Proceedings of the 4rd International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2026.32313

Bayesian Updating for Multi-Source Data Fusion and Its Application in Refining Finite Element Models of Concrete-Faced Rockfill Dam: The Case of the Polihali Dam

Xiaobing Wu

College of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou, China
wuxiaobing_ncwu@outlook.com

Abstract. This study proposes a finite element model correction method for the Polihali concrete-faced rockfill dam based on multi-source data fusion Bayesian updating. This solves some major problems with traditional model fixing, there isn't enough way to measure how certain we are of the parameters, combining lots of different types of data is difficult, and it's slow for figuring out what parameters mean when there are a lot. According to the fusion of four types of monitoring data on 329 based on standardization formation. A Bayesian inversion framework is centered on the internal friction angle and the core parameters (φ) and elastic modulus (E) were constructed. NUTS algorithm is used for MCMC sampling and convergence diagnosis, and the optimal parameters are corrected for the ANSYS finite element model. After Bayesian update, the posterior mean of is 53.86°, with the reduction of uncertainty at 61.8%; the posterior mean of is 119.60MPa, with the reduction of uncertainty at 95.9%: A -0.045 correlation coefficient between the two parameters would avoid a compensate effect. The modified model correctly predicted the stress and strain, the maximum equivalent stress was 1.9263MPa and the maximum deformation amount was 282.47mm, this is in keeping with the mechanical principles. This kind of method overcomes the main problem of traditional model modification very well. It has a strong and healthy probabilistic foundation for creating improved engineering models, performing structural safety evaluations, and optimizing structural monitoring systems. It provides a good reference for similar panel rockfill dams.

Keywords: bayesian updating, multi-source data fusion, finite element model, concrete-faced rockfill dam, uncertainty quantification

1. Introduction

As a type of water conservancy core dam, rockfill dams have nonlinear, compressible and spatial variable rockfill material that easily lead to uneven deformation, which will have a direct influence on its structure safety and long-time economic viability [1]. Finite element numerical simulation has become the main method to study the stress - strain situation of the dams [2]. But traditional methods depend on laboratory tests and experience formulas to decide on things. Due to differences

DOI: 10.54254/2755-2721/2026.32313

in field placement conditions and differences in the space of material placement, there are great discrepancies between the prediction results of the initial model and the monitoring data, which is difficult to apply in risk warnings and precise decision-making [3].

The development of multi-source monitoring technologies has provided a lot of data for improving engineering models [4]. But data type, accuracy, and spatiotemporal scale all have different integration problems [5]. Bayesian updating process adds prior parameter distribution information and information from monitoring data that is a likelihood, and as a result it gets parameter inference and uncertainty. This fundamentally overcomes the main deficiency of the previously mentioned deterministic correction method that is unable to take into account the parameter uncertainties [6]. Other existing studies are Junru Li et al. proposed a Bayesian inversion framework, which is based on single deformation data and does not take advantage of the complementary advantages of multi-source data [7], Dian-Qing Li et al. [8] proposed a multi-source data fusion Bayesian method, but it has low sampling efficiency and incompleteness of uncertainties quantification. As a result, the probabilistic benefit of Bayesian is still not full, the possibility of technology from Bayesian hasn't been unfolded amid the complex background of the rockfill dam [9].

This paper researches the studying finite element model refinement through Bayesian updating via multi-sourced data fusion with polihali rockfill dam as an example. Then establish a refined ANSYS model and conduct conditioning of various available monitoring data collected during the process of building a dam. Next created a Bayesian inversion framework, then performed probabilistic updating of the internal friction angle and elastic modulus of the rockfill material using the NUTS algorithm [10]. Finally,i used the updated inputs to improve the newly improved finite element model and predict it,offering robust,sound and useful probabilistic bases as well as technical routes for enhancing the steel structures and assessing and improving their monitoring.

The innovations of this paper are as follows: The first is the integration of four kinds of different kinds of instruments' data to do some sort of automatic temperature correction, time alignment, and some kind of interpolation. Second, it uses no-U-turn sampler (NUTS) to do efficient MCMC sampling and automatic convergence diagnosis, better than the ordinary metropolis-hastings algorithm [11]. Third, it can do comprehensive uncertainty evaluation and provide posterior distribution analysis, parameter correlation analysis and comparison between prior and posterior to better ensure the reliability of the updating results and achieve accurate and quick Bayesian parameter inversion for multi-source data.

2. Theory of Bayesian update with multi-source data fusion

2.1. Theory of multi-source data fusion in engineering monitoring

The most important thing for multi-source data fusion is to reduce data redundancy and data bias by using consistent and complementary information from different monitoring data and making it into a standard data set for the input of Bayesian inversion. Theoretical basis is the information fusion consistency test and data orthogonality principle. It uses different data sources' credibility and correlation to make different data have the same representation and can be used together [12].

Heterogeneous multi-source monitoring data and environmental disturbances are major problems for fusion. Data fusion theory pays attention to the consistency and synergy standards, and uses statistical tests to filter out abnormal data and guarantee the observation consistency of co-sourced physical quantity. It is integrated complementary information coming from various kinds of data in order to show complete mechanical answers to the dam and environment combined impacts.

DOI: 10.54254/2755-2721/2026.32313

The fused multi-source data for the concrete-faced rockfill dam in this study include the following: settlement, pore water pressure, temperature, and pipe readings. Their interrelations are as follows: temperature affects the accuracy of the pore pressure sensor, backfill elevation affects the size of the overburden load, and settlement and pore pressure affect the rockfill material properties and are used as multi-dimensional constraints for the inversion parameters [3].

In terms of dealing with the data heterogeneity and interference problems, a standardized conditioning workflow was established: "abnormality deletion—systematic error correction—time alignment and interpolation complementation".

First, the outliers are spotted using the 3σ rule, data surpassing or falling below ±3 times the standard deviation from the average is put under reasonableness testing, and data that is invalid is thrown away [13].

Subsequently, large temperature effects on some instrument measurements are caused by the thermal elasticity of the sensor element [14]. Use a linear correction model to remove the influence of temperature:

$$u_{corrected}=u_{measured}+\beta_T(T-T_{ref}) \tag{1}$$

Where $u_{measured}$ is the raw measured pore pressure; T is the current measurement temperature; T_{ref} is the reference temperature, typically the temperature at the first measurement; β_T is the temperature correction coefficient.

To ultimately handle the problem of inconsistent timestamps due to different sampling rates, the study uses the following methods: First, the timestamps were converted into relative time $\Delta t=t-t_0$ in days with the first measurement time being used as the starting point. Second, we combined the timestamps from every instrument into one single continuous timeline. Finally, for the missing points on the unified timeline, they are filled using the following formula for linear interpolation:

$$y(t)=y_i+\frac{y_{i+1}-y_i}{t_{i+1}-t_i}(t-t_i) \tag{2}$$

Where $t_i\leq t\leq t_{i+1}$, y_i and y_{i+1} represents the measurement data at times t_i and t_{i+1}, respectively.

Based on "preserve key content, remove redundancy, and standardize representation", conditioned data are combined into a standardized dataset of valid dataset. Statistical characteristics: The mean describes the center point of the dataset, the standard deviation describes the spread, and the min and max define what is allowed for our data. This kind of standardized database provides rich high-dimensional observable input for the subsequent Bayesian inversion, so as to make the estimation of the parameters precise [15].

2.2. Bayesian inversion theory

2.2.1. Bayes' Theorem and inversion

Bayes' Theorem is the main theoretical foundation for parameter inversion and uncertainty quantification, which can achieve the probabilistic inference of uncertain parameters with prior information and observed data. It is a rigorous theoretical basis for solving the uncertainty problem in engineering [16]. Update previous beliefs on parameters using likelihood information to get a

DOI: 10.54254/2755-2721/2026.32313

posterior probability density function that is closer to the real distribution characteristics of the parameters. It can be expressed in the following form.

$$p(\theta|D)=\frac{p(D|\theta)\cdot p(\theta)}{p(D)} \tag{3}$$

Where $p(\theta|D)$ is the probability distribution of parameter θ under the constraint of the observed data D, representing the posterior distribution; $p(D|\theta)$ quantifies the fit between the model prediction θ and the observed data, representing the likelihood function; $p(\theta)$ is the prior knowledge of parameter from the previous engineering experience and literature, is the prior function; $p(D)$ is the normalization term and does not need to be calculated in MCMC sampling, which represents the probability of data being observed.

