Review of manuscript

Bayesian data selection to quantify the value of data for landslide runout calibration

by V Mithlesh Kumar , Anil Yildiz, and Julia Kowalski

The authors present a framework for the friction and measurement noise parameter estimation in landslide observations. These values are essential for landslide runout models in order to obtain reliable values for runout length and velocity. That makes the work very valuable and beneficial. They use a Bayesian approach to estimate the information gain based on various observations. The amount of information gain is measured by the Kullback-Leibler-Divergence of the prior and posterior parameter distributions in the Bayesian inference. Based on the KL divergence the most relevant observations for calibration of a certain parameter is chosen. This is a promising approach that has the potential to improve landslide runout predictions.

General remarks:

The first question that comes to my mind is if the results are easily transferable from one region to another. Once a landslide occurred and observations for the calibration are available I suppose the danger of a landslide at the same location is rather low (my understanding, I am not a specialist in this field). If so what are the values of the information gathered at one site? Can they be used at other landslide-threatened locations? So, the applicability of the results should be pointed out a bit better.

There is no description of the geometry (or anything else) of the landslide model the authors used to generate the synthetic data for the experiments. This way it is hard to estimate if it is a very idealised case that is difficult to transfer to reality or not. Also, the friction parameters in the model are uniform on the entire sliding surface. How would the results look like if the sliding surface was complicated and parameters vary within the model?

How did you choose the prior distributions for the friction and discrepancy parameters? They look almost evenly distributed, but not quite. Why?

Regarding information gain: KL divergence is just a measure for the dissimilarity of distributions. Its value could be high even though the resulting MAP is far from the true value. So it is a little bit misleading (or optimistic) to call it information gain. The authors need to elaborate a bit why the value of KL divergence alone lets one infer the improvement of the parameter estimation.

Almost all axis labels are missing in the figures. Also labels of curves in the legends. This makes reading (and reviewing) the manuscript rather inconvenient since some guess work is involved. This is a major mistake and must be fixed!

Specific remarks:

l. 43: “parameters’” since rest of sentence is plural.

l. 65: better: “uncertainties“

l. 113: What does “high variance related issues” mean exactly?

l. 162 &164: the equations are not consistent. Exchange y^r and y in eq. (3)?

l. 177: why does the appearance of epsilon change? Typo or different meaning? If the latter, explain!

l. 210: The operator det(\Sigma) should not be written in italic.

l. 218: Better describe MCMC method used. How is stationarity ensured?

l. 238: What is the structure of \bar{y}^{*}? Is it an element of $\{y_1, y_2, \ldots, y_n\}$?

l. 255: Figure 5 is referenced before Figure 4.

l. 264: Further explain and justify usage of Gaussian emulator.

l. 270: PsimPy is written in typewriter-like font here but in normal font earlier. Please be consistent.

l. 284: Section 2.4.3 (Data selection) describes the process of computing the KL divergence for all combinations of observations and parameters. However, nothing is written about selecting data (or observations) based on the resulting matrix shown in Fig. 4. This needs to be described.

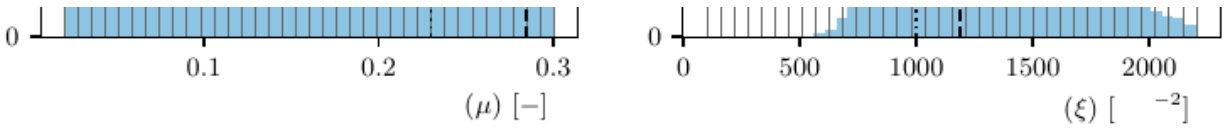
l. 316: Subscripts that are not variables or indices should not be italic but normal font (i.e. F_{\mathrm{res}}). Furthermore, it is inconvenient to use \theta for the angle here, since it is used as variable for the parameters earlier in the manuscript.

l. 360: Experiment numbers (1–9) do not match numbers in table.

l. 361: Italic subscripts that are not variables or indices.

l. 364, Figure 6: Some information are missing / not printed (at least in the provided pdf file): legend incomplete (no text next to lines);

no axis labels; unit of \xi not printed:



Same issues with all the other figures of prior and posterior distributions. When it is unclear which line represents which quantity, it is nearly impossible to estimate the information content and correctness of the figures! This especially difficult for figure 11.

l. 366: Subscript “max” should not be italic.

l. 368: Are six significant digits for \xi justified here?

l. 372: Where are the KL divergence values given / shown? → reference Table 2

l. 387: “x(t)” here is just text, not typeset as a formula. Please be consistent.

l. 393: Figure 10: axis labels are missing. Are these examples taken from a set of 100 experiments? If so please state that fact. Otherwise the reader wonders how from 6 values you get such a smooth curve in figure 11a.

l. 396: How is it visible that initial time steps have greater information content? That does not become clear. You show graphs for varying numbers of time steps but do not compare different onsets of the time series. Please clarify!

l. 399: better: “number of velocity time steps”. Otherwise it could be interpreted as KL divergence vs. specific time steps. Same in next sentence on line 400: better “Increasing number of time steps…”

Figure 11a: What is the reason for the shape of the KL div curve? Why is there a plateau between 300 and 500 steps and then an increase again?

I suppose figure 11 shows number of time steps used (a) and time steps (b) but that’s a guess since x-axis labels are missing. If that is so the start of the plateau corresponds to about 200 steps. It would be interesting to see that length also in figure 10. In figure 10 the three uppermost curves (1000, 1100, 1200 steps) look very similar. In figure 11a, however, the KL divergence is still increasing significantly in that range. Why is that?

l. 410: Figure 12 is missing axis-labels again. A plot of KL divergence vs. time step would be informative, too. Which time window is used here for the observations?

l. 413/414: you refer to section 2.1 where the discrepancy, i.e. measurement error is called \epsilon. Here you write \sigma which seems to refer to the STD of applied noise as given in Table 1. Be consistent with naming scheme!

l. 512: Figure 11b shows the max. acceleration (assuming it is the blue line) at about 200 (no unit given). Figure 10 shows a narrow distribution only for 1000 steps or more. So how do you come to this optimality statement? That does not seem reasonable. There is not even a curve for 200 steps in figure 10.