We express our gratitude to the Reviewer, which pointed out some weaknesses of the manuscript and gave us several insights to improve it.
In the following, we provide a point-to-point reply (AA – authors' answers) to every referee comment (RC).

-
General comments

RC – The paper briefly communicates the improvement of a previous version of a landslide early warning decision tree (SIGMA) by adding soil moisture information. Two separate methodologies are presented. The first consists in cutting-off the application of SIGMA if mean daily soil moisture (MSM) averaged on the given Territorial Unit (TU) is below a threshold value. The second uses the time series of soil moisture measured at a point within the decision tree of SIGMA. The topic fits within the scope of NHESS. The paper is globally well-written, though language is improvable. However, I have some concerns about the real improvement obtained by using soil moisture information, and I think that the authors should prove the improvement by more in-depth tests. In particular, the authors should address the following points:

AA – During the revision stage, we will improve the writing and we will address all the points mentioned by the Referee.

-

RC – • As far as I understand, MSM is available from TOPKAPI for all (or most of) the 25 TUs. Why the authors apply it only to 7 selected TUs? This could be an ad hoc choice to make the methodology work well.

AA – A few words to explain the difference in the test sites between the two experiments (MSM experiment and Sigma-U experiment): during the first stage of the research we had at our disposal only soil moisture data from 7 TUs (years 2009-2014). There, we tested the MSM approach. Results were deemed encouraging, therefore when we obtained an increased dataset of soil moisture data (7 more TUs, but limited to the years 2011-2014) we directly developed and tested a more elaborate approach (the Sigma-U experiment).
Since in this work we are describing two distinct experiments, we decided to use two distinct datasets, related to test sites of different extension.
Concerning this issue, we ask the suggestion also from the Handling Editor: we have the raw data and we could extend the back-analysis also to the additional 7 TUs, but in there, the timespan would be limited to the years 2011-2014.

-

RC – • Soil moisture measured at an arbitrary point (where are the punctual measurements located?), may be totally unrelated to soil moisture at landslide locations. Hence the improvement showed by the authors may be just a case. For a

more robust testing, the authors should apply some sort of “jack-knife” validation test.

AA – Unfortunately, a misunderstanding occurred. We do not use measurements. In the manuscript, we were very careful to use the term “punctual estimates”, as values are not actually measured (e.g. by instruments or monitoring stations): they are estimated by TOPKAPI model. We use “punctual” to stress that we are not performing a distributed assessment [e.g. on a pixel-by-pixel basis]: since the original EWS uses only a rainfall measuring station for each territorial unit, we need only a soil moisture value for each territorial unit. That’s a value averaged for the whole TU, consistently with the “reference rain gauge” approach in which a rain gauge provides a rainfall value considered representative for a whole territorial unit. This could be clarified in the “materials and method” section, which could be edited as follows:
“For most of the hydrographic basins of the region, ARPAE-ER (…) provides the mean soil moisture value at hourly time step. These are values estimated by the TOPKAPI (…) model (…), which is a rainfall-runoff model that can provide high resolution hydrological information. We use these data to estimate the mean daily soil moisture (MSM) value for each TU…”

Please, consider also that we need to use only data readily available online to be used in real time in the EWS. ARPAE-ER does not provided distributed soil moisture data, it provides aggregated soil moisture data and they are just what we need for our objectives.

-

RC – Another point is that I do not see the rationale of considering the standard deviation of a random variate as an indication of its magnitude. The standard deviation is a measure of dispersion. The magnitude could be rather expressed by comparing the difference between the value and the mean with the standard deviation.

AA – Maybe we over-simplified the description of the original model SIGMA. In the revised version of the manuscript, more space will be devoted to the description of SIGMA approach and the passage from daily time series to sigma curves.

-

RC – For the reasons above I think that this brief communication should undergo major revisions before its publication.

AA – We thank the referee for the constructive comments, we will work hard on the revised version of the manuscript.

-

Specific comments

RC – P3 from L18. “A back analysis. . .”. Why only 7 TUs are used for the test?