Bayesian inversion, this is how we estimate parameters and handle uncertainty; this happens through the Bayes theorem. The basic operating principle of Bayesian inverse problem solving is to consider unknown parameters as random variables and use the combination of the likelihood function and prior distribution to derive the posterior probability distribution of the parameter, thereby transforming the inverse problem into a typical statistical inference problem, using Bayesian update to improve the knowledge about the parameter. It is not only to get the best parameter values, but also it can get the estimation error by using its posterior distribution, therefore it is suitable for solving ill-posed inversion problems.

The multi-source data fusion Bayesian updating technical workflow is shown in Fig. 1. It consists of 7 main steps: data preparation, Bayesian model, MCMC sampling, convergence, analysis of posterior distribution, validation of model and output of results.

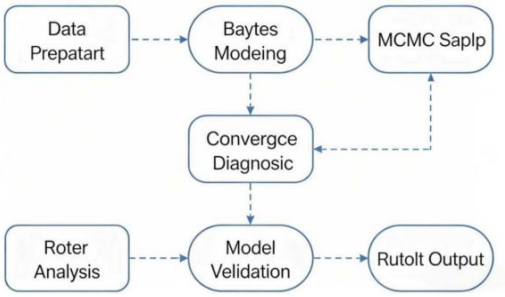

Figure 1. Technical road map of multi-source fusion and Bayesian updating

The pseudo-code for the complete Bayesian inversion algorithm and workflow developed in this paper can be seen in Algorithm 1 (in the Appendix). To improve algorithms on implementation to focus more on numerical stability. For instance, the original likelihood function employed a product form, which could cause numerical underflow when dealing with large datasets ($L(\theta)=\prod_{i=1}^{n} p(D_i|\theta)\to 0$). By adopting a log-likelihood function for computation, the product operation is transformed into a summation ($\log L(\theta)=\sum_{i=1}^{n} \log p(D_i|\theta)$), which effectively prevents numerical underflow.

2.2.2. Forward model design

The forward model serves as the core predictive component of Bayesian updating, providing prior predictive information for parameter inversion. Based on multi-source integrated data derived from

DOI: 10.54254/2755-2721/2026.32313

concrete-faced rockfill monitoring, this study first defines known conditions: settlement monitoring $S_{obs}=\{S_1,S_2,...,S_n\}$; pore water pressure observation $u_{obs}=\{u_1,u_2,...,u_n\}$; boundary conditions including backfill elevation $h(t)$ and temperature $T(t)$. Subsequently, by applying the known conditions and Bayes' theorem, the posterior probability distributions are defined for the rockfill internal friction angle (φ) and elastic modulus (E) $p(\theta|D)$; the parameter vector to be inverted $\theta=[\varphi, E]^T$; and the observation data vector $D=[S_{obs},u_{obs}]^T$, where T denotes transposition.

To ensure the practicality and computational efficiency of the forward model, this study makes the following reasonable assumptions: First, the rockfill material is treated as an isotropic, homogeneous, elastoplastic medium, simplifying the characterization of material mechanical properties. Second, the small deformation theory is adopted to reduce the complexity of solving the mechanical equations. Third, the creep and aging effects of the rockfill are neglected, focusing on the short-term mechanical response patterns during the construction period. Fourth, the evolution of pore water pressure follows Terzaghi's consolidation theory to accurately characterize the variation of pore pressure under seepage-stress coupling [17].

This study employed an elastic-theory-based simplified model to simulate dam settlement deformation patterns, with the deformation equation defined as:

$$S(t)=\frac{(1-\nu^2)\cdot\sigma(t)\cdot H}{E}\cdot\left(1-e^{-\alpha t}\right) \tag{4}$$

Where $S(t)$ is the cumulative settlement at time t, E is the elastic modulus of the rockfill material, ν is the Poisson's ratio of the rockfill material, set to 0.35 based on engineering experience in this study, H is the thickness of the compressible soil layer, α is the consolidation coefficient, related to material permeability, t is time, $\sigma(t)=\rho g h(t)$ is the overburden stress at time t, where ρ is the density of the rockfill material, g is the gravitational acceleration, and $h(t)$ is the fill elevation at time t. The leading term characterizes the elastic compression, whereas the exponential term reflects the time effects in consolidation.

Use a simple pore pressure dissipation model to describe the changes to the pore water pressure and it is given by:

$$u(t)=u_0+\Delta u(t)\cdot e^{-\beta t} \tag{5}$$

Where $u(t)$ is the excess pore water pressure at time t and u_0 is the initial pore water pressure; $\Delta u(t)=\frac{K_0}{1+K_0}\cdot\sigma(t)$ is the increment of excess pore water pressure due to loading; t is time; K_0 is the coefficient of passive earth pressure; β is the pore pressure dissipation coefficient, related to the material's permeability coefficient and consolidation coefficient. The exponential term describes the temporal dissipation pattern of the pore pressure.

Key parameters K_0 , α , and β are calculated using the inversion parameters as follows:

DOI: 10.54254/2755-2721/2026.32313

$$\begin{cases} K_0=1-\sin\varphi \\ \alpha\approx\frac{c_v}{H^2} \\ \beta\approx\frac{k\cdot E}{(1-\nu^2)\gamma_w H} \end{cases} \tag{6}$$

Where φ is the internal friction angle of the rockfill material, c_v is the consolidation coefficient, k is the permeability coefficient, γ_w where is the unit weight of water, and the other parameters retain their previous definitions.

The algorithmic pseudo-code for the forward model is shown in Algorithm 2 (see Appendix). This forward model is constructed based on elasticity theory and Terzaghi's consolidation theory, where φ primarily influences pore water pressure calculations and E primarily affects settlement calculations.

2.2.3. Inversion theory construction

Drawing upon the research of Gibson and McAuley [18], this study first employs a normal distribution with imposed physical constraints. By truncating the normal distribution, the parameter samples were retained within the physically feasible domain.

$$\begin{gathered} \varphi \sim N(\mu_\varphi, \sigma_\varphi^2) = N(50, 2.5^2)[degrees] 30^\circ \leq \varphi \leq 60^\circ \\ E \sim N(\mu_E, \sigma_E^2) = N(100, 10^2)[MPa] 50\ MPa \leq E \leq 200\ MPa \end{gathered} \tag{7}$$

Second, the likelihood function is constructed by assuming a Gaussian distribution of the observational errors. The single-variable likelihood functions for settlement and pore pressure are :

$$\begin{gathered} S_{obs,i}\sim N(S_{sim,i}(\theta),\sigma_S^2) \\ u_{obs,i}\sim N(u_{sim,i}(\theta),\sigma_u^2) \end{gathered} \tag{8}$$

Where $S_{obs,i}$, $u_{obs,i}$ is the observed settlement and pore water pressure at the i -th observation point; $S_{sim,i}(\theta)$, $u_{sim,i}(\theta)$ is the simulated settlement and pore water pressure at the i -th observation point, calculated via the forward model based on parameters θ ; σ_S^2 , σ_u^2 is the variance of the settlement observation error and pore water pressure observation error.

Based on the above single-variable likelihood functions, the joint likelihood function is constructed as follows:

$$p(D|\theta,\sigma_S,\sigma_u)=\prod_{i=1}^{n} N(S_{obs,i}|S_{sim,i},\sigma_S^2)\cdot N(u_{obs,i}|u_{sim,i},\sigma_u^2) \tag{9}$$

To avoid numerical underflow, the log-likelihood function is employed as follows:

$$\log p(D|\theta)=-\frac{n}{2}\log(2\pi)-n\log(\sigma)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{obs,i}-y_{sim,i})^2 \tag{10}$$

DOI: 10.54254/2755-2721/2026.32313

Where the standard deviation of the observation error is σ (σ_S for settlement data and σ_u for pore water pressure data); $y_{obs,i}$ is the observed value at the i -th observation point; $y_{sim,i}$ is the simulated value at the i -th observation point.

