Response to Reviewer #1

We thank Referee #1 for her/his time and helpful suggestions to clarify the manuscript. The comments made by the Reviewer are in black, and our responses are in blue.

General comments:

This paper demonstrates a useful application of a "bottom up" approach to understanding drivers of landslide behavior. I caveat that I am reviewing mostly in my capacity as one familiar with analytic methods to support decision making under deep uncertainty, and have minimal landslide modeling expertise. Therefore, my comments are mostly restricted to this realm, and I trust other landslide modeling experts will provide review on that front.

On this subject, my impression is that, for the scope the paper has limited itself too, I believe the treatment given to the subject matter is excellent. This includes the blending of distributions and deep uncertainty, and utilization of the full feature-set of CART, and introduction of an auxiliary variable to simplify the tree structure with minimal loss of performance.

In general, I believe the paper should be revised to more clearly discuss the limited nature of the findings, or better yet, explore a few other key parameters that would greatly enhance the ability to speak to issues of generalizability – the most clear being, slope, site, cell size, and time step. (To be clear, I don't necessarily think all of these things need to be explored for the paper to be publishable, but they should be spoken to. For many, the authors could just explain why they are confidence the qualitative nature of the findings will not change.)

Authors' reply: We thank the Reviewer for the positive comments. The main contribution of our study is the development of a methodology to evaluate landslide hazard, which accounts for uncertainties associated with slope characterisation (geometric, geotechnical and hydrological properties) and deep uncertainties related to climate change. Furthermore, we demonstrate how by accounting for interactions between variables, classification trees can be simplified substantially to improve communication of results without any loss of predictive accuracy. This methodology is generic and applicable to any slope, subject to the physically-based model being able to capture adequately the critical processed affecting slope stability in that area.

However, the Reviewer is correct to highlight that our findings, i.e. the specific factors controlling slope stability, are likely to vary from slope to slope and therefore it is difficult to make any specific claims out whether the factors we identify as major drivers of slope failure would also be important in other sites. Nevertheless, our proposed methodology could be easily applied to other sites to assess the dominant factors affecting slope stability. We discuss these points briefly on Sect. 5 (page 10, lines 30-34 of the original manuscript) and on Sect. 6 (page 12, lines 28-31 of the original manuscript). However, to clarify this for the reader, in the revised manuscript we will reword the last paragraph of Sect. 6 as follows:

"The factors identified as drivers of slope failure in this study are specific to the slope investigated and for the assumptions made about the system representation, i.e. the choice of the CHASM model and its resolution (spatial and temporal). While this limits the ability to generalise our findings to other sites, it is important to note that our proposed methodology can also be easily applied to other sites to assess the dominant factors affecting slope stability. Future work will seek to expand our analysis to a broader range of slope conditions found in a wider study region, for example to analyse effects of variable slope angles and heights on slope stability."

With regard to the Reviewer's additional comment about our choice model cell size and time step, we acknowledge that we have not discussed this in the original manuscript and it would be helpful to consider this in more detail. Therefore, we will add the following text at the end of paragraph 1 of Sect. 3.1 explaining our decision to fix the discretisation parameters (time step, cell size, etc):

"These discretisation parameters are selected based on preliminary model simulations, and our previous experience of applying CHASM to similar slopes, to ensure the numerical stability and conservation of mass (water) of the hydrological component throughout the simulation period. These choices of discretisation parameters minimised the number of failed model runs during the Monte Carlo simulation process."

It would also be useful to more clearly orient the reader early-on to how the Monte Carlo sampling mixes with the deep uncertainty sampling.

Authors' reply: We agree with the Reviewer. To orient the reader early-on regarding how the Monte Carlo sampling mixes with the deep uncertainty sampling, in the revised manuscript we will add the following text at the beginning of Sect. 3 (page 5, line 26 of the original manuscript):

"Such combinations are generated via random sampling from a set of probability distributions that characterise the uncertainty in the slope properties, and uniform distributions with very wide ranges for the intensity and duration of future rainfall events, so as to reproduce the (practically) unconstrained nature of 'deep' uncertainties."

Specific comments:

P2:

L19 – 21: I understand how censored or biased availability of data limits statistical models relative to physically-based models, though would be nice (but not required) to speak to the issue of whether physically based models have large domains of applicability if their development and testing is often limited to the same regimes of data availability. Put more plainly: Yes, you can run physically based models in the domains where you might not have good enough data to build a statistical model, but might you still be tenuously extrapolating?

Authors' reply: This is an interesting discussion point, but maybe conflates the slightly different requirements for applying statistical models versus physical-based models in new regions (i.e. extrapolation). The usefulness of a physically-based model inevitably depends on how reliably and realistically it can represent the key processes that determine slope stability in a given location. For a statistical model there is a need for sufficient landslide inventory data as well as some physical characterisation of the region by topography, soil etc. Of course, if a physically-based model is used that is unsuitable for the chosen case study then that model will be unlikely to provide useful information about the drivers of slope failure. As the Reviewer correctly notes, one can still run physically-based models in such data limited/scarce conditions. One of the aspects of our study is how to evaluate the impacts of different uncertainties on model predictions (i.e. slope stability).

To briefly emphasise this point for readers, in the revised manuscript we will add the following text:

"Nevertheless, when using physically-based models to assess slope stability it is important to note that care should always be taken to ensure that the selected model adequately represents the key processes determining slope stability for the chosen study site."

P4:

Research question 2 seems overly vague and of ambiguous relevance without sharpening it a bit. The discussion does that better, but maybe you could reverse engineer that question a little to be more precise.

Authors' reply: In the revised manuscript we will rephrase question 2 to:

"Does deep uncertainty in future landslide triggers, such as climate change, exceed other uncertainties, such as those related to slope properties, in determining our ability to predict slope failure?"

