Revision letter: response to Reviewer #2

egusphere-2025-4531
Bayesian data selection to quantify the value of data for landslide runout calibration

January 26, 2026

We would like to thank Reviewer #2 for their valuable comments and suggestions, which improved the quality of the manuscript substantially. We present herein our responses to the comments of Reviewer #2.

Italicized line, figure, section numbers refer **to the revised manuscript**.

The changes in the revised manuscript with respect to the original manuscript can be traced in the `diff.pdf` file. Note that the line numbers in this file deviate from the pure manuscript PDF due to the tracked changes displayed.

Legend

> Shows original reviewer comment

Response Contains direct response and explanations.

Resulting changes in manuscript Summarizes resulting changes in the paper.

> Contains a direct quote of new or modified manuscript content with significant changes with respect to the original manuscript.

1 Comments of Reviewer #2

1.1 Comment 1

> Please define "calibration" as it has different definitions. For example, see https://en.wikipedia.org/wiki/Calibration_(statistics) and https://doi.org/10.1214/23-BA1404

Response

Thank you for this comment. We use the term "calibration" to refer to the process of inferring model parameters from observational data within a Bayesian framework. We have now added an explicit definition in Section 1 (Introduction) to avoid ambiguity.

Resulting changes in manuscript

To address the reviewer's feedback we have added the definition of calibration (*p.2 lines 33-34*).

1.2 Comment 2

> I'm also concerned with use of KL-divergence. I don't think author's rebuttal on this is sufficient. As the authors have used only uniform prior, the measure used is equivalent to entropy. Entropy is sensible measure for sharpness of the distribution. Instead of talking about KL-divergence, I suggest the authors would talk about entropy and use entropy also in case of non-uniform priors. Using entropy would focus on maximizing the sharpness of the posterior, while using KL can lead also maximal shift of the posterior.

Response

Thank you for bringing this to our attention. We agree that in the context of this study, where we use uniform priors, KL divergence between the posterior and the prior is equivalent to the entropy of the posterior. Accordingly, we have modified the method subsection 2.3 (now titled "Data selection using information-theoretic metrics") to mathematically demonstrate this equivalence and explicitly acknowledge it in the context of our work.
Our primary objective in this work is to quantify the information gained during calibration with a given observation. Based on our literature review, KL divergence between the posterior and prior is commonly used to measure the information gain during calibration[Haeusel et al., 2026, Chowdhary et al., 2024, Huber et al., 2023, Baptista et al., 2022]. Thus, we adopted the KL divergence framework to follow the standard convention and facilitate comparison with existing literature. For the landslide models under consideration, we typically have weakly informative priors for the parameters, since they are conceptual parameters that cannot be physically measured. In this context, maximizing KL divergence between posterior and prior is equivalent to minimizing the entropy of the posterior distribution.
Additionally, we have modified the discussion section to address the relative merits of entropy versus KL divergence for cases with non-uniform priors. In such settings, entropy can be a more suitable metric when posterior shifts are potentially misleading, since it focuses solely on posterior sharpness. In contrast, KL divergence is preferable when posterior shifts are informative, as it captures both posterior concentration and shift. However, both these metrics have limitations under model misspecification and prior-data conflict; thus, we recommend using posterior predictive validation as a complementary assessment tool.

Resulting changes in manuscript

To address the reviewer feedback we have modified *Section 2.3* (*p. 9, lines 233-260*). We introduced entropy and explicitly state its mathematical equivalence to KL divergence under uniform priors, and we added citations showing that KL divergence is the standard metric in the literature for quantifying information gain.

2.3 Data selection using information-theoretic metrics

In the Bayesian framework, our prior beliefs about the parameters θ are updated by observations, but the extent of this update varies depending on which observations are used. To quantify how informative each candidate observation is, we measure the information gained during calibration with observation y. Observations reduce parameter uncertainty by updating the prior distribution $P(\theta)$ to the posterior distribution $P(\theta \mid y)$. Thus, the information gained during this updating process corresponds to the reduced uncertainty. A fundamental information-theoretic measure of uncertainty is the entropy. For a random variable θ with probability distribution $P(\theta)$, entropy $H(\theta)$ is given as:

$$H(\theta) = -\int P(\theta) \log P(\theta)\, d\theta \tag{1}$$

Higher entropy indicates greater uncertainty about the parameter. Therefore, the change in entropy from prior to posterior quantifies the information gained from an observation. A standard measure to quantify information gain based on change in entropy is KL divergence [Kullback and Leibler, 1951], also known as relative entropy. In the Bayesian setting, KL divergence between the posterior distribution ($P(\theta \mid y)$) and the prior distribution ($P(\theta)$), as defined in Equation (2), admits a direct interpretation as information gain, since it quantifies the reduction in uncertainty about parameter θ achieved by conditioning on observation y. Higher KL divergence indicates greater reduction in uncertainty from prior to posterior, implying that the observation provides greater information about the parameters. Accordingly, we use this quantity as a criterion for data selection, consistent with its standard use in Bayesian optimal experimental design and inverse problems [Haeusel et al., 2026, Chowdhary et al., 2024, Huber et al., 2023, Baptista et al., 2022].

