Response to Referee #2:

Dear Dr. Ben Mirus,

We would like to thank you for the careful reviews and constructive comments, which really lead to an improvement of our manuscript. We are especially grateful for reminding us of the publication (Mirus et al.,2018b; https://doi.org/10.3390/w10091274). We are very excited to find there are some similar methods used in the threshold definition, like the hybrid threshold. Some inspirations have been taken from your work to improve our manuscript, like Figure 4 in the revised manuscript.

All the comments are addressed point by point, and the changes are tracked in the marked manuscript. It is worth noting some revisions requested by the other reviewer are also included in the revised manuscript. It is believed that all necessary changes are made to address every point of the concerns.

If any further information is needed, please don't hesitate to contact us.

Yours Sincerely

Binru Zhao

Water and Environmental Management Research Centre
Department of Civil Engineering
University of Bristol, Bristol, BS8 1US, UK
Email: bz17336@bristol.ac.uk

CENERAL COMMENTS

I agree strongly with the background information and justification of the study, but the objectives could be clarified as they raise some questions. The authors aim to investigate two issues:

1) The role of antecedent wetness information in landslide threshold definition, which seems to be the focus of their prior work (now published in Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2019.04.062).

Although there are some differences between the new study and the authors' prior work, there are some notable similarities (e.g., overall topic, study area, data used, techniques for analysis). Thus, before considered for publication, the new work should be revised to include reference to their previous work. Specifically, the authors must provide context for how this new work goes beyond their prior contributions. One notable difference, that could be highlighted, is the framework used in the present study to evaluate the value of different types of information content in landslide thresholds.

2) Whether or not it's necessary to explicitly consider antecedent wetness, or if it's acceptable to use only the recent rainfall condition instead.

However, in the approach they use, the proxies for antecedent wetness (e.g. API) are calculated with rainfall and temperature data only. So essentially, this study is merely comparing whether it is worthwhile to take antecedent rainfall and somehow transform it into a wetness index before developing landslide thresholds. It seems that this has already been addressed in prior studies references in the introduction (e.g., Glade, 2000; Godt et al., 2006). So even though the study uses measured soil moisture (at 10cm depth) to calibrate the recession parameter for the API, it is still a calculation with rainfall only, which is limiting. The manuscript cites our recent paper (Mirus et al., 2018a) in which we used actual soil moisture data and found similar improvements, but it does not recognize the follow up publication (Mirus et al.,2018b; https://doi.org/10.3390/w10091274) in which we further evaluated the appropriate timescales of antecedent saturation vs. recent rainfall. As such, the discussion should better recognize the limitations of the API approach in the context of other contributions in the literature (see specific comments below).

Reply: Many thanks for these suggestions. All these concerns have been addressed in the revised manuscript, and the specific responses are shown in the reply to specific comments below.

SPECIFIC COMMENTS

Point 1:
P3.L10 – In our more recent paper (Mirus et al., 2018b), we used ROC characteristics to evaluate different durations of antecedent saturation vs. recent rainfall for landslide thresholds, as well as to illustrate the impact of different choices in ROC skill metrics for hydro-meteorological threshold optimization. This is worth noting in the introduction.

Reply: Many thanks for reminding us of this publication. We have cited it in the introduction and discussion section.

In the introduction:

"Mirus et al. (2018b) and Mirus et al. (2018a) accounted for the antecedent wetness condition with direct subsurface hydrological measurements, which are then combined with the rainfall information to define the threshold for landslides. The derived thresholds show improved performances in landslide alert systems."

In the discussion:

"Mirus et al. (2018a) explored a wide range of timescales when developing hydro-meteorological thresholds for landslide initiation. They found that using 3 days as the separation works well for two sites in the Pacific Northwest of the United States. Besides, 3 days are widely used to separate the antecedent condition from the recent condition in the previous studies (Chleborad et al., 2008;Scheevel et al., 2017;Mirus et al., 2018b)."

Point 2:
P6.L15-16 – This is more or less the thresholds we identified in the aforementioned paper in Water (Mirus et al., 2018b). So it is interesting the timescales are similar.

Reply: The hybrid threshold in our study is similar to the thresholds identified in the paper in Water (Mirus et al., 2018b), and we find that the description 'bilinear threshold' is better, so this description is also used in the revised manuscript through citing your works. As for the timescales of 3 days, we take the inspiration from another paper (Mirus et al., 2018a), however, your exploration on the effect of timescale in the paper in Water (Mirus et al., 2018b) really provide more supports to the choice of the 3 day, which has been mentioned in the revised manuscript.

Point 3:
P7.L30 – typo: : : should be “no landslide” not “on landslides” occur.

Reply: Agreed and revised.

Point 4:

P8.E3&E4 – Should mention that HR and FAR are more commonly referred to as the TP_rate and FP_rate.

Reply: Agreed. In the revised manuscript, 'Hit Rate' has been replaced with 'Ture Positive Rate', and 'False Alarm Rate' has been replaced with 'False Positive Rate'.

Point 5:

P8.L8 – Is this the same as the optimal point criteria (often referred to as the radial distance)?