AA – As we explained in response to a previous comment, the MSM experiment was performed on the first dataset we had at our disposal: 7 TUs, years 2009-2014. The SigmaU experiment was performed on 7+7 TUs, years 2011-2014.

-

RC – P3 L19 “from 320 to 231” these numbers differ from those in table 1. That’s okay because, as far as I understand, the number of TUs considered is different in the two cases. Maybe the authors should explain better this point.

AA – The main difference is not the TU number, it is that we are making comparisons between different approaches. The text highlighted in the referee comment (P3 L19) is placed in section 3.1 and it is about the difference between Sigma and the MSM approach. Table 1 is referred to section 3.2 and it is about the difference between Sigma and the Sigma-U approach. Since we are comparing Sigma with two different approaches, it is normal that numbers are different. We believe that the misunderstanding arose because figures and tables are listed at the end of the manuscript. In the final edited paper the table will be placed at the right point in the text and we think that will be sufficient to avoid misunderstandings.

-

RC – P4 L3: I understand that the SIGMA model has already been published by the authors, but the rationale of using standard deviation is not clear. The authors should possibly explain better this point. (See general comments).

AA – Maybe we over-simplified the description of the original model. In the revised version of the manuscript, more space will be devoted to the description of the SIGMA approach and the passage from daily time series to sigma curves.

-

RC – P4 from L14 “The results of the back-analysis clearly show an overall improvement. . .” The authors should apply a more in-depth test for assessing that the performances truly improve, by applying a “jack-knife”/”leave one out” validation test. This consists in the following: a) calibrate the decision tree based on all rainfall events except one (left-out); b) test the performance of the calibrated decision tree on the rainfall event left-out; c) repeat steps a) and b) until all rainfall events are covered as left-outs, d) summarize the results (e.g. by ROC indices) of all the left-outs. This may be done for all TUs. Other similar validation tests may be applied (See e.g. Haykin, 1997).

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice Hall, Upper Saddle River, New Jersey.

AA – In this manuscript, we use a different method, more simple and more straightforward than suggested by the Referee, but still we believe it could provide a rigorous quantitative assessment of the performances/improvements. We formulate a hypothesis (EWS can be enhanced by substituting antecedent rainfall with soil moisture) and we empirically verify that it is met in our testing dataset. We also shows basic statistics (count of hits and errors). As we stated in the conclusion, before actual implementation in the EWS, additional data should be gathered for a more robust calibration should, possibly including one ad-hoc threshold for each TU (and not the same threshold for the whole region). At that time, the approach suggested by the Referee will provide a valuable contribution. However, we agree with the reviewer that the sentence "*The results of the back-analysis clearly show an overall improvement*" is too "definitive" and would need more robust support. Therefore, we believe that the text should be modified with something like "*The results of the back-analysis are encouraging, as the count of both false alarms and missed alarms is lower in SIGMA-U than in SIGMA*".

-

RC – P1 L17 Possibly update references on landslide triggering thresholds by adding, e.g.: Peruccacci et al, 2017; Peres and Cancelliere, 2016; Leonarduzzi et al., 2017.
Leonarduzzi, E., Molnar, P. and Mcardell, B. W.: Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data, doi:10.1002/2017WR021044, 2017.
Peres, D. J. and Cancelliere, A.: Estimating return period of landslide triggering by Monte Carlo simulation, J. Hydrol., doi:10.1016/j.jhydrol.2016.03.036, 2016.
Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M. and Guzzetti, F.: Rainfall thresholds for possible landslide occurrence in Italy, Geomorphology, 290, 39–57, doi:10.1016/j.geomorph.2017.03.031, 2017.

AA – Thank you for the suggestion, we will perform an update to the references cited in the manuscript. Please note that the "rainfall thresholds" topic is very broad, therefore we focused the introduction on literature thresholds that are implemented into EWS and on thresholds considering antecedent rainfall as a proxy for soil moisture conditions. We were forced to a very strict focus because the manuscript typology demand a limitation to max. 20 references. We will carefully go through the suggested papers and include them in the introduction, when appropriate.