The observation errors σ_S and σ_u were also treated as parameters to be estimated, with a half-normal distribution used as their prior distribution. The specific settings were as follows:

$$\begin{aligned} \sigma_S &\sim \text{HalfNormal}(10)\ [\text{mm}] \\ \sigma_u &\sim \text{HalfNormal}(50)\ [\text{kPa}] \end{aligned} \tag{11}$$

2.3. MCMC sampling algorithm

2.3.1. Markov Chain Monte Carlo (MCMC) sampling theory

In the Bayesian inversion, because the posterior is hard to solve with higher dimensions and complex distributions, numerical methods will become the main method. The grid method faces problems with the curse of dimensionality and variational inference loses important posterior info MCMC sampling on the other hand uses the periodicity of Markov chains and the Monte Carlo integration principle. Create a stationary Markov chain to gain the posterior sample without needing to make simplification assumptions over the posterior distribution. This achieves no need for exploration of the global parameter space, has a high computational efficiency in a high-dimensional situation, and has the correct characterization of the true posterior distribution, hence the method of choice for complex engineering Bayesian inversion [19].

2.3.2. NUTS algorithm improvement and application

This study uses the NUTS sampler(see Appendix, algorithm 3: No-U - Turn Sampler (NUTS) which is an adaptive improvement of the Hamiltonian Monte Carlo (HMC) algorithm.

a) Algorithm principle: The key of NUTS is to convert the parameters to the dynamical evolutions of the Hamiltonian systems. It defines the Hamiltonian: $H(\theta,p)=U(\theta)+K(p)$, where the $U(\theta)= -\log p(\theta \mid D)$ is the potential energy function, and $K(p)=\frac{1}{2}p^T M^{-1}p$ is the kinetic energy function (p is the momentum variable, obeying $p\sim N(0,M)$, where M is the mass matrix).

Sampling proceeds by solving the Hamiltonian equations numerically with a leap-frog method:

$$\begin{cases} \frac{d\theta}{dt} = M^{-1}p \\ \frac{dp}{dt} = -\nabla_\theta U(\theta) = \nabla_\theta \log p(\theta|D) \end{cases} \tag{12}$$

This achieves the coordinated evolution of parameters and momentum, avoiding the inefficient sampling caused by random walks in traditional MCMC algorithms.

b) Key improvement features: No-turn-around criterion is presented which automatically truncates trajectory extension by checking the inner product of the parameter and the momentum, it removes the need for manual length setting; utilizes a dual averaging approach: during the warm up phase, it dynamically regulates its integration step size according to the objective acceptance rate (0.9) so as to balance the chance of correct sampling with the sampling rate; it updates matrix M based on the sample covariance during the warm-up period so that after warm-up, the posterior

DOI: 10.54254/2755-2721/2026.32313

distribution's covariance structure is matched, reducing the strength of functional coupling. Compared to classical algorithms, NUTS gets higher acceptance rate(95%), lower sample auto-correlation, fewer manual tuning needs and more robustness.

c) Gradient computation challenge: NUTS algorithm needs to be calculated during the sampling process. In order to efficiently compute the gradient correctly, there are two ways to do so – if the forward model is simple, its gradients can be derived analytically, and if the forward model is difficult, the library uses automatic differentiation to provide gradients efficiently (such as PyMC).

d) Algorithm application: Considering the parameter dimensions and posterior shape characteristics of the problem, NUTS algorithm sample parameter settings are done. First, the number of each chain sample is set to be 2000. Second, set the warmup iterations to 1000. Third, set the number of independent chain to be 2. After this, set the target acceptance rate to 0.9 and then set the initial value of the quality matrix to.

2.4. Quantification of uncertainty

2.4.1. Posterior statistics

Posterior statistic is an essential statistics for calculating the parameter estimation statistic and its related uncertainty. Based on the posterior sample set obtained with MCMC sampling, $\{\theta^{(1)},...,\theta^{(M)}\}$ (where M is the effective sample's size), we calculate the posterior statistic using the following formula:

$$\begin{aligned}\hat{\theta}&=\frac{1}{M}\sum_{i=1}^{M}\theta^{(i)}\\ \text{mode}&=\arg\max_{\theta} p(\theta|D)\end{aligned} \tag{13}$$

Specifically, the posterior mean $\hat{\theta}=\frac{1}{M}\sum_{i=1}^{M}\theta^{(i)}$ reflects the most plausible values of the parameters. It is the main criterion for parameter estimation and is zero-biased and efficient. The Maximum posterior estimate $=\arg\max_{\theta} p(\theta|D)$ corresponds to the peak location of the posterior probability distribution, which contains constraints from both prior information and observed information.

The uncertainty measure is defined as:

$$\begin{aligned}\text{Std}(\theta_j)&=\sqrt{\frac{1}{M\text{-}1}\sum_{i=1}^{M}(\theta_j^{(i)}\text{-}\hat{\theta}_j)^2}\\ \text{HDI}_{95\%}&=[\theta_{2.5\%},\theta_{97.5\%}]\end{aligned} \tag{14}$$

Specifically, the posterior standard deviation $\text{Std}(\theta_j)\&=\sqrt{\frac{1}{M\text{-}1}\sum_{i=1}^{M}(\theta_j^{(i)}\text{-}\hat{\theta}_j)^2}$ is the degree of spread of parameter prediction. The standard deviation is less, the more certain it is. HDI is 95%, it means that the parameters will have a 95% chance of being inside of the HDI. Compared to traditional confidence interval, HDI better reflects the degree of concentrating on probabilities for the posterior distribution in the case of asymmetric distribution.

DOI: 10.54254/2755-2721/2026.32313

2.4.2. Parameter correlation and model accuracy assessment

The parameter correlation analysis is for us to find out the statistic relation of different inversion parameters. It can avoid the unreliable inversion result caused by compensate effect between the inversion parameters. To evaluate the accuracy of the model there has to be the calculation of the Pearson correlation coefficient:

$$\rho(\varphi, E)=\frac{Cov(\varphi,E)}{Std(\varphi)\cdot Std(E)} \tag{15}$$

Where $Cov(\varphi,E)=\frac{1}{M-1}\sum_{i=1}^{M}(\varphi^{(i)}-\widehat{\varphi})(E^{(i)}-\widehat{E})$ represents the covariance. The correlation coefficient ranges from -1 to 1: if $|\rho|\approx 0$: Parameters are approximately independent, and they are not strongly correlated, so we can estimate them independently and avoid "compensation effects."; if $|\rho|\approx 1$: Strong correlation shows that "compensation effect"is present and accurate parameter is estimated related to. It needs to improve by increasing the variety of observed data, or strengthening the prior.

The coefficient of designation for model accuracy assess is defined as R^2 :

$$R^2=1-\frac{\sum_{i=1}^{n}(y_{obs,i}-y_{sim,i})^2}{\sum_{i=1}^{n}(y_{obs,i}-\bar{y}_{obs})^2} \tag{16}$$

Where $y_{obs,i}$ is the i-th observed value; $y_{sim,i}$ is the i-th simulated value; $\bar{y}_{obs}=\frac{1}{n}\sum_{i=1}^{n}y_{obs}$, i is the mean of observed values; n is the number of observations. The range of R^2 is $(-\infty,1]$. The closer R^2 is to 1, the better the model fit. When $R^2<0$, it is also worse to use the simulated model to predict than simply using the average of the observed values. It is usually because of unrealistic assumptions for the model or incorrect parameter estimation.

While it's easier to think of how close the fit is, in reality, it would be impossible to have some measure on how well the fit is since it's the difference in absolute values of simulated and observed data. Therefore, we use RMSE as a supplement measurement metric. The real value of RMSE is to measure the mean difference between the simulated and observed values in the same physical unit as the observation. It is more sensitive to the sample with a big deviation, it can grasp the whole error of the prediction from the model efficiently. And it is very important as evidence so that the model can see if it's enough for real engineering. Its calculation formula is:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{obs,i}-y_{sim,i})^2} \tag{17}$$

Smaller RMSE means that the model fits well, because it is easier to imagine that the predicted values of the model should be closer to the engineering truth. Support for the certification and engineering utility checks of the following model improvements.