$$D_{\mathrm{KL}}(P(\theta \mid y) \,\|\, P(\theta)) = \int P(\theta \mid y) \log \left(\frac{P(\theta \mid y)}{P(\theta)} \right) d\theta \tag{2}$$

In the context of our work, we focus on calibrating landslide runout models with conceptual parameters that cannot be measured physically. For these parameters, we have limited prior information such as physically plausible bounds derived from literature and domain expertise. We therefore use uniform priors within these bounds, following standard practice in landslide runout calibration [Aaron et al., 2019, Navarro et al., 2018, Moretti et al., 2017]. For uniform priors, the prior entropy $H(P(\theta))$ is constant, so maximizing KL divergence is equivalent to minimizing the posterior entropy $H(P(\theta|y))$ (see Appendix A for derivation). In our numerical experiments with uniform priors, maximizing KL divergence is therefore equivalent to selecting observations that minimize posterior entropy, yielding the sharpest and most informative posterior distributions.

Additionally we included the limitation on the use of KL divergence as measure of improved parameter calibration in the presence of model misspecification and prior-data conflict, particularly when using non-uniform priors, (*p. 30, lines 551-559*).

Second, when model misspecification or prior-data conflict occurs, posteriors can contract to incorrect regions of parameter space, where higher KL divergence does not necessarily indicate better calibration. This limitation is exacerbated when using non-uniform priors. For non-uniform priors, KL divergence between the posterior and the prior no longer reduces to posterior entropy alone, as it captures both the sharpness of the posterior and its shift from the prior distribution. This introduces a trade-off: when a posterior contracts to an incorrect region, KL divergence can assign high information gain to a misleading result because it rewards both concentration and shift. Conversely, entropy focuses solely on posterior sharpness and does not account for whether the posterior has shifted toward more plausible parameter regions. When applying this framework in such scenarios, practitioners should therefore complement these information-theoretic metrics with robust validation methods such as posterior predictive checks.

1.3 Comment 3

> I found the Figure 2 and 3 schematics confusing, as by first look it looks like information flows from priors to likelihoods, which doesn't make sense. I don't have good suggestion how to change them, but wanted to mention this if the authors would have other ideas.

Response

Thank you for pointing this out. We agree that current schematics in Figure 2 and 3 could be misinterpreted as showing information flowing from the prior to the likelihood. These figures are intended to illustrate the Bayesian updating process, where the prior distribution is updated by the observational data through the likelihood function. To clarify this relationship, we have explicitly added a "Bayesian Updating" box that takes the prior and likelihood as inputs and produces the posterior. The revised figures are reproduced here as fig. 1 and fig. 2 for your convenience.

Resulting changes in manuscript

To address the reviewer's comment we have added the revised figures to the manuscript.

1.4 Comment 4

> The author's mention which MCMC algorithm is used, so the authors could also mention which convergence diagnostics were used from ArviZ package.

Response

Thank you for the suggestion. To assess the convergence of the MCMC chains, we relied on both qualitative and quantitative diagnostics from the ArviZ package. Qualitatively, we examined trace plots of the MCMC chains to check for proper mixing and to ensure the chains adequately explored the parameter space. Quantitatively, we used the Gelman-Rubin statistic (also called the potential scale reduction factor, $\hat{R}$), which assesses whether multiple chains have converged to the same distribution by comparing the variance between chains to the variance within chains. We adopted the standard threshold of $\hat{R} < 1.01$ to confirm adequate convergence as proposed by Vehtari et al. [2021]. The traces plots for selected experiments are provided here for reference as figs. 3 to 8 and the corresponding $\hat{R}$-values are reported in table 1. The $\hat{R}$ values reported in table 1 satisfy the convergence criterion $\hat{R} < 1.01$ for all but two cases, which attained $\hat{R} = 1.01$ and $\hat{R} = 1.011$ respectively. For these marginal cases, we found that the effective sample size (ESS) was adequate, extending the chain length reduced both $\hat{R}$ values below the 1.01 threshold, confirming convergence.