L8: "has a track-record: …" – causes the reader to expect citations here (I see some are provided farther below).

Authors' reply: Thank you for pointing this out. To address the Reviewer comment we will add in brackets "(references provided in Sect. 2.2)".

P5:

L4-8: Would be strengthened by providing some context on what "good" classification rates are in this arena, both from a scientific perspective and a policy-relevance perspective. It would be great to also include at least one more validation exercise if such exists.

Authors' reply: We agree. In the revised manuscript, we will add a comparison with two spatially distributed (GIS-based) studies based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) and associated statistic, Accuracy.

We will expand paragraph 1 in Sect. 2.2 by adding the following text:

"In a validation exercise in Hong Kong CHASM was shown to be numerically robust and capable of correctly classifying 77% of failed slopes and 68% of stable slopes (i.e. true positives, TP=77%; false positives, FP=23%; true negatives, TN=68%; false negatives, FN=32%) for a specified rainfall event (Anderson, 1990), corresponding to an accuracy of 72.5% ((TP+TN) / (total observed failed slopes + total observed stable slopes)). This is comparable with, or exceeds, the performance of spatially distributed GIS-based models such as a statistical analysis of landslide susceptibility in Central America (using the Hurricane Mitch landslide inventory) that attained an accuracy of 68% (Kirchbaum et al, 2011); and a physically-based infinite-slope analysis (using modified SINMAP) in the data-rich region of Calabria, Italy, that achieved a maximum TP rate of 71% and accuracy of 72% (Formetta et al., 2014, p. 639). This level of accuracy might typically enable Disaster Risk managers and practitioners (such as engineers and planners) to identify slopes exhibiting potentially high hazard and prioritise further investigation and/or risk reduction accordingly. CHASM has been extensively used by slope stability researchers and practitioners to assess landslide hazards along roads and in urban and rural areas, and to propose appropriate mitigation in Malaysia, Indonesia, the Eastern Caribbean, United Kingdom and New Zealand (Anderson et al., 1997; Lloyd et al., 2001; Wilkinson et al., 2002a; Wilkinson et al., 2002b)."

An additional validation exercise was carried out as part of Holcombe's (2006) PhD thesis. This study included some slopes for which shear box testing had been carried out. It therefore achieved a higher accuracy than Anderson (1990) and is not directly comparable to our current study (or spatially distributed approaches) where larger population of slopes are defined by weathering grade or soil type. However, for the Reviewer's information we provide the details here as follows: Number of slopes=30 TP=86%, TN=73%, FP, 18%, FN=27%, accuracy=77.5%, POD=82%.

References:

Formetta, G., Capparelli, G., Rigon, R., and Versace, P. Physically based landslide susceptibility models with different degree of complexity: calibration and verification. In: Ames, D. P., Quinn, N. W. T., and Rizzoli, A. E. (Eds.), Proceedings of the 7th International Congress on Environmental Modelling and Software, June 15-19, San Diego, California, USA, 2014.

Holcombe, E. A. (2006) Modelling landslide risk on highway cut slopes in developing countries, PhD thesis, University of Bristol.

Kirschbaum, D.B., Adler, R., Hong, Y., Kumar, S., Peters-Lindard, C.; Lerner-Lam, A.: Advances in landslide nowcasting: evaluation of a global and regional modeling approach. Environ Earth Sci 66: 1683-1696 doi:10.1007/s12665-011-0990-3, 2012.

L12: "matric suctions" – is technical. If fine for assumed audience, that's fine, just flagging.

Author's reply: In the revised manuscript, we will add in brackets "negative pore water pressure", after matric suctions, to make it clearer.

L26: Minor: Strengthen by putting a number before "uncertain input factors."

Authors' reply: Following the Reviewer suggestion, we will add the number 28 in the revised manuscript.

P6:

L5-7: Since this is couched as bottom up and deep uncertainty oriented, might orient reader to this use of distributions and how that does or does not integrate with the deep uncertainty mode of analysis – which often (but not necessarily) operates without distributional assumptions.

Authors' reply: In the revised manuscript we will change the text to make this point clear. The sentence starting on line12, page 6 (original manuscript), will be modified to:

"While the uncertainties in slope properties are characterised by probability distributions based on past experience of applying the model to the study area, variability of potential future rainfall is difficult to define with a probability distribution, hence the term of 'deep' uncertainty."

L7-8: For reproducibility, would be good to describe where the "checks undertaken" are explicitly listed.

Authors' reply: In the revised manuscript we will refer the reader to Table 1, where all checks undertaken are listed. Those are:

- Saturated soil moisture content greater than residual soil moisture content (added in footnote d Table 1)
- Residual soil content greater than 0 (footnote d Table 1)
- Van Genuchten suction-moisture curve n greater than 1 (footnote c Table 1)
- Effective cohesion greater than 0 (footnote f Table 1)

P11:

L11: This statement suggests there is a significant interaction effect between climate and other source of uncertainty, or that one conducting the sensitivity analysis independently would not think to add the two elementary effects of climate and other variables.

Authors' reply: Indeed, that was our intention when we wrote this sentence. It is common for studies to consider either one or the other. As we note on page 11, Melchiorre and Frattini (2012) are the only ones, to our best knowledge, to have considered both.

Figure 6: Seeing this visually makes me think the authors might also want to consider linear discriminate analysis – not at all a required change, just for consideration.

Authors' reply: CART is one, among others, option to analyse our results. We chose CART in this particular study due to its simplicity in communicating the results, particularly to non-specialists. Linear discriminate analysis is another option but we are not sure it could be applied in this case since it assumes that the independent variables are normally distributed, which is not the case in our study.