2.5. Convergence diagnosis

2.5.1. Gelman-rubin statistic

MCMC sampling's convergence directly relates to the reliability of posterior statistical estimates. If there is no sampling convergence, what is obtained from the sampled value with parameter estimate and uncertainty is biased and the inversion results are invalid. So here, we need to use a bunch of

DOI: 10.54254/2755-2721/2026.32313

different convergence diagnostic indicators and visualizations to check that every part of the sampling is converging [20]. As for calculation formula:

$$\widehat{R}=\sqrt{\frac{\widehat{V}}{W}} \tag{18}$$

Where $W=\frac{1}{m}\sum_{c=1}^{m}s_c^2$ is within-chain variance; m is the number of independent chains; s_c^2 is the sample variance of the c-th chain; $\widehat{V}=\frac{n-1}{n}W+\frac{1}{n}B$ is the total variance estimate; n is the effective sample size per chain; $B=\frac{n}{m-1}\sum_{c=1}^{m}(\overline{\theta_c}-\overline{\theta_{..}})^2$ is the between-chain variance; $\overline{\theta_c}$ is the sample mean of the c-th chain; $\overline{\theta_{..}}$ is the total sample mean across all chains.

Convergence criteria: if $\widehat{R}<1.01$: Multiple chains are sufficiently mixed, indicating good convergence. $1.01<\widehat{R}<1.1$: Convergence is acceptable, but results should be interpreted with caution. $\widehat{R}>1.1$: No convergence; increase the number of samples, adjust sampling parameters, or improve the model structure.

The Gelman-Rubin diagnostic algorithm developed in this study is presented in the appendix (Algorithm 4: Gelman-Rubin Diagnostics).

2.5.2. Effective Sample Size

Due to auto-correlation in MCMC samples, the actual effective information content is lower than the total sample size. Effective Sample Size (ESS) quantifies sample quality by measuring the equivalent number of independent samples, calculated as follows:

$$ESS=\frac{M}{1+2\sum_{k=1}^{\infty}\rho_k} \tag{19}$$

Where M is the total sample size; ρ_k is the lag- k . The auto-correlation coefficient ρ_k is calculated using the following formula:

$$\rho_k=\frac{Cov(\theta^{(i)},\theta^{(i+k)})}{Var(\theta)} \tag{20}$$

Convergence criterion: if ESS>1000: the effective sample size is sufficient to ensure the precision of posterior statistic estimates; 400≤ESS≤1000: the effective sample size is acceptable, but attention should be paid to the estimation precision in the tails of the posterior distribution; $ESS<400$: the effective sample size is insufficient, requiring an increase in the number of samples or optimization of the sampling algorithm (e.g., adjusting parameters such as the quality matrix or stride).

The effective sample size algorithm developed in this study is presented in the appendix (Algorithm 5: Effective Sample Size).

2.5.3. Convergence diagnostic process and results

To ensure comprehensive and reliable convergence diagnostics, the following diagnostic workflow is established: First, plot trajectory plots and auto-correlation function plots to visually assess sampling convergence, statistical stability, and auto-correlation decay characteristics. Well-converged trajectories must satisfy: multiple chain trajectories sufficiently overlapping (convergence), exhibiting random walk behavior around the posterior mean (statistical stability), and

DOI: 10.54254/2755-2721/2026.32313

covering all high-probability regions of the posterior distribution (exploratory coverage). Second, compute the Gelman-Rubin statistic ($\widehat{R}$) and effective sample size (ESS) to quantitatively validate convergence, using the same criteria as previously described. If the criteria are not met, implement measures such as increasing the sample size, extending warm-up iterations, optimizing the acceptance rate, or refining the initial values of the quality matrix. Finally, repeat the above steps until all metrics satisfy the convergence criteria.

3. Modification of the Polihali Concrete-faced Rockfill Dam finite element model

3.1. Project background and finite element model fundamentals

The Polihali Concrete-faced Rockfill Dam is the core project of the Lesotho Highlands Water Project Phase II. And it is located approximately 130km northeast from Maseru, Lesotho and about 20km west from Mokhotlong and is located on the Senqu river, at 28°51'50"E, 29°17'20"S. The dam is mainly made up of a maximum height of 171.3m, a crest length of 921m and a total volume of embankment is 14.98 million m^3. Construction is currently in the embankment filling phase. The project area lies within the Great Karoo Basin, with the dam foundation resting on the slightly horizontal Lesotho Formation, formed by accumulated basaltic lava flows.

A three-dimensional finite element model was established using ANSYS: The main rockfill zone (3B) and secondary rockfill zone (3C) of the dam body were characterized using a coupled Cam-Clay model and Porous Elasticity model, while the mountain soil employed an isotropic elastic model. The element size was set to 1m based on the maximum rockfill particle size, with refined meshing in the contact area between the dam foundation and the mountain. Boundary conditions: dam body - horizontally fixed at foundation base, vertically free; mountain flank - fixed constraint; dam crest and flanks-free. Applied gravitational acceleration of 9.81 m/s^2 and hydrostatic pressure. Simulated layered filling process with 10 load steps. Key initial model parameters are shown in TABLE 1 and TABLE 2.

3.2. Bayesian update for multi-source data fusion

3.2.1. Multi-source data conditioning: and fusion data structure

Four types of monitoring instruments were embedded in the main body of the Polihali Panel Rockfill Dam. First, physical quantities directly influencing key material parameters were selected for fusion. After outlier data removal, the core instrument information is shown in TABLE 3.

Table 1. Rockfill area material parameters

No.	3B Rockfill Area aterial		
	Property	Value	Unit
1	Density	3895	kg/m^3
2	Young's Modulus	3.4007×10^{10}	Pa
3	Poisson's Ratio	0.23391	
4	Bulk Modulus	2.13×10^{10}	Pa
5	Shear Modulus	1.378×10^{10}	Pa
6	Initial Internal Friction Angle	0.010472	radian
7	Initial Cohesion	4.2×10^{7}	Pa

DOI: 10.54254/2755-2721/2026.32313

Table 1. (continued)

8	Dilatation Angle	0.010472	radian
9	Residual Inner Friction Angle	0.005236	radian
10	Residual Cohesion	2×107	Pa
	3C Rockfill Area Material		
1	Density	4000	kg/m³
2	Young's Modulus	5.2819×10^{10}	Pa
3	Poisson's Ratio	0.1238	
4	Bulk Modulus	2.34×10^{10}	Pa
5	Shear Modulus	2.35×10^{10}	Pa
6	Initial Inner Friction Angle	0.010472	radian
7	Initial Cohesion	4.2×10^{7}	Pa
8	Dilatation Angle	0.010472	radian
9	Residual Inner Friction Angle	0.005236	radian
10	Residual Cohesion	2×10^{7}	Pa

Table 2. Main dam foundation soil parameters

No.	Foundation Soil Material		
	Property	Value	Unit
1	Density	2300	kg/m³
2	Thermal Expansion Coefficient	1.4×10^{-05}	C^{-1}
3	Young's Modulus	3×10^{10}	Pa
4	Poisson's Ratio	0.18	
5	Bulk Modulus	1.5625×10^{10}	Pa
6	Shear Modulus	1.2712×10^{10}	Pa
7	Tensile Yield Strength	0	Pa
8	Compressive Yield Strength	0	Pa
9	Ultimate Tensile Strength	5×10^{6}	Pa

Table 3. Instrument codes and data description

Full Name	Instrument codes and data description			
	Instrument code	Physical quantity	Sampling frequency	Data volume
Settlement Gauges	SSG	Settlement (mm)	Irregular	134
Vibrating Wire Pressure	VWP	Borehole Pressure (kPa) and Temperature (°C)	2-4 hours	131
Hydraulic Settlement Cells	HSC	Pipe reading (mm)	Irregular	23
Foundation Piezometers	FP	Pore Pressure (kPa) and Temperature (°C)	Irregular	43

DOI: 10.54254/2755-2721/2026.32313

Second, since pore pressure sensors are susceptible to temperature drift, this study needs temperature correction. Based on engineering practice experience, the β_T for the VWP sensor in this study is −1.3 kPa/°C, and the β_T for the FP sensor is −0.363 kPa/°C.

