**Responses to comments_RC1**

Dear Referee,

We are so thankful for your overall positive feedback on the manuscript and for the important comments and suggestions. We have therefore addressed them as follow:

**Specific comments**

**Comment1:** Section 3.3.3 This section describes the hydrological model-derived soil moisture. I feel that maybe the manuscript can benefit from a few more words about the model and its calibration. Perhaps just 5-10 lines may be sufficient, as I understand that you do not want to break too much the flow of the manuscript. Otherwise go for an appendix/supplementary material.

**Response:** A short paragraph about the Wflow model and its calibration will be added in Section 3.3.3 of the final version of the manuscript

**Comment 2:** Equation 7. I really appreciate the approach of normalizing the soil water content to make comparisons between models and observations. However more details should be given on this: which values of theta_max and theta_min have been found for the various soil moisture products?

**Response:** We agree that the normalization of soil water content $\boldsymbol{\theta}$ (theta) was made for easy comparison of the observed, model-derived and satellite-based soil moisture products. However, for all compared soil moisture products, the $\boldsymbol{\theta_{max}}$ and $\boldsymbol{\theta_{min}}$ were 1 and 0 respectively which led to almost similar values of Se (effective soil moisture) and $\boldsymbol{\theta}$ (actual soil water content). We will add such information in Section 3.4.1 of the manuscript for clarification.

**Comment 3:** Section 3.1 The landslide inventory is made of 32 useful landslides. These are a bit few (see analyses in https://link.springer.com/article/10.1007/s10346-021-01704-7). A comment on this may be added. However, for the manuscript this is not a big issue as it focuses on Rwanda which is an area for which only a few studies exist.

**Response:** This short paragraph on the constraints related to the used small sized sample (32 hazardous landslides events) will be added to Section 4.2.4. "Ideally one would have a landslide inventory of about 200 landslides events in order to have a precise estimation of threshold parameters (Peres and Cancelliere, 2021). However, the landslide inventory used for this study counts for only 32 hazardous landslides. Although, the reliance on this limited sample size is likely to lead to a bias towards the larger landslide events and those with impact to society, this landslide inventory is the most comprehensive currently available in the study area".

**Comment 4:** Section 4.2.3 : A comment on the limitations of the analysis related to the constraint of using a bilinear threshold form may be added (see e.g., https://www.mdpi.com/2073-4441/13/13/1752/htm, where other forms are suggested).

**Response**: The limitations and constraints of using the bilinear format have been shortly presented in Section 4.2.4. However, additional comment on the constraint of using a bilinear threshold will be added in Section 4.2.4 referring to Conrad et al., (2021).

**Minor comments/technical corrections**

**Comment 1:** LL 364-365 This is unclear: I imagine that the critical level for landslide occurrence is sort of fixed and then it is reached more or less easily based on the prior rainfall and the time lag.

**Response**: It is true that the critical level for landslide occurrence is more or less fixed when other geological and geomorphological condition are kept constant and it is reached more or less easily depending on the prior rainfall expressed in terms of antecedent soil moisture and the time lag between the landslide triggering rainfall and the soil hydrological response. LL 364-365 will be paraphrased accordingly.

**Comment 2:** LL 500 The authors apply a threshold of 10 mm on satellite products to make them better agree with observations. This is a sort of a "bias correction", about which a lot of literature exist. Perhaps make a fast literature review and add some references. (This could be useful also for future work)

**Response**: The threshold definition of a rainy day (10 mm) improved the similarities between the satellite-based and gauge-based landslide thresholds and thus considered as a bias correction between the two sources of rainfall data. Similarly, bias correction methods were adopted by other researchers to ensure for the high accuracy between ground- and satellite-based rainfall data (Bhatti et al., 2016; Vernimmen et al., 2012). This paragraph and references will be added to LL 500

**Comment 3:** Fig. 1 is perhaps a little bit messy (especially in B/W).

**Response**: Fig 1 will be improved to ensure for a better visibility and readability.

**Comment 4:** LL 512 the authors write "inter-event time" as the minimum dry interval between rainfall events. Perhaps add "minimum", even if I understand that IET is aligned with previous literature in the field of landslides.

**Response**: The word "minimum" will be added and thus "minimum inter-event time"

**Comment:** L120 a ")" is missing after Mukungwa.

**Response**: a ")" will be added after Mukungwa

**Comment:** L586 thus "can be" (?) very useful (something is missing in the sentence).

**Response**: The word "can be " will be added "thus can be very useful for landslide hazard…."

**Responses to comments_RC2**

Dear Referee,

Thank you for your feedback on our manuscript. The raised comments and suggestions are of great value for the improvement of the manuscript. We have considered all as follow:

**Comment:** The title contains the term "landslide hazard", which is misleading if referred to the manuscript contents. Hazard is generally intended to be an "off line" property of a territory, which is a function of susceptibility, temporal and magnitude (size) probability.