Resulting changes in manuscript

We have specified the diagnostics used to assess convergence of MCMC chains in the manuscript (*p. 13, lines 328-329*). Corresponding trace plots and the table containing $\hat{R}$-hat values are provided in the Appendix B of the revised manuscript.

1.4.1 Comment 5

> Explicitly define the prior in case study, now it seems based on the plots that it's uniform, but the priors were not explicitly defined in the text. This make huge difference for the use of KL, as it's then equivalent to entropy which measures just the sharpness. Without explicitly mentioning the prior in the text, it takes more time and effort from the reader to see what has been actually used.

Response

Thank you for spotting this. We agree that interpretation of KL divergence results changes based on the choice of prior. We have therefore revised Section 3.3 (Design of Numerical Experiments) to provide

a clear description of the prior distributions used in the case study. As you correctly noted, we employ uniform priors, which we have now explicitly stated along with appropriate justification and citations.

Resulting changes in manuscript

Added description of priors used in Section 3.3 (*p.18, lines 408-410*), repeated below for convenience.

> In all the 8 experiments we assume uniform priors for the friction parameters. The bounds for the priors are chosen from the literature as discussed in Section 2.3, for $\mu \in [0.02, 0.3]$ and $\xi \in [100, 2200]$ m/s^2. Additionally in experiment 7 and 8, we assume an uniform prior for the discrepancy parameters $\sigma_{vel_{TS}}$ and $\sigma_{pos_{TS}}$ with bounds $[1, 5]$ and $[0, 100]$ respectively.

2 Figures from manuscript

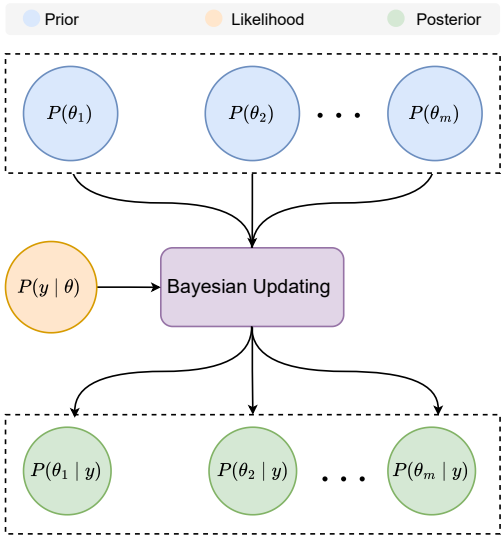

Figure 1: Schematic illustration of the Bayesian inference process. The likelihood function acts as the core driver of updating the prior distribution to posterior distribution based on observational data. The resulting posterior depends strongly on type and quality of the observation.