-

RC – Perhaps the introduction may take into account that the importance of including soil moisture information in landslide triggering thresholds has been stressed by a recent NHESS invited perspective by Bogaard and Greco, 2017.
Bogaard, T. and Greco, R.: Invited perspectives. A hydrological look to precipitation intensity duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Nat. Hazards Earth Syst. Sci. Discuss., 1–17,doi:10.5194/nhess-2017-

241, 2017.

AA – Indeed, the Referee suggests a very interesting article. We will make reference to it in the introduction and also in the conclusion, since we believe that our works expands by a small step the classical rainfall threshold approach towards the direction expressed by Bogaard and Greco: instead of using only rainfall, we try to better encompass the hydrology of the territorial units by using soil moisture. On this regard, we will also make reference to a work very recently submitted to the same special issue by Kanli et al. (2017), which shares a similar perspective.
Canli, E., Mergili, M., and Glade, T.: Probabilistic landslide ensemble prediction systems: Lessons to be learned from hydrology, Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2017-427, in review, 2017.
We were not aware of these papers when we submitted the first version of our manuscript.

-

RC – Tab. 1: also the number of landslides and true positives and negatives should be shown, and commented in the text

AA – In the revised version of the manuscript, we will expand the table as suggested and provide the necessary comments in the text.
While performing calculations about true positives, we noticed an error: the total number of landslides (hits+missed) was not the same in SIGMA and SIGMA-U. After a thorough check of the used spreadsheet, we identified an error in the formulas: in a few words, 5 TUs were erroneously not included in the calculations for Table 1. We corrected the formulas and the links and re-calculated the statistics, which now result even better than the mistaken ones reported in the previous version of the manuscript:
"…*false alarms issued at warning level 1, which are negligible, decreased by 8%, while the very important warning level 3 was erroneously issued 11 times instead of 21 (-48%). False alarms at the intermediate warning level 2 were reduced from 287 to 197 (-31%). Missed alarms are reduced as well: while SIGMA missed 88 alarms, SIGMA-U missed 69 alarms (-22%). This corresponds to a total of 134 missed landslides instead of 214 (-37%). Overall, SIGMA-U hits 789 landslides out of 923 (85.5%), outperforming SIGMA that hits 709 landslides (76.8%)."*
We apologize for the error and we express our gratitude to the Referee that made it possible to notice it and to correct it.

-

RC – Technical corrections
P1 L12 maybe replace "were" with "are"
P1 L22 "thresholds" instead of "threshold"
P1 L23 "landslide occurrence" instead of "the landslide occurrence"
P1 L25 "as a proxy" instead of "a proxy"
P2 L2 "landslide" instead of "the landslide"

P2 L29 here introduce the acronym MSM
P2 L27 “rainfall-runoff” instead of “inflow-outflow”
P4 L14 “importantly” instead of “important”
P4 L29 “is by large” maybe can be improved

AA – All suggested corrections have been included in the revised text.

-
RC – Fig. 1: Where soil punctual measurements were taken?

AA – As explained before, we do not use measurements. In the manuscript, we use the term “punctual estimates”. Values are not actually measured, they are estimated by TOPKAPI model. We use “punctual” to stress that we are not performing a distributed assessment [e.g. on a pixel-by-pixel basis]: since the original EWS uses only a rainfall measuring station for each territorial unit, we need only a soil moisture value for each territorial unit.
In the conclusions section, we only hypothesize the possibility of using actual measures in the future developments of the research (of course, provided the funds are renewed and the research plan is approved).
Maybe in the introduction a sentence could be misleading *(“This work explores the possibility to exploit punctual soil moisture values estimated at few discrete points and to correlate them with landslide triggering over wide areas”*). We could change it to avoid misunderstandings *(“This work explores the possibility to exploit punctual soil moisture estimates and to correlate them with landslide triggering over wide areas”*). Maybe, the term “*aggregated*” could be used instead of “*punctual*”, if Editor and Referees consider it clearer.

-

RC – Fig. 2: on the upper-left: there must be a mistake in the orientation of the arrows

AA – Thank you for identifying this issue. The figure was adjusted.

-

On behalf of all authors,
Samuele Segoni