Third, addressing the variations in type, accuracy, and spatiotemporal scales among multi-source data, multi-source monitoring data fusion analysis was conducted in a Python3.10 environment: Firstly, a fill elevation-temperature time series diagram (Fig. 2) was plotted to visually represent the evolution trends of construction progress and environmental temperature. Secondly, a settlement monitoring time series diagram (Fig. 3) was generated based on SSG monitoring data to reflect the temporal evolution pattern of settlement. Thirdly, Fig. 4 shows the pipe reading time series based on HSC monitoring data to assist in settlement verification. Finally, Fig. 5 compares pore water pressure from multiple instruments by overlaying the VWP pore water pressure time series, the temperature-corrected FP pore water pressure time series, and the average pore pressure curve, conducting multi-source data consistency testing to ensure data reliability.

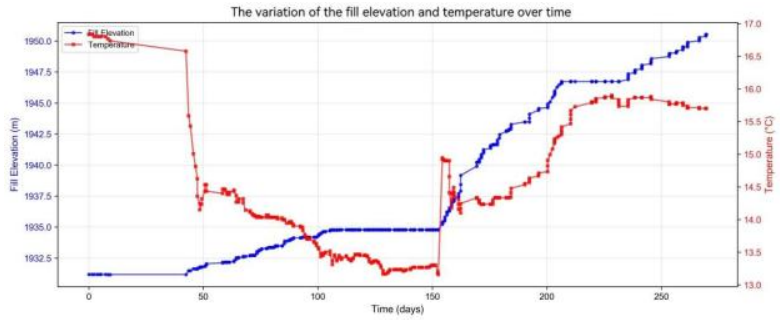

Figure 2. Variation of fill elevation and temperature over time

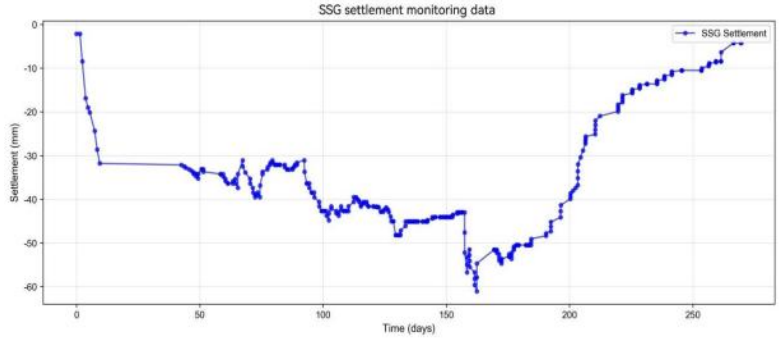

Figure 3. Settlement monitoring data over time

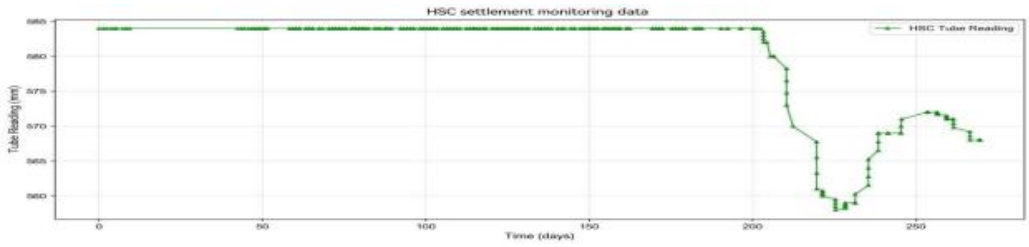

Figure 4. Hydraulic settlement cell tube reading-time monitoring data plot

DOI: 10.54254/2755-2721/2026.32313

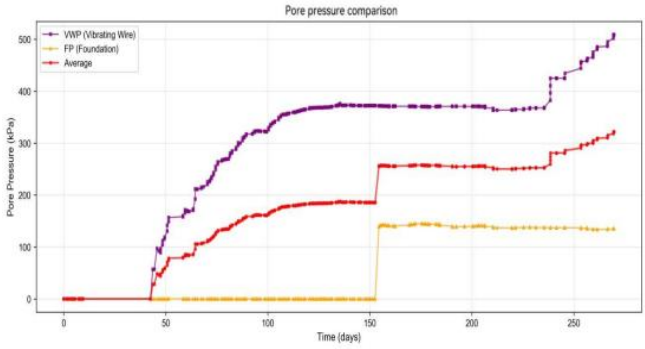

Figure 5. Pore water pressure comparison from multiple instruments

The reliability of the fused dataset was validated through consistency checks: Fig. 5 and Fig. 6 show a 0.82 fit between SSG settlement data and HSC tube readings, confirming the orthogonality and credibility of settlement data. Fig. 7 indicates a Pearson correlation coefficient of 0.87 between VWP and FP, demonstrating good consistency among pore pressure data from different sensors.

Through the aforementioned data conditioning and fusion workflow, this study ultimately obtained 329 valid records of merged data. TABLE 4 lists the statistical characteristics of each variable in the merged dataset.

Table 4. The merged data processing

Variable Name	The Merged Data Processing					
	Unit	Number	Mean	Standard deviation	Maximum	Minimum
Time days	day	329	144.7	64.7	0.0	269.7
Temperature average	°C	329	14.4	1.0	13.2	16.8
Fill elevation	m	329	1938.4	5.9	1931.2	1950.5
Settlement SSG	mm	329	-35.7	13.8	-61.0	-2.1
Pore pressure average	kPa	329	197.3	73.30	0.0	322.2

TABLE 4 indicates that the mean settlement value is -35.7 mm with a standard deviation of 13.8 mm, reflecting the overall pattern and fluctuation range of dam body settlement during the construction period. The mean pore water pressure is 197.3 kPa with a standard deviation of 73.3 kPa, consistent with the increasing trend of backfill elevation and conforming to the pore pressure evolution pattern under Terzaghi's consolidation theory.

Based on the above integrated data, this study developed a multi-source data fusion algorithm. (refer to Appendix 6 for the pseudo-code)

3.2.2. Bayesian update application

The processed multi-source fusion data underwent MCMC sampling using the NUTS algorithm. Following the aforementioned configuration, 2000 valid posterior samples were ultimately obtained after the pruning warm-up phase, ensuring the accuracy and reliability of posterior statistical estimates. The MCMC sampling pseudo-code using the NUTS algorithm is provided in the appendix. Practical test shows that the NUTS sampling process of this project will take very short time, compared with the traditional Metropolis-Hastings algorithm, the efficiency improvement is several magnitude, it can meet the rapid parameter inversion need in engineering.

DOI: 10.54254/2755-2721/2026.32313

Also did the convergence diagnosis. Based on the above diagnostic process, the MCMC convergence results obtained by the user are shown in the Figures 6,7, which were created automatically by the Arvi Z library and serve as core charts for judging the quality of MCMC sampling.

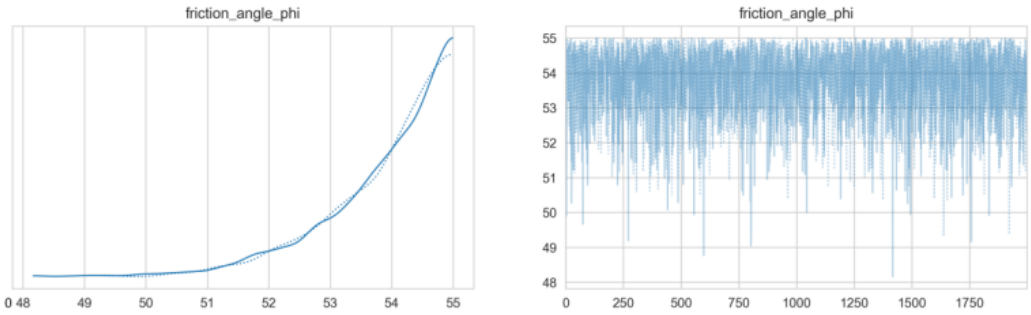

Figure 6. MCMC trajectory and posterior distribution of internal friction angle (φ)