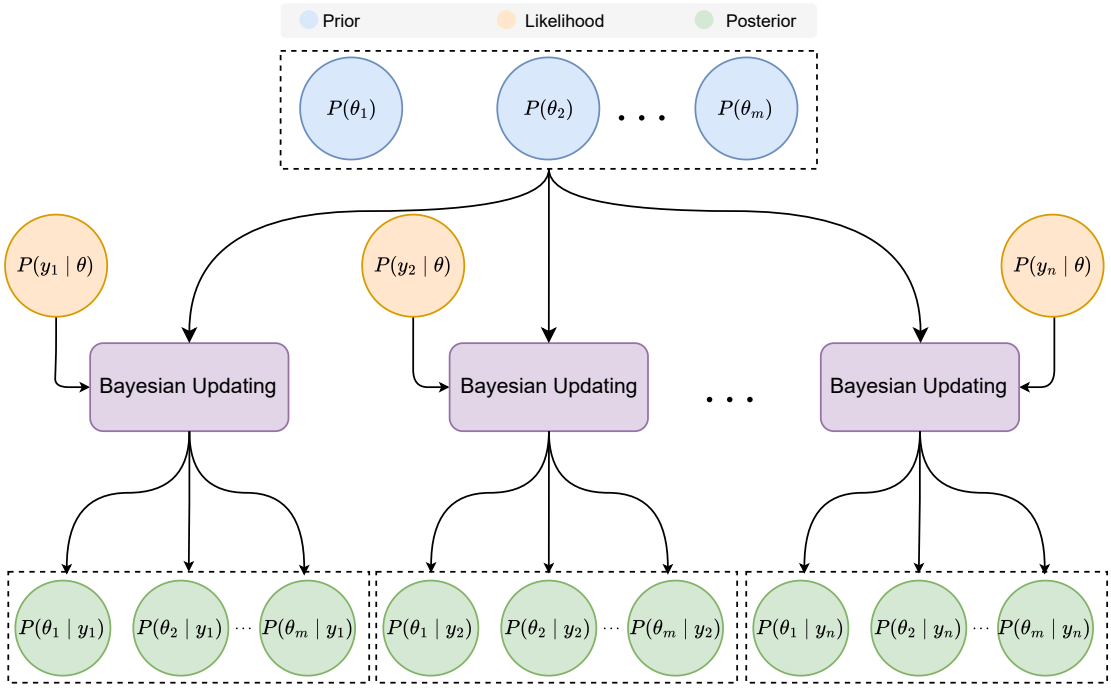

Figure 2: Schematic illustration of the Bayesian data selection process using information-theoretic metrics. Multiple calibration routines are performed in parallel, each using the same likelihood function but leveraging different observations to update the prior distributions. By comparing each resulting posterior distribution against the prior, we can quantify the information gained during calibration relative to observations.

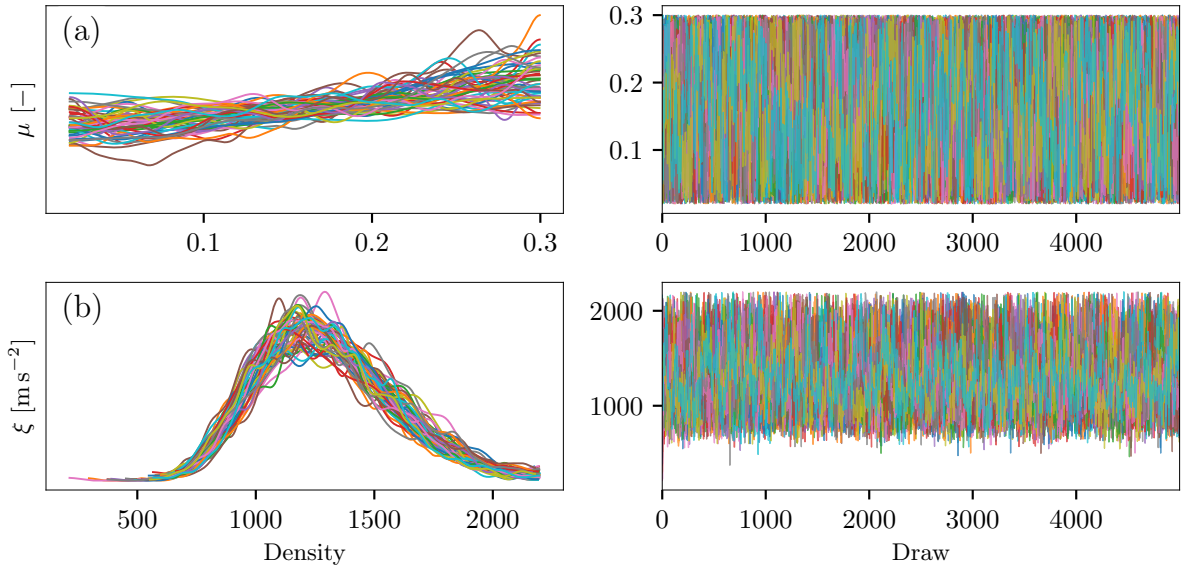

Figure 3: MCMC trace plots for (a) Dry Coulomb friction coefficient (μ) and (b) Turbulent friction coefficient (ξ) based on calibration with maximum velocity ($U_{\max}$) as observation. Each color represents an independent chain. Left panels show kernel density estimates of the marginal posterior distributions. Right panels show trace plots across iterations, demonstrating chain convergence and adequate mixing.

