017) Probabilistic analysis of tunnels considering uncertainty in peak and post-peak strength parameters. Tunn Undergr Space Technol 70:375–387. https://doi.org/10.1016/j.tust.2017.09.013
55. USACE (1997) Engineering and design—introduction to probability and reliability methods for use in geotechnical engineering—ETL 1110-2-547. USACE, New York, p 14
56. Wang Y, Akeju OV (2016) Quantifying the cross-correlation between effective cohesion and friction angle of soil from limited site-specific data. Soils Found 56:1055–1070
57. Wang Y, Aladejare AE (2015) Selection of site-specific regression model for characterization of uniaxial compressive strength of rock. Int J Rock Mech Min Sci 75:73–81. https://doi.org/10.1016/j.ijrmms.2015.01.008
58. Wang Y, Aladejare AE (2016) Evaluating variability and uncertainty of geological strength index at a specific site. Rock Mech Rock Eng 49:3559–3573. https://doi.org/10.1007/s00603-016-0957-5
59. Wang Y, Cao Z (2013) Probabilistic characterization of Young's modulus of soil using equivalent samples. Eng Geol 159:106–118. https://doi.org/10.1016/j.enggeo.2013.03.017
60. Wang Q, Fang H (2018) Reliability analysis of tunnels using an adaptive RBF and a first-order reliability method. Comput Geotech 98:144–152
61. Wisz MS, Hijmans RJ, Li J et al (2008) Effects of sample size on the performance of species distribution models. Divers Distrib 14:763–773
62. Wyllie DC, Mah C (2004) Rock slope engineering. CRC Press, Boca Raton
63. Zhang J, Shields MD (2018) On the quantification and efficient propagation of imprecise probabilities resulting from small datasets. Mech Syst Signal Process 98:465–483. https://doi.org/10.1016/j.ymssp.2017.04.042
64. Zhang J, Huang HW, Juang CH, Su WW (2014) Geotechnical reliability analysis with limited data: consideration of model selection uncertainty. Eng Geol 181:27–37. https://doi.org/10.1016/j.enggeo.2014.08.002
65. Zhao H, Chen B, Li S et al (2021) Geoscience frontiers updating the models and uncertainty of mechanical parameters for rock tunnels using Bayesian inference. Geosci Front 12:101198. https://doi.org/10.1016/j.gsf.2021.101198
66. Zhou X, Chen J, Chen Y et al (2017) Bayesian-based probabilistic kinematic analysis of discontinuity-controlled rock slope instabilities. Bull Eng Geol Environ 76:1249–1262. https://doi.org/10.1007/s10064-016-0972-5