GENERAL COMMENT

The contribution "Use of delayed ERA5-Land soil moisture products for improving landslide early warning" by N. Palazzolo and co-Authors is interesting and addresses questions relevant with the scope of NHESS. In this paper the researchers conducted a series of direct experiments. They aimed to determine how effective it is to use antecedent soil moisture data (ranging from 0 to 15 days old) from a global reanalysis model, combined with rainfall data, to predict landslide triggers using machine learning. As one might expect, the findings indicate that this historical soil moisture data enhances model performance, though its benefit diminishes as the time lag increases. Results, and discussion sections are short compared to amount of work done. They should be increased. The theoretical background is well-argued but not complete. Review of literature seems completed. The description of study area is sufficiently complete. The description of methodology and successive parts of paper are well organized but not complete. The readability of the whole paper is good with a good English. In general, the synthetic approach of the study is very clear, but I think that going into more detail on several aspects (later reported) could be useful for the readers. It can be published on NHESS journal only after minor revision.

SPECIFIC COMMENTS

I have some specific comments that should be addressed before the manuscript can be accepted for publication in NHESS journal.

1 Introduction (Line 30-40)

My main comment regards the conceptual background of the comparing of the performance of different models. It would be necessary to better understand from the authors whether it is possible to directly compare the performance of the models with and without soil moisture (on the empirical rainfall thresholds) information based exclusively on the comparison with the TSS index or whether other strategies exist in this direction.

2 Material and methods

(Line 48-49) Have you tried using landslide data from other catalogues? (i.e. Peruccacci et al., 2023)

(Line 55-60) The Authors reconstruct triggering and non-triggering rainfall events.

I didn't understand the following list of things:

The rainfall event or landslide triggering condition (MPRC automatically reconstructed by the tool) normally represents a subset of the entire rainfall event while non-triggering rainfall events are always considered as a whole. Did the authors take this difference into account when creating the input data for the neural network? How many triggering conditions coincide with the entire rainfall event? If not, how many of them differ significantly in terms of duration and cumulative rainfall, and what justifies the choice of these two sets of triggered and non-triggered rainfall events? The authors should provide a better argument and commentary on the method used.

The number of reconstructed non-triggering rainfall events is at least two orders of magnitude greater than those that trigger landslides. What procedure was used to account for this substantial difference? Are the samples used for the training, validation and testing phases balanced or unbalanced? What observations are used in this last case, and how is sampling performed? Have statistical tests been conducted to confirm whether these samples are representative of the population? Please specify the setting more precisely.

2.3 Artificial Neural Network models (ANNs)

(Line 80-90) ANNs Input variables/data

The reconstruction of the rainfall conditions that triggered landslides is performed starting from the assumption that in the place where the landslide occurs the rainfall is the same as that measured on a representative rain gauge (specific criteria dependent on a parameterized variable that considers the distance was used by CTRL-T tool). As regards the choice of cells that contains information associated with the soil moisture status, it seems that the cell that includes the representative rain gauge chosen automatically by the tool has been considered.

Under this hypothesis (soil moisture associated with the cell that includes rain gauge) what considerations can be made from a physical point of view? Is it conceptually correct to look for a relationship between the saturation state of a soil different from that in which the triggering of a landslide occurs? If the cell that contains the landslide is considered, do the results change? Considering the spatial resolution of the ERA5-Land product (about 9km x 9km) probably in most cases the cell containing the landslide or the rain gauge is the same, but I believe that deepening this aspect by making descriptive statistics could be useful. Moreover, for an operational use, the forecast value returned as output by the model would describe the possibility that a landslide could happen in that specific cell for which the soil moisture value was provided as input. Therefore, in my opinion, it should perhaps also be used in the model's learning phase.

3 Study area and data

(Line 140-150) CTRL-T parameters setting

What is the length of the “warm” spring–summer period (C_W) and “cold” autumn–winter period (C_C). More in detail, what values were assigned to the variables sws (beginning of warm season) and ews (end of warm season), i.e. C_W from May to October and C_C from November to April correspond sws=5 (May), ews=10 (October). Please specify the setting more precisely.

Add the *sws* and *ews* parameters in the Table 1 (cfr. comment refered to Line 140-150)

4 Results and discussion

(Line 160-165)

Figure 4 shows box plots of the results for the autocorrelation conditions. The text provides a good description of the behaviour of the variations in terms of lag k. However, a visual analysis of the ρ(k) values (k = 1, 2, ..., 15 days) in the different layer depth cases (b) 7–28 cm, (c) 28–100 cm, and (d) 100–239 cm reveals that the values are indistinguishable (perhaps due to the scale). Consequently, it is unclear why layer 2 was selected for subsequent analysis (b). Please explain better.

Line 175-190

How does performance change for all combinations of layer use? For example, what is the maximum TSS (mean) value for D-E-S1, D-E-S3, D-E-S4, D-E-S all-1, and so on? Has a systematic analysis been conducted to better understand the effect of each input data component introduced in network training? If it is not very time-consuming, I would suggest trying this approach.

Figures description

Description of Figure 2

In the structural diagram of the ANN, the symbol for dichotomy on the third layer is different from those for the three preceding nodes on layer 2.

I suggest to:

remove “---and---” because it’s implicit in the figure

replace “depth” with “cumulated rainfall” (also extended to the text)

Specify the terms S (ERA5-Land depht)

Specify which is the input, hidden and output level.

Description of Figure 3

Add geographical grid and coordinate labels.

I suggest to change the colour filling of the peninsular part of Italy (i.e. the monochrome palette) because the scale is different respect to the study area.

I suggest increasing the size of the dots and changing the starting colour of palette from “blue” to “green” for greater contrast.

Add the description of the symbols in the caption.

Description of Figure 4

Formatting borders, label text, and ticks in black instead of grey

Add the description of the continuous red and the dashed grey curves in the caption (also extended to the text).

Remove the labels and headings from the x-axis for (a) and (b) and from the y-axis for (b) and (d). This should make the figure clearer by increasing its size.

Description of Figure 5

Formatting borders, label text, and ticks in black instead of grey

Change dimensions of ‘(a)’ and ‘(b)’ as in figure 4.

TECHNICAL CORRECTIONS

Below a list of some more detailed comments and suggestions referred to specific parts of the text.

(Line 54): replace "(*D*, in hours)" with "*D* (h)"

(Line 55): replace "(*E*, in millimeters)" with "*E* (mm)"

(Line 116): remove the space in "non-triggering"

(Line 120): remove the space in "(missed alarms)"

(Line 123): Change the 'T' term format

(Line 126): enter the comma in the range [0,1]

(Line 129): replace "southern Italy" with "Southern Italy"

(Line 31): replace "700" with "700 mm"

REFERENCES:

Peruccacci S., Gariano S.L., Melillo M., Solimano M., Guzzetti F., Brunetti M.T. (2023) The ITAlian rainfall-induced LandslIdes CAtalogue, an extensive and accurate spatio-temporal catalogue of rainfall-induced landslides in Italy. Earth Syst Sci Data, 15, 2863–2877, doi: 10.5194/essd-15-2863-2023