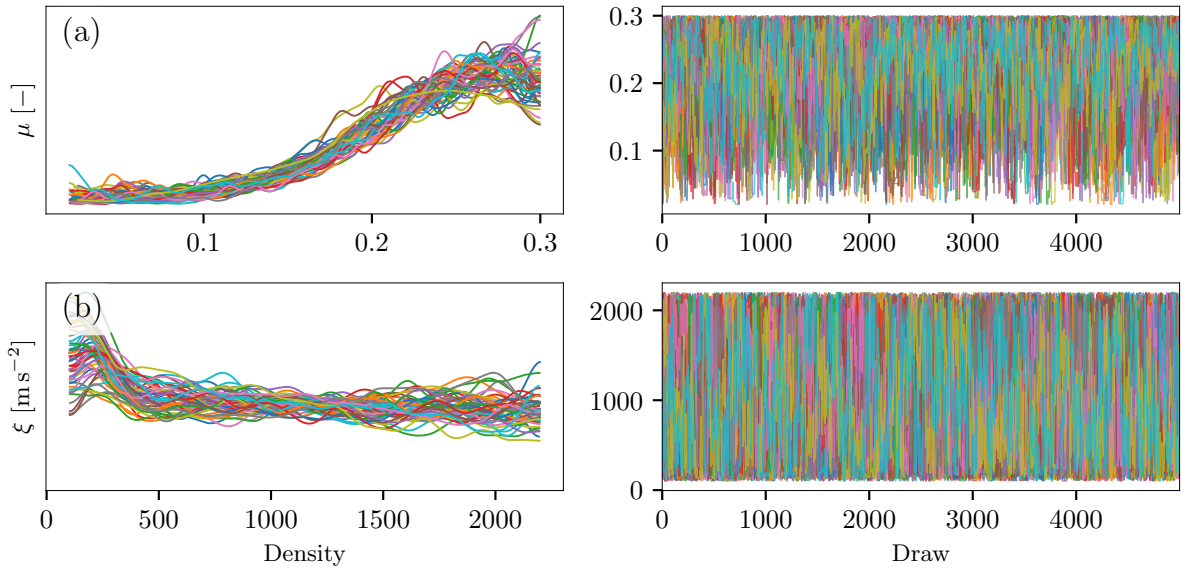

Figure 4: MCMC trace plots for (a) Dry Coulomb friction coefficient (μ) and (b) Turbulent friction coefficient (ξ) based on calibration with runout distance (X_{end}) as observation. Each color represents an independent chain. Left panels show kernel density estimates of the marginal posterior distributions. Right panels show trace plots across iterations, demonstrating chain convergence and adequate mixing.

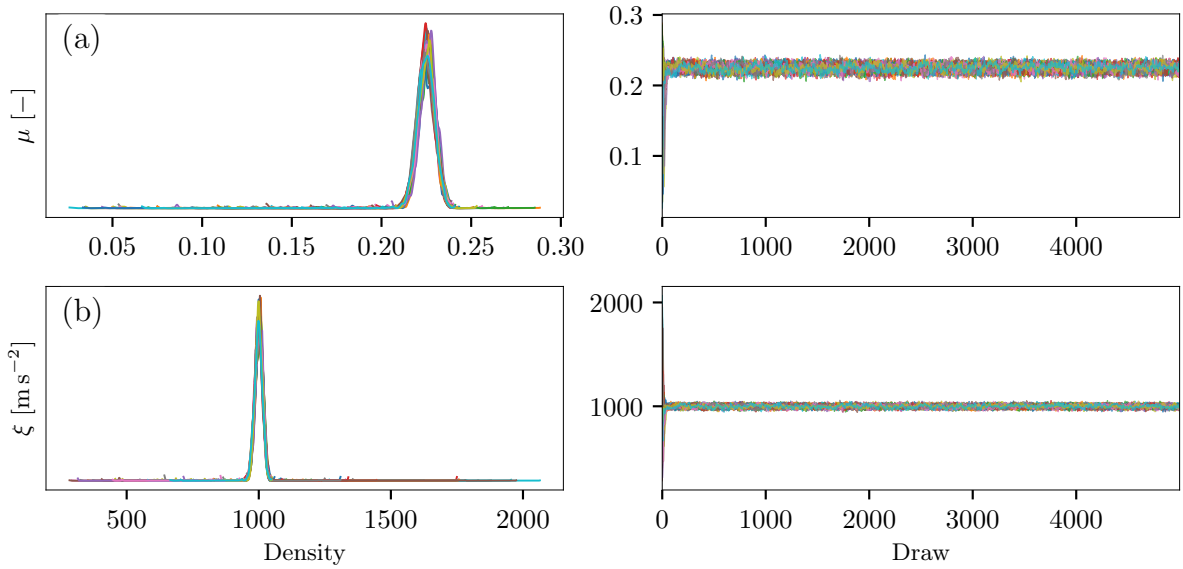

Figure 5: MCMC trace plots for (a) Dry Coulomb friction coefficient (μ) and (b) Turbulent friction coefficient (ξ) based on calibration with velocity time series as observation. Each color represents an independent chain. Left panels show kernel density estimates of the marginal posterior distributions. Right panels show trace plots across iterations, demonstrating chain convergence and adequate mixing.