Reply: Yes, this is the same as the optimal point criteria. However, in this study, the distance is between the perfect point and the threshold point (there are 12 cases for each landslide threshold, represented with the point), the Euclidean distance is considered more accurate.

Point 6:

P8.L10 – This is confusing, why would you restrict the value of HR? In ideal circumstances HR should reach unity. Do you mean that there are multiple threshold values with HR=1, but rather than minimize the Euclidean distance, you do not allow HR<1.0 and the optimization focuses on reducing FAR instead?

Reply: Yes, for the 12 cases of each threshold, there are multiple threshold cases with HR = 1, and the optimal threshold is determined among these cases by minimizing FAR. This sentence has been rephased as follows:

"Sometimes owing to the danger of the missed alarms, the optimal one is chosen among thresholds with TPR as 1. In this case, the smaller the FAR value, the better the prediction performance. "

Point 7:

P10.L7-8 – Is a contingency a "null" event (i.e. day with no landslide)?

Reply: No, the contingency here means all possible events, including events with landslides and with on landslide. We have replaced 'contingency' with 'event' in the revised manuscript.

Point 8:

P10.L14 – Maybe not that unexpected, since the comparison between soil moisture and API is fairly poor (Figure 2).

Reply: This statement is under the assumption that the API could indicate the soil moisture condition, so the results are considered unexpected.

Point 9:

P10.L20-26 – These findings seem consistent with Figure 7 in Zhao et al. (Journal of Hydrology, 2019). Consider discussing the differences and similarities with your prior work.

Reply: Agreed. The corresponding discussion has been added:

"First, the comparison of the 3-day rainfall threshold and the hybrid threshold shows that including wetness information in the hybrid threshold could improve the false positive rate, compared with the 3-day rainfall threshold which only considers the recent rainfall information. As the only difference between these two types of thresholds is the incorporation of the wetness information, the improvement in the false positive rate is due to this factor. The work of Zhao et al. (2019) also demonstrates that integrating antecedent soil moisture conditions could improve the predictive capability of the cumulated event rainfall-rainfall duration (ED) thresholds, especially in terms of reducing false positives. However, the improvement directly contributed by the added soil wetness information is unexplored. This study is the first time to investigate this issue. The right plot in Figure 7 shows the proportion of the reduced false positives that is caused by the added antecedent wetness information, which could reach 35% for APIv1 and 52% for APIv2. Such high proportion of reduced false positives further illustrates the crucial role of the antecedent wetness information in affecting the landslide threshold's predictive capability. We also explored the extent to which the false positive rate is improved under different critical values of the 3-day cumulated rainfall. It is found that the false positive rate is improved more distinctly when a lower critical value of the 3-day cumulated rainfall is used. By including the antecedent wetness condition, events whose antecedent wetness condition is dry could be excluded from false positives, and thus reduce false positive rate. Given the dry wetness condition is more frequent in the dry season compared with the wet season, it is implied that incorporating the antecedent wetness condition to the landslide threshold is more advantageous in reducing false positives for the dry season."

Point 10:

P13.L6-7 – Assuming this improvement from APIv1 to v2 mostly reflects the better representation of soil moisture (Figure 2), this suggests that even better representation of soil moisture than either API version would be even better for threshold performance. Thus, one should recommend using a better model (e.g. Godt et al.,2006), which accounts for monthly variations in ET and an exponential decline to reflect faster drainage during wetter conditions. Or even more appropriate would be to use measured soil moisture or a better model of soil moisture (Mirus et al., 2018a,b).

Reply: Many thanks for this suggestion. We have added the corresponding text:

"The API_{v2}'s better representativeness of the soil moisture is also reflected in the threshold performance, where the thresholds based on API_{v2} present better prediction results than those based on API_{v1}. Therefore, it is implied that the better representation of the soil moisture could also benefit the threshold's prediction performance. The representation of the soil moisture could be improved by using the measured soil moisture (Mirus et al., 2018a;Mirus et al., 2018b) or other indexes estimated with a better model, like the water balance model proposed by Godt et al. (2006), which could account for the monthly variations in evapotranspiration and an exponential decline to reflect faster drainage during wetter conditions."

Point 11:

P13.L21-22 – What do you mean by a physical-based approach? Consider providing references that account for either the seasonality or the antecedent wetness explicitly (e.g., Napolitano et al., 2015, https://doi.org/10.1007/s10346-015-0647-5; Thomas et al., 2018, https://doi.org/10.1029/2018GL079662). Also, consider revising to "physics-based" or "physically based" rather than "physical-based."

Reply: Agreed. In the revised manuscript, 'physical-based' has been replaced with 'physics-based', and these references have been added:

"To better understand the role of the antecedent wetness condition and the recent rainfall in the occurrence of rainfall-induced landslides, a physics-based approach is expected. The understanding of the physic process could help construct the threshold which is more in line with the practice. For instance, Napolitano et al. (2015) explored the effect of seasonal variations of antecedent-hydrological conditions on rainfall triggering of debris flows by carrying out a hydrological and slope stability model. The results show the opposing winter and summer antecedent hydrological conditions exert a significant control on intensity and duration of rainfall triggering events. Thomas et al. (2018) designed thousands of storm patterns and coupled them with a physics-based hydrological and slope stability model for various antecedent wetness conditions, the pore water pressure and factor of safety metrics were then analysed. The proposed physics-based approach facilitates the exploration of the relative impact of plausible variations in soil hydraulic and mechanical properties on thresholds."