**Response:** The word "Hazard assessment " will be removed from the title to fit with the manuscript content as suggested. The Title will be "Potential of satellite-derived hydro-meteorological information for landslide initiation thresholds in Rwanda".

**Comment:** Regarding the discussion of threshold results, in my experience, false negatives (i.e., missed alarms) are more important than false positives (i.e., false alarms), because the consequences of missed alarms (e.g. deaths and injuries) are certainly more severe than those caused by false alarms (e.g., the unnecessary evacuation of a school). Therefore, looking at the number of missed alarms in Fig. 8b and 8e, I would not rely too heavily on these thresholds in a LEWS (Landslide Early Warning System). I suggest you to review the Discussion and Conclusions.

**Response:** We agree that the consequences of offering false alarms (FPR) are less harmful on the short-term than missed alarms (FNR) which implies that the best threshold should maximize the rate of true positives TPR (true alarms) while minimizing the FNR. However, the thresholds in Fig. 8b and 8e are classical thresholds ED relying exclusively on rainfall (Trigger), leading to the high rate of missed alarms and thus less important for a robust LEWS development. Similar to this study, previous studies (Bogaard and Greco, 2018; Peres et al., 2018, Marino et al., 2020; Thomas et al., 2020; Zhuo et al., 2019, Mirus et al., 2018a; Thomas et al., 2019; Uwihirwe et al., 2020, 2021) indicated that the consideration of the prior subsurface hydrological conditions reduce the number of missed alarms FNR as well as the number of false alarms FPR relative to the exclusive use of rainfall-only thresholds. In Figure 8a, 8c and 8d, 9a, 9b and 9c, we integrated the hydrological information (i.e. antecedent soil moisture) in landslide thresholds to improve the rate of TPR and reduce the rate of FNR and FPR. The main goal of hydro-meteorological thresholds (Cause-trigger) is to maximize the rate of true positives TPR (true alarms) i.e. minimize the FNR but at the same time reducing the rate of false positives FPR (False alarms). The used statistical metrics (TSS and RAD) are also in line with this concept aiming at maximizing the rate of true positives TPR while minimizing the rate of false positives FPR. Once TPR is maximized, the FNR is also minimized though difficult and or impossible to have a perfect threshold model with zero FNR and FPR. We will add a discussion point about this information in Section 4.2.4 and in conclusion part. We will also correct the number of false negative (FNR) in Figure 8f and 9d.

**Minor revisions**

**Comment:** Figs. 1, 2, 3, and 4 are similar and repetitive, and the 5-km buffers locally obscure the information from high-granularity maps in Figs. 1 and 2. For a better readability, my suggestion is to merge Figures (perhaps Fig. 1 with Fig. 3, and Fig. 2 with Fig.4) selecting two maps (perhaps elevation and geomorphology), and grouping sensor and landslide information, in Fig.2-4. By the way, I think that the year of occurrence in not that important for this analysis given the relatively low number of failures. Other information on the landslide sites (mean terrain slope in ROIs, aquifer type) could be provided in a Table with the list of failures.

**Response:** We will merge Fig.1&2 containing quite similar information and keep Fig. 3&4 showing different extent of our study area (ROIs) to keep the flow of the methodology. The mean terrain slope (Map) will only be kept in text and be removed in Fig 2.

**Comment:** I suggest using "cumulated event rainfall" for E, "event duration" for D, and "rainfall mean intensity" for I.

**Response:** We will replace " rainfall event volume E" by "cumulated event rainfall E" ; and "event intensity E/D" by "rainfall mean intensity E/D" as Suggested. The "event duration" for D is same as suggested and will be kept.

**Comment:** Tables 4 and 5 are a bit confusing. If I understand correctly, the first group of 5 columns refers to the whole landslide area, while the second group only refers to the modeled catchments. If so, please amend the tables accordingly.

**Response:** We will amend the Tables' captions accordingly even though same explanation have been provided in the footnotes of Table 4 and Table 5.

**Comment:** In Section 4.2.3 on hydro-meteorological thresholds, I would suggest to calculate also classical ED thresholds, which could provide competitive skill scores.

**Response:** In Figure 8b, 8e and 8f the classical ED thresholds are presented despite the weak prediction capability (TPR=50%) and weak skill score.

**Comment:** Figs. 8 and 9. You should improve the quality, especially that of 8d and 9d. Please, also avoid too small characters in the legend a, and instead use the caption to explain symbols and colours.

**Response:** We will improve the quality of Figure 8 and 9 by increasing the font size of the legend and explaining some of symbols in the Figures' captions.