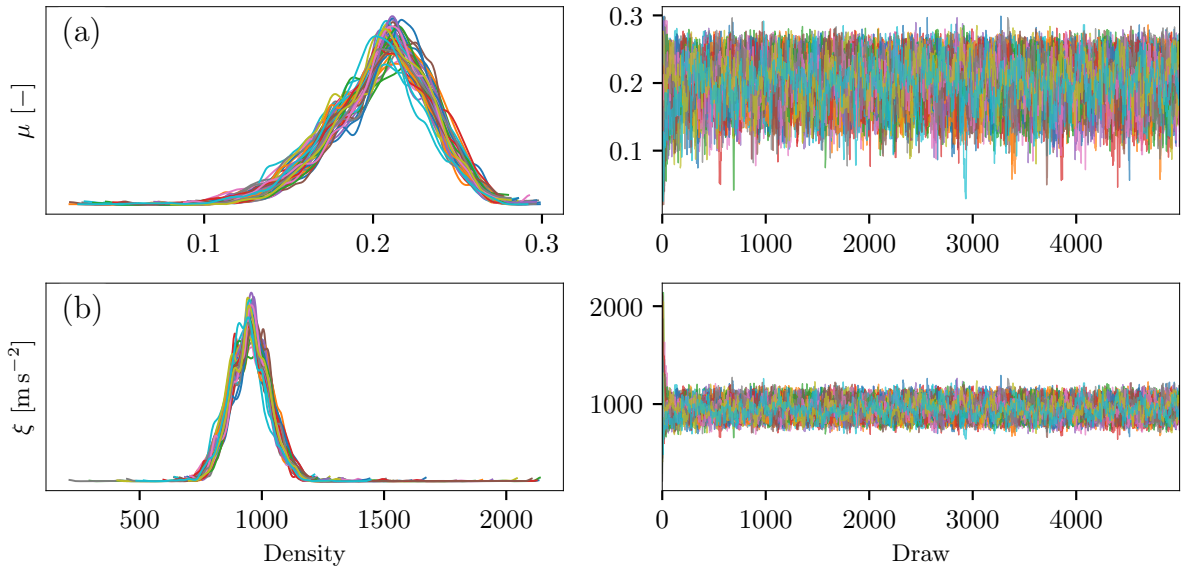

Figure 6: MCMC trace plots for (a) Dry Coulomb friction coefficient (μ) and (b) Turbulent friction coefficient (ξ) based on calibration with position time series as observation. Each color represents an independent chain. Left panels show kernel density estimates of the marginal posterior distributions. Right panels show trace plots across iterations, demonstrating chain convergence and adequate mixing.