Point 12:

P13.L30 – yes, see suggestions to cite in above comments (L6-7).

Reply: Agreed. The mentioned references have been added in the manuscript.

Point 13:

P14.L1-4 – Indeed, in our more recent paper (Mirus et al., 2018b) we explored a wide range of timescales and still found that 3 days does work quite well for different cities in the Pacific Northwest of the United States. Of course, different regions should expect different durations of recent rainfall to correlate with shallow landslide occurrence, which is an important point to mention.

Reply: Many thanks for this suggestion. The corresponding text has been added:

"First, when separating the antecedent wetness condition from the recent rainfall, 3 days are selected as the boundary. Although there may be many other selections for this separation, the initial exploration we present here is intended as a proof-of-concept. We start by using 3 days as the separation to explore the role of the antecedent wetness condition and the recent rainfall in landslide thresholds. Mirus et al. (2018a) explored a wide range of timescales when developing hydro-meteorological thresholds for landslide initiation. They found that using 3 days as the separation works well for two sites in the Pacific Northwest of the United States. Besides, 3 days are widely used to separate the antecedent condition from the recent condition in the previous studies (Chleborad et al., 2008;Scheevel et al., 2017;Mirus et al., 2018b). Despite this, different regions should expect different durations of recent rainfall to correlate with shallow landslide occurrences."

Point 14:
P14.L7-8. This was also shown by Godt et al. (2006) with a better model and by Mirus et al. (2018a,b) using actual measured soil moisture.

Reply: Agreed. The mentioned publication has been cited in the revised manuscript.

Point 15:
P14.L16-24 – Although I agree with these conclusions, they do not represent a particularly novel or unexpected finding in the context of prior published work (see references list and papers cited in this review). As such, perhaps the paper is more suitable as a technical note, than as a research paper.

Reply: The conclusions have been rephased as follows:

"We presented a framework to explore the role of the antecedent wetness and recent rainfall information in the thresholds for landslides. The comparative study is carried out among four types of landslide thresholds. By including different variables that are responsible for landslide occurrences, these thresholds could represent different cases, like whether to include the antecedent wetness condition or whether to consider the recent rainfall explicitly. The important role of the antecedent wetness information in landslide thresholds is further reinforced. The false positives could be reduced by incorporating the antecedent wetness information in the threshold definition, where the proportion of reduced false positives could reach as high as 50%. It is beneficial for the threshold's predictive capability to include the antecedent wetness information and the recent rainfall condition more explicitly. It is also found the reliability of the soil moisture measurement is a key factor affecting the threshold's predictive capability. The proposed results provide a timely complement to the exploration on hydro-meteorological landslide thresholds. It is the empirical approach that we used to investigate the relative impact of different information in landslide thresholds, a physics-based approach is also expected to explore this issue, which would benefit the development of the hydro-meteorological thresholds in landslide early warnings."

As for the novelty of the work, although the landslide thresholds used in this study is not novel, this is the first time to design such a framework to carry out the comparative study. The designed comparison could help to investigate the relative impact of different information in the landslide threshold and facilitate the achievement of this study's objective. For the results proposed in this work, although the important role of the antecedent wetness information in the landslide initiation has been recognized in multiple studies, results on its effect on the threshold's prediction performance is rarely. In our study, we explore the direct impact of the added antecedent wetness information on threshold' prediction performance, results show the false positives could be reduced by incorporating the antecedent wetness information in the threshold definition, where the proportion of reduced false positives could reach as high as 50%. We also explored whether it is necessary to explicitly consider the antecedent wetness information and the recent rainfall condition in landslide thresholds. To our knowledge, this concept is firstly discussed in detail, although it is more or less involved in some researches. It is also found the reliability of the soil moisture measurement is a key factor affecting the threshold's predictive capability. This statement is well recognized; however, we support it with the research results. Based on the above and the length of the manuscript, we think our paper is more suitable as a research paper.

Point 16:
P18.T2 – Not sure this table is strictly necessary or beneficial.

Reply: This table has been deleted in the revised manuscript.

Point 17:
P19.T3&T4 – need to provide key for terms in headings, especially d, is that Euclidean distance?

Reply: Yes, d is the Euclidean distance. The headings have been modified.

Point 18:
P21.F2 – It seems that API in both cases is a very bad predictor of measured soil moisture. Why not use the actual measured soil moisture as we did in other studies (Mirus et al., 2018a,b)?

Reply: The reason of not using the actual measured soil moisture is the unavailability of such type of data. Considering the data's availability and completeness, we choose API as the proxy of soil moisture due to its less requirement for data. Its poor representation of soil moisture is really a limitation of our study. If more accurate data is available in the future, we will carry out further explorations on this topic.