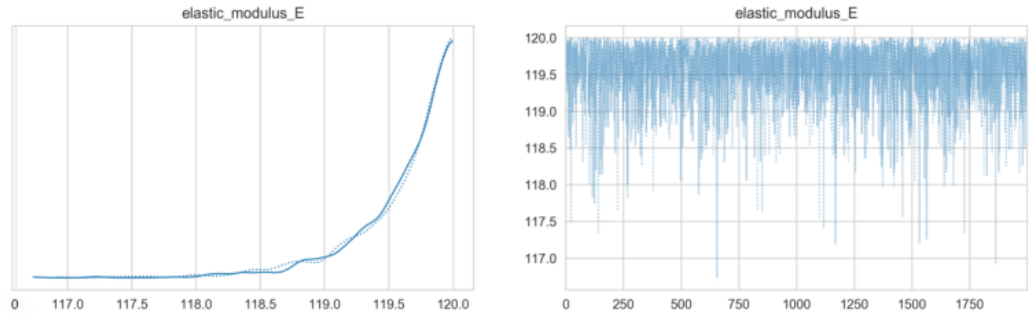

Figure 7. MCMC trajectory and posterior distribution of elastic modulus (E)

The first figures in Fig. 6 and Fig. 7 respectively show the MCMC sampling trajectory plots of φ and E: The multi-chain trajectories overlap and take a random walk around the posterior mean that contains all high-probability parts of the posterior. This meets the criteria for mixing, statistical stability and periodicity, meaning it has the best sampling results. Fig. 6 and Fig. 7, specifically the second panels, display the posterior distributions for φ and E, respectively. These are generated following convergence diagnostics conducted using Algorithm 4: Gelman - Rubin diagnostic algorithm, and Algorithm 5: Effective sample size algorithm (pseudo-code in Appendix).

In this paper, the plot of trajectory for parameters shows very good mixing and stat stability. The convergence diagnostics show that the GELMAN-RUBIN statistics for internal friction angle and elastic modulus E are, and their effective sample sizes ESS(φ)=2476 and ESS(E)=2852, and the auto-correlation coefficient decays quickly to zero. All metrics meet the convergence requirements, which means that this MCMC sampling process has achieved complete convergence. Sampling is of trust and meets requirements for use in future posterior statistics and quantification of uncertainties.

3.3. Inversion results

Based on the Bayesian-Update-method for fusion of multi-source data, the key- mechanical-parameters of rockfill-material were inverted: Present the posterior distribution histograms of the internal friction angle (Fig. 8) and the elastic modulus (Fig. 9). The red dashed lines in each figure indicate the posterior mean and the red shaded areas are the 95% confidence intervals. The

DOI: 10.54254/2755-2721/2026.32313

probability density peaks indicate the most probable parameter value, and the width is characteristic of how uncertain we are about the parameter values.

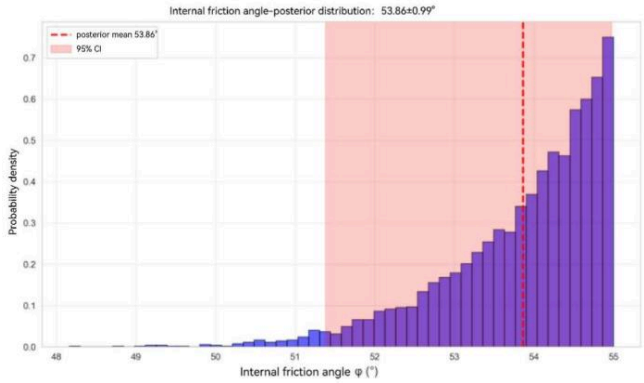

Figure 8. Posterior distribution of the internal friction angle (φ)

The probability density distribution of the internal friction angle (φ) is illustrated in Fig. 8, it displays a posterior mean of 53.86°, a standard deviation of 0.99°, and a 95% confidence interval. In comparison to the previous mean of50.00 °, uncertainty was cut short 61.8% due to Bayesian updating.

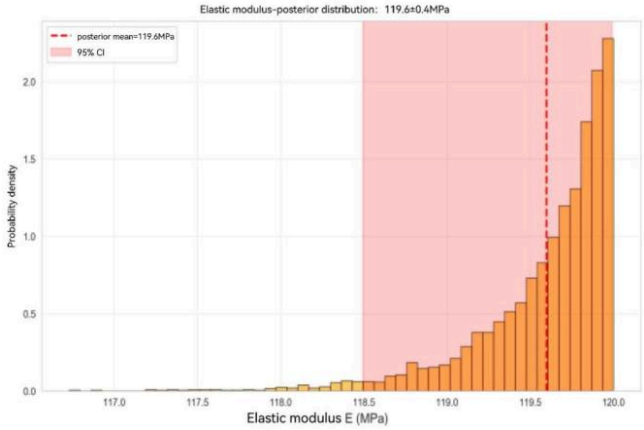

Figure 9. Posterior distribution of elastic modulus (E)

The probability density distribution of the elastic modulus E is shown in Fig. 8, indicating a posterior mean of 119.60 MPa, a standard deviation of 0.41 MPa, and a 95% confidence interval. Compared to the prior mean of 100.00 MPa, the parameter uncertainty is reduced by 95.9%.

The above posterior distribution histograms fully validate the significant advantages of the multi-source monitoring data fusion Bayesian updating method in enhancing parameter estimation accuracy and reducing parameter uncertainty.

To diagnose the parameter correlation between φ and E and identify multi-modality, a joint parameter distribution was established as shown in Fig. 10. This figure employs a hexagonal density plot to display their joint posterior distribution. The green asterisk represents the posterior mean (P.M), while color intensity indicates sample density—deeper red signifies greater sample concentration.

DOI: 10.54254/2755-2721/2026.32313

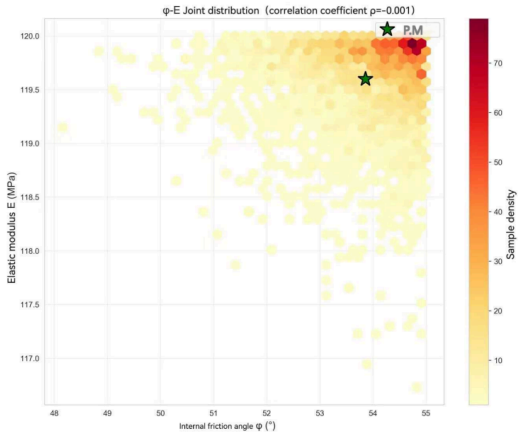

Figure 10. Joint distribution: internal friction angle and elastic modulus

Statistical results see (Fig. 10) that we have a -0. 045 ρ correlation with and. This value tends to zero which means these parameters hardly have any connection. This effectively avoids the 'g-compensation effect' between parameters, ensuring the uniqueness and reliability of parameter estimation.

Model validation based on the Bayesian posterior distribution yielded the scatter plot of observed versus simulated values in Fig. 11.

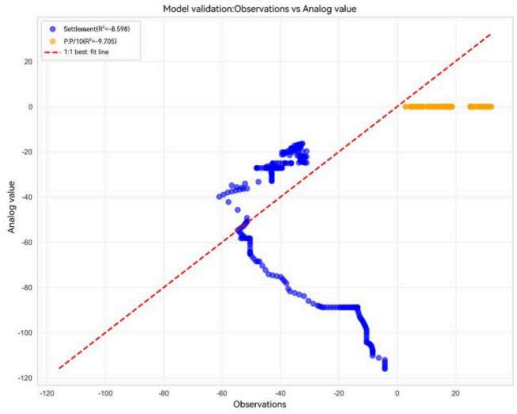

Figure 11. Model validation

The settlement model exhibits a coefficient of determination R^2=−8.598 and a root mean square error RMSE=213.69 kPa. The negative R^2 value stems from the simplified forward model's insufficient consideration of the complex constitutive properties of rockfill material. However, the RMSE indicates controllable systematic bias, validating the reliability of the Bayesian parameter inference model. The updated material parameters for rockfill zones 3B and 3C are presented in TABLE 5.