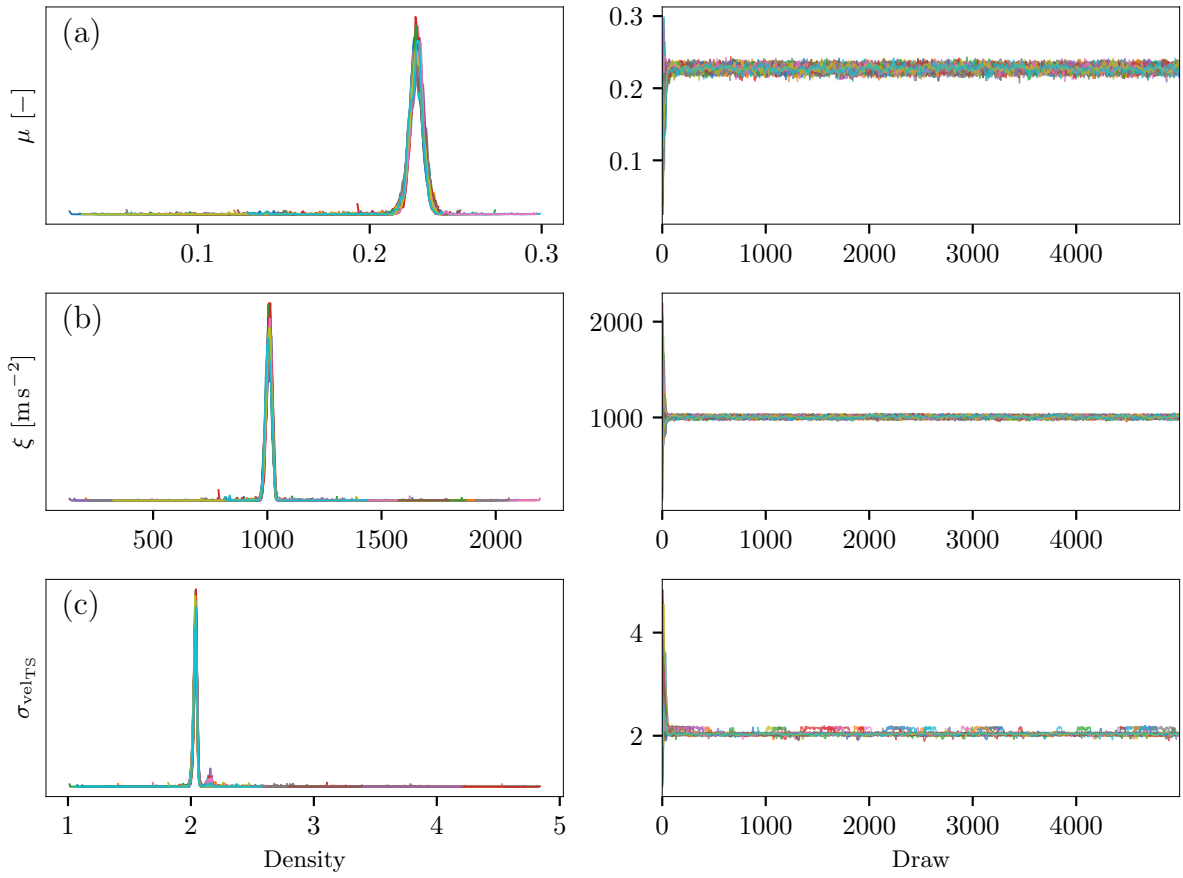

Figure 7: MCMC trace plots for (a) Dry Coulomb friction coefficient (μ) and (b) Turbulent friction coefficient (ξ) (c) Velocity discrepancy parameter ($\sigma_{\mathrm{vel_{TS}}}$) based on calibration with velocity time series as observation. Each color represents an independent chain. Left panels show kernel density estimates of the marginal posterior distributions. Right panels show trace plots across iterations, demonstrating chain convergence and adequate mixing.

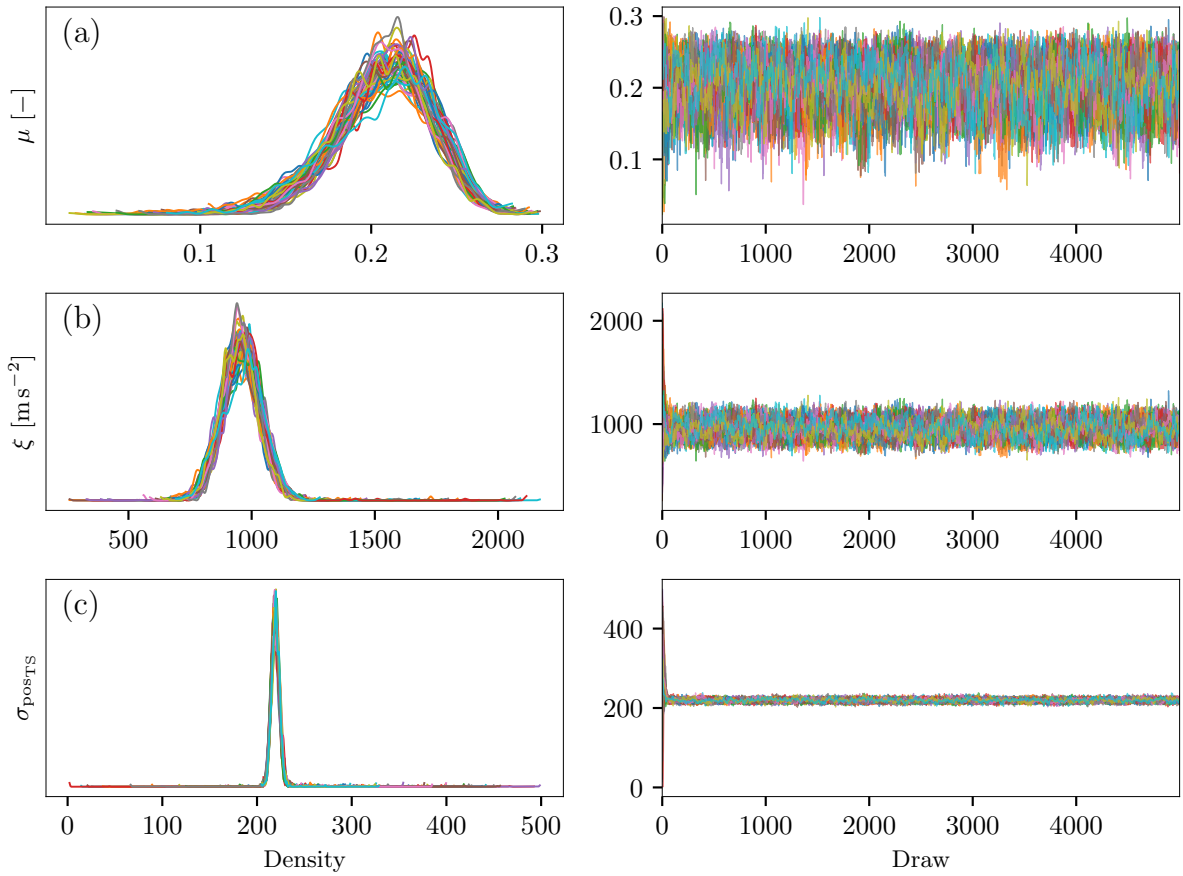

Figure 8: MCMC trace plots for (a) Dry Coulomb friction coefficient (μ) and (b) Turbulent friction coefficient (ξ) (c) Position discrepancy parameter ($\sigma_{\mathrm{pos_{TS}}}$) based on calibration with position time series as observation. Each color represents an independent chain. Left panels show kernel density estimates of the marginal posterior distributions. Right panels show trace plots across iterations, demonstrating chain convergence and adequate mixing.

Observations	R_hat		
	Dry Coulomb friction coefficient	Turbulent friction coefficient	Discrepancy parameter
Maximum velocity $U_{\max}$	1.007	1.007	–
runout distance X_{end}	1.008	1.009	–
Velocity time series u(t)	1.006	1.005	–
Position time series x(t)	1.010	1.010	–
Velocity time series u(t) (discrepancy)	1.007	1.008	1.011
Position time series x(t) (discrepancy)	1.009	1.009	1.007

Table 1: Gelman-Rubin statistics ($\hat{R}$) for MCMC sampling of friction coefficients and the discrepancy parameter. $\hat{R}$ values indicate successful chain convergence for all parameters across different observations.

References

Lea J. Haeusel, Jonas Nitzler, Lea J. KÃűglmeier, and Wolfgang A. Wall. Multi-physics-enhanced bayesian inverse analysis: Information gain from additional fields. *Computer Methods in Applied Mechanics and Engineering*, 452:118735, 2026. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2026.118735. URL https://www.sciencedirect.com/science/article/pii/S0045782526000095.

Abhijit Chowdhary, Shanyin Tong, Georg Stadler, and Alen Alexanderian. Sensitivity analysis of the information gain in infinite-dimensional bayesian linear inverse problems. *International Journal for Uncertainty Quantification*, 14(6):17–35, 2024. ISSN 2152-5080. doi: 10.1615/int.j.uncertaintyquantification.2024051416. URL http://dx.doi.org/10.1615/Int.J.UncertaintyQuantification.2024051416.

Holly A. Huber, Senta K. Georgia, and Stacey D. Finley. Systematic bayesian posterior analysis guided by kullback-leibler divergence facilitates hypothesis formation. *Journal of Theoretical Biology*, 558:111341, 2023. ISSN 0022-5193. doi: https://doi.org/10.1016/j.jtbi.2022.111341. URL https://www.sciencedirect.com/science/article/pii/S0022519322003320.

Ricardo Baptista, Lianghao Cao, Joshua Chen, Omar Ghattas, Fengyi Li, Youssef M. Marzouk, and J. Tinsley Oden. Bayesian model calibration for block copolymer self-assembly: Likelihood-free inference and expected information gain computation via measure transport, 2022. URL https://arxiv.org/abs/2206.11343.

S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951. URL https://api.semanticscholar.org/CorpusID:120349231.

J. Aaron, S. McDougall, and N. Nolde. Two methodologies to calibrate landslide runout models. *Landslides*, 16:907–920, 2019. doi: 10.1007/s10346-018-1116-8.

Maria Navarro, Olivier P. Le Maître, Ibrahim Hoteit, David L. George, Kyle T. Mandli, and Omar M. Knio. Surrogate-based parameter inference in debris flow model. *Computational Geosciences*, 22(6):1447–1463, Dec 2018. ISSN 1573-1499. doi: 10.1007/s10596-018-9765-1. URL https://doi.org/10.1007/s10596-018-9765-1.

L. Moretti, A. Mangeney, F. Walter, Y. Capdeville, T. Bodin, E. Stutzmann, and A. Le Friant. Constraining landslide characteristics with bayesian inversion of field and seismic data. *Geophys. J. Int.*, 2020:1–15, 2017. doi: 10.1093/gji/ggaa056.

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2):667 – 718, 2021. doi: 10.1214/20-BA1221. URL https://doi.org/10.1214/20-BA1221.