Table 5. Updated material parameters for rockfill areas 3B and 3C

No.	3B Rockfill Area Material		
	Property	Value	Unit
1	Density	3895	kg/m³
2	Young's Modulus	120.01	MPa

DOI: 10.54254/2755-2721/2026.32313

Table 5. (continued)

3	Poisson's Ratio	0.3	
4	Bulk Modulus	1.0001×10^{8}	Pa
5	Shear Modulus	4.6158×10^{7}	Pa
6	Initial Internal Friction Angle	54.85	degree
7	Initial Cohesion	4.2×10^{7}	Pa
8	Dilatation Angle	0.010472	radian
9	Residual Inner Friction Angle	0.005236	radian
10	Residual Cohesion	2×10^{7}	Pa
	3C Rockfill Area Material		
1	Density	4000	kg/m^3
2	Young's Modulus	119.19	MPa
3	Poisson's Ratio	0.25	
4	Bulk Modulus	7.946×10^{7}	Pa
5	Shear Modulus	4.7676×10^{7}	Pa
6	Initial Inner Friction Angle	52.87	degree
7	Initial Cohesion	4.2×10^{7}	Pa
8	Dilatation Angle	0.010472	radian
9	Residual Inner Friction Angle	0.005236	radian
10	Residual Cohesion	2×10^{7}	Pa

Based on the updated parameters, the ANSYS finite element model was rerun, yielding the isostatic stress and deformation distribution contour plots for the dam body as shown in Fig. 12 and Fig. 13.

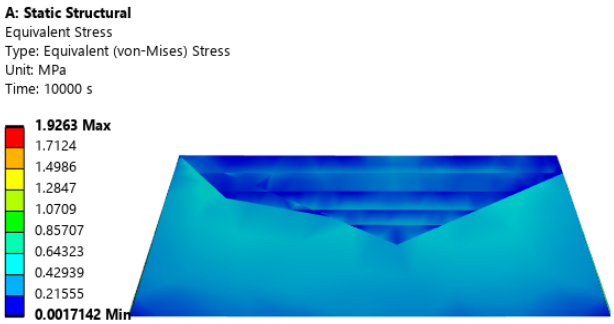

Figure 12. Dam Von Mises stress contour map

As shown in Fig. 12, the dam's Von Mises stress, measured in megapascals (MPa), characterizes the equivalent shear stress state of the structure under external loading. The maximum equivalent stress of 1.9263 MPa is concentrated in the central region, gradually decreasing toward the sides to a minimum of 0.0017142 MPa. This value is significantly below the yield strength of the rockfill material, indicating the dam operates entirely within the elastic range. The stress distribution meets the mechanical rule that stress concentrations happen where deformation concentrations do. This is a visual example of internal forces being transmitted internally through the structure and showing how the load is being transferred by the center of the dam to the outside restraint zones. This kind of

DOI: 10.54254/2755-2721/2026.32313

stress distribution has a main basis for evaluating the structural strength reserve of the dam and the local stress risk.

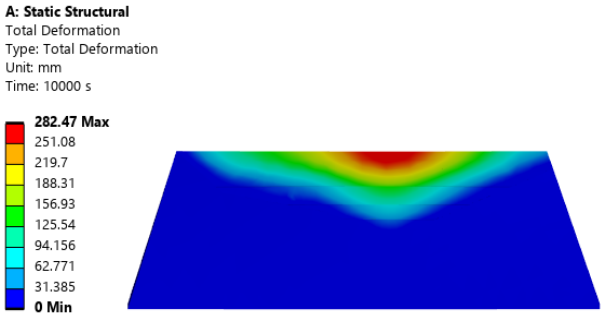

Figure 13. Deformation distribution contour map of the dam body

As shown in fig.13 total dam deformation is in mm, it represents the overall displacement response of the structure subjected to external load. The deformation shows symmetrical distribution and its maximum displacement 282.47mm which is at the middle of the dam, and it gradually goes down to zero at the both end of the dam. It also meets the rigid connection boundary condition between the dam and the slope as well as the symmetric external load configuration, which is consistent with beam bending deformation theory. The displacement field results are visually consistent with the dam's stiffer response to the targeted load, providing an academic baseline for assessing the strength reserve as well as analyzing deformity resistance.

4. Conclusions

And this study overcomes some of the most critical technical difficulties in converting conventional finite element models by developing a multi-source data fusion Bayesian update model for model improvement, taking the Polihali concrete-faced rockfill dam as an example. And the conclusions are as follows:

The first is to establish a multi-source heterogeneous monitoring data fusion system. After the correction of the temperature, the time was realigned and the linear interpolation was used to integrate four types of instrument into 329 standardized fused data points. It did a good job on solving traditional problems about data redundancy, bias, and spatial-temporal mismatch, and gave useful data backing for Bayesian updating.

Second, we built a Bayesian inversion framework incorporating multi-source data. Use NUTS algorithm to obtain the probabilistic estimation and uncertainty determination of internal friction angle (φ) and elastic modulus (E) inside the rockfill. Post-inversion the posterior uncertainties of and E decreased by 61.8%, 95.9% respectively. The correlation coefficient ρ= -0.045 avoided the "compensation effect", which was the most accurate core parameter for the model.

And using the inverted optimal parameters to correct the ANSYS finite element model, the corrected model can predict the correct stress and deformation, in line with engineering reality. The solution is the dam where the maximum equivalent stress is 1.9263Mpa and the maximum deformation is 282.47mm. It is operated in the elastic area. Deformation distribution conforms to mechanical principle, it is safely evaluated numerically for a dam.

Finally the proposed multi-source data fusion Bayesian updating approach has made a leap from point estimation to probability distribution estimation, and does an effective job with respect to multi-source data fusion, high dimensional parameter inversion and also the uncertainty evaluation.

DOI: 10.54254/2755-2721/2026.32313

The way being strong, being friendly makes some sensors do throw things geloades make sense. It's readily transferable to other similar Earth-rock Dams and Geotech. structures, showing good application potential.

Future works will put stress on adding more complex constitutive models to create more precise forward models and adding in machine learning algorisms for better data fusion, all to improve precision and fusion speed.

References

[1] E. Wu, P. Wei, and J. Zhu, et al., "State-dependent model incorporating particle breakage for rock-fill material established by fractional calculus, " International Journal of Geomechanics, vol. 26, no. 2, 2026, pp. 04025344.
[2] E. Erayman, U. S. Cavus, and M. Yildiz, "Post-construction safety assessment of a high concrete face rockfill dam, " Proceedings of the Institution of Civil Engineers–Geotechnical Engineering, vol. 177, no. 2, 2022, pp. 119–134.
[3] Z. Ai, G. Ma, and G. Zhang, et al., "Multi-source monitoring data filtering assisted deformation analysis model updating of ultra-high rockfill dam, " Computers and Geotechnics, vol. 171, 2024, pp. 106323.
[4] Y. Xu, M. Yuan, X. Chen, Y. Shi, J. Liu, W. Yang, and Y. Li, "Detection technology of concealed structures in hydraulic engineering based on the fusion of non-destructive multi-source heterogeneous data, " CT Theory and Applications, vol. 28, no. 2, 2019, pp. 187-194.
[5] S. Luo, D. Yuan, and B. Wei, et al., "Dam multi-source heterogeneous monitoring data fusion and synchronization method based on time series analysis, " Engineering Structures, vol. 338, 2025, pp. 120623.
[6] G. Sevieri, A. De Falco, and M. Andreini, et al., "Hierarchical Bayesian framework for uncertainty reduction in the seismic fragility analysis of concrete gravity dams, " Engineering Structures, vol. 246, 2021, pp. 113001.
[7] J. Li, Z. Wu, and J. Chen, et al., "FEM-Bayesian Kriging method for deformation field estimation of earth dams with limited monitoring data, " Computers and Geotechnics, vol. 148, 2022, pp. 104782.
[8] D. Q. Li, H. H. Zang, and X. S. Tang, et al., "Efficient Bayesian updating for deformation prediction of high rock slopes induced by excavation with monitoring data, " Engineering Geology, vol. 342, 2024, pp. 107772.
[9] Z. Li, E. Y. Khailah, and X. Liu, et al., "Exploring purpose-driven methods and a multifaceted approach in dam health monitoring data utilization, " Buildings, vol. 15, no. 15, 2025, pp. 2803.
[10] H. Zhao, S. Li, and X. Zang, et al., "Uncertainty quantification of inverse analysis for geomaterials using probabilistic programming, " Journal of Rock Mechanics and Geotechnical Engineering, vol. 16, no. 3, 2024, pp. 895-908.
[11] A. Mootoovaloo, J. Ruiz-Zapatero, and C. García-García, et al., "Assessment of gradient-based samplers in standard cosmological likelihoods, " Monthly Notices of the Royal Astronomical Society, vol. 534, no. 3, 2024, pp. 1668-1681.
[12] X. Xi, Y. Nie, and Y. Zhou, et al., "An information fusion approach based on weight correction and evidence theory, " Journal of Computational Science, vol. 83, 2024, pp. 102456.
[13] Y. Mao, J. Li, and Z. Qi, et al., "Research on outlier detection methods for dam monitoring data based on post-data classification, " Buildings, vol. 14, no. 9, 2024, pp. 2758.
[14] M. Liu, Z. Wang, and P. Jiang, et al., "Temperature compensation method for piezoresistive pressure sensors based on gated recurrent unit, " Sensors, vol. 24, no. 16, 2024, pp. 5394.
[15] A. Amavasai, H. Tahershamsi, and T. Wood, et al., "Data assimilation for Bayesian updating of predicted embankment response using monitoring data, " Computers and Geotechnics, vol. 165, 2024, pp. 105936.
[16] X. Zhao and A. Curtis, "Variational prior replacement in Bayesian inference and inversion, " Geophysical Journal International, vol. 239, no. 2, 2024, pp. 1236-1256.
[17] C. A. Rodríguez, Á. M. Rodríguez-Pérez, and R. López, et al., "A finite element method integrated with Terzaghi's principle to estimate settlement of a building due to tunnel construction, " Buildings, vol. 13, no. 5, 2023, pp. 1343.
[18] L. A. Gibson and K. B. McAuley, "Bayesian parameter estimation using truncated normal distributions as priors for parameters in fundamental models of chemical processes, " The Canadian Journal of Chemical Engineering, vol. 103, no. 2, 2025, pp. 649-665.
[19] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P. C. Bürkner, "Rank-normalization, folding, and localization: An improved R^ for assessing convergence of MCMC, " Bayesian Analysis, vol. 16, no. 2, 2021, pp. 667-718.
[20] L. Liu, H. Chen, and S. Wang, et al., "A comparative study of single-chain and multi-chain MCMC algorithms for Bayesian model updating-based structural damage detection, " Applied Sciences, vol. 14, no. 18, 2024, pp. 8514.

DOI: 10.54254/2755-2721/2026.32313

Appendix

Algorithm 1: Complete Bayesian Inversion Framework

Input: Multi-source monitoring data (raw Excel files)
Prior knowledge (parameter ranges, engineering experience)
Output: Posterior distributions p(θ|D), parameter estimates, diagnostics
PHASE 1: DATA PREPARATION
1: Load raw data from {SSG, VWP, HSC, FP}
2: Apply temperature correction (Algorithm 1)
3: Time alignment and interpolation
4: Merge multi-source data → D_merged
PHASE 2: BAYESIAN MODELING
5: Define prior distributions: p(φ), p(E), p(σ)
6: Define forward model: f(θ) → {S_sim, u_sim} (Algorithm 2)
7: Construct likelihood: $p(D|\theta) = \prod N(y\_obs \mid y\_sim, \sigma^2)$
8: Compute posterior: $p(\theta|D) \propto p(D|\theta) \times p(\theta)$
PHASE 3: MCMC SAMPLING
9: Initialize NUTS sampler with adaptive step size
10: Run sampling: Θ ← NUTS(p(θ|D), N=2000, chains=2) (Algorithm 3)
11: Discard burn-in samples (first 1000 iterations)
PHASE 4: CONVERGENCE CHECK
12: Compute Gelman-Rubin $\widehat{R}$ for all parameters (Algorithm 4)
13: Compute ESS for all parameters (Algorithm 5)
14: If $\widehat{R} > 1.01$ or ESS < 1000: increase sampling, goto step 10

Algorithm 2: Forward Model (Settlement & Pore Pressure)

Input: θ = [φ, E], fill elevation h(t), time series t
Output: S_sim (settlement), u_sim (pore pressure)
1: Extract parameters: φ (friction angle), E (elastic modulus)
2: Compute K_0 = 1 - sin(φ) // Lateral earth pressure coefficient
3: for each time step t_i do
4: σ[i] = ρ × g × h[i] // Overburden stress
5: $S[i] = -[(1-\nu^2) \times \sigma[i] \times H / E] \times (1 - \exp(-\alpha \times t\_i))$
6: $u[i] = u\_0 + [K\_0/(1+K\_0)] \times \sigma[i] \times \exp(-\beta \times t\_i)$
7: end for
8: return S_sim, u_sim
Note: α, β are consolidation/dissipation coefficients
Complexity: O(n), where n = number of time steps

Algorithm 3: No-U-Turn Sampler (NUTS)

Input: D = {S_obs, u_obs} (observed data)
prior(θ) (prior distributions)
N_samples, N_tune, N_chains
Output: Θ = {θ^(1), ..., θ^(N)} (posterior samples)
1: Initialize: θ^(0) ~ prior(), step size ε, mass matrix M
2: // Tuning phase (adapt ε and M)
3: for iter = 1 to N_tune do
4: Sample momentum: p ~ N(0, M)

DOI: 10.54254/2755-2721/2026.32313

5: Build Hamiltonian trajectory using leapfrog integration
6: Adapt step size ε (target acceptance rate = 0.9)
7: Estimate mass matrix M from sample covariance
8: end for
9: // Sampling phase
10: for iter = N_tune+1 to N_tune+N_samples do
11:Sample momentum: p ~ N(0, M)
12:Compute: log p(θ|D) = log prior(θ) + log likelihood(D|θ)
13:Build trajectory until U-turn detected
14:Sample candidate θ' from trajectory
15:Accept/reject via Metropolis-Hastings
16: end for
17: return posterior samples Θ
Key components:
- Leapfrog integrator: energy-preserving trajectory construction
- No-U-Turn criterion: automatic path length selection
- Dual averaging: adaptive step size tuning
Complexity: O(N × L × C), L = trajectory length, C = model cost
Algorithm 4: Gelman-Rubin Convergence Diagnostic
Input: m chains, each with n samples {θ_c^(i)}
Output: $\hat{R}$ (potential scale reduction factor)
1: Compute within-chain variance: W = (1/m) Σ_c s_c^2
2: Compute between-chain variance: B = (n/(m-1)) Σ_c $(\bar{\theta}_c - \bar{\theta}_{..})^2$
3: Estimate total variance: $_i$ = [(n-1)/n]W + (1/n)B
4: Compute $\hat{R} = \sqrt{(\hat{V} / W)}$
5: If $\hat{R} < 1.01$, convergence is achieved
Interpretation: $\hat{R} \approx 1$ indicates chains have converged
Algorithm 5: Effective Sample Size (ESS)
Input: MCMC samples Θ = {θ^(1), ..., θ^(N)}
Output: ESS (effective sample size)
1: Compute sample variance: σ^2
2: Compute autocorrelation function: ρ[k] for k = 1,
3: Compute autocorrelation time: τ = 1 + 2Σ ρ[k]
4: Compute ESS = N / τ
5: If ESS > 1000, then samples are sufficient
Interpretation: Accounts for correlation between samples
PHASE 5: POSTERIOR ANALYSIS
15: Compute posterior statistics: mean, standard deviation, 95% HDI
16: Analyze parameter correlations
17: Quantify uncertainty reduction
PHASE 6: MODEL VALIDATION
18: Predict using posterior mean: $\hat{y} = f(\bar{\theta})$
19: Compute metrics: R^2, RMSE, residual plots
PHASE 7: OUTPUT GENERATION
20: Generate visualizations
21: Save posterior samples and reports

DOI: 10.54254/2755-2721/2026.32313

22: Export parameters for FEM software (ANSYS)
Algorithm 6: Multisource Data Fusion
Input: Raw data {SSG, VWP, HSC, FP}, temperature coefficients β
Output: Unified data matrix D_merged
1: Load and preprocess all instrument data
2: Apply temperature correction:
$u_{corrected} = u_{measured} + \beta \times (T - T_{ref})$
3: Merge timestamps from all sources → T_union
4: Interpolate missing values (linear interpolation)
5: Fuse multi-source data by averaging
6: Return D_merged
Complexity: O(n log n), where n = number of data points