Revision of
Landslide Susceptibility Assessment Tools v1.0.0b - Project Manager
Suite: A new modular toolkit for landslide susceptibility assessment.
(Jewgenij Torizin, Nick Schüßler, Michael Fuchs)

General Comments

The manuscript under revision presents a standalone software dedicated to the landslide susceptibility models' preparation and evaluation, as well as to produce susceptibility maps.
The topic is well introduced and justification is well contextualized. The objectives are:

1) To introduce LSAT PM
2) To provide easy access to a selection of state-of-the-art methods representing groups of different approaches such as weights of evidence, logistic regression, artificial neural networks and analytical hierarchy process.
3) To share the experience in the implementation of heuristic and data-driven landslide susceptibility assessment methods.

In general, I find the tool really interesting and I acknowledge the tremendous work that is behind its implementation. However, the manuscript needs more emphasis in the detailed description of the functionalities or tools offered by the software. Many important aspects of the modelling process are not clear. Also, I would focus the examples shown in section 5 to support the software description, instead of detailed discussions about the results.
The conclusion section should be improved.
In addition, although I'm not a native English speaker, I believe the text can be improved. I suggest to submit the new version of the manuscript revised by an English speaking person.
I think this work deserves to be published, but before that the authors should address some important specific questions that are listed in the attached PDF file.

Specific questions

In my opinion, the popularity gained by the data-driven methods is not related only to the advances in remote sensing (*Line 34*), but also, to the increased data availability (landslide inventories and digital layers of thematic maps) and accessibility to statistical packages.

I don't understand why you say that "LSAT PM's core is the weights of evidence (WoE) method" (*Line 71*). I understand that your suggested workflow begins with WoE method for exploration of the data and then the application of other methods for comparison. But, with this sentence it seems that WoE must be run before any other method. Is it true?

Vector inputs are limited to the inventory? How can I manage a vector Geologic map in LSAT PM? As far as I could test the software, the environmental factors can be only input as raster. In this case, you should clearly specify it. And also, if this is the case, why do we need Geology and Land Use vector maps in the test dataset?

In Table 1 caption, I would change "tools" by "functionalities" or something more generic. In the end, it describes the GUI panels organisation, where some of them are tools.
Also, not all the items in the table can be found in the content bar: Context Menu, Reproject and Model Info.

Figure 1 should be improved. At least, letter "a" and the north arrow are not visible in Fig. 1a; Coordinates can be fixed better (latitude numbers vertically oriented and without exiting the margins); Add a location map. Also, the names of the geological formation without a short description are not useful for the reader.

After *Line 160* I would specify the different splitting options for the inventories (random, spatial, temporal), in case the software allows to do it.

In *Lines 168-170* you say that "vector data are unsuitable for spatial analysis", and I don't agree. Maybe, linear and point-like vector data can be unsuitable, but a land-use vector map is fully suitable for spatial analysis, in my opinion. Please re-formulate this part.

In *Lines 180 -182,* why do you say that "Contingency analysis is the only tool in the tool domain Raster data"? According to Tab. 1 and Fig. 3, there are other tools (Euclidean distance, Combine …). I would reformulate the complete paragraph starting with something like "The contingency analysis tool helps to explore ..."

In section 4, specify and describe better the data requirements and outputs obtained at each step (contingency table, result tables). Which specific information they contain?

I would remove *Lines 244-245*, because at this point the model builder is not introduced yet, and they confuse a little bit.

In section 4.5 the explanation about the sampling error assessment needs more details. Provide details on how it can be done using LSAT PM.

When is it used the validation sample? After having tested the software, I understood that in the Model Builder module you have the option to generate the ROC curve respect to the desired inventory partition (training, test or even a group of subsamples). This is a crucial step of the evaluation that is not clearly explained in the manuscript. Please, improve this part.

The Zoning module is used to reclassify the susceptibility maps in few, and more understandable, susceptibility classes. According to your description in section 4.6, it seems that you follow the approach of Chung and Fabbri (2003). However, this approach is not based in ROC curve, but in the prediction and success rate curves, which are completely different things. I believe that if you want to set the classes in a way that you can ensure the proportions of landslide areas that should fall within the specific zone, then a prediction rate curve has to be used, and not a ROC curve. I did some tests and I realized that the curve prepared with the Model Builder and the curve prepared by the Zoning module are identical. In my opinion, this has to be fixed before publishing the software.
In addition, if you suggest some classification thresholds by default (50; 30; 15; 4; 1), you should explain more in detail the implications of this values in the interpretation of the maps. Because, the suitability of such thresholds can be discussed.

The current section 5 should be reorganized. I would include the test data description in this section. Then, it should be just a section where the potential of the tools explained in section 4 are illustrated. The procedures to build the models should be described very briefly (maybe using tables or conceptual plots/figures) and making reference to the section 4, where more detailed explanations can be found. In short, this section should give examples of (i) what we can get as outputs and (ii) how we should interpret them. In this regard, I believe that some interesting outputs are missing, such as susceptibility maps or variables evaluation reports and contingency tables.

In section 6, which is the difference between hybrid model and model ensemble? Did you perform hybrid models in section 5?

In general, I find the conclusion section a little bit incomplete. Before going through the future implementations that are planned, I think that a real recap summary is missing. Something more detailed than only three lines.

Technical corrections

Line 16 <u>Rephrase</u> "not included in" instead of "not created with".
Line 17-18 <u>Rephrase</u> "The software distribution package is subject to continuous further development
    and is provided with comprehensive documentation as well as a dataset for testing purposes"
    instead of "The software distribution package includes comprehensive documentation. A dataset for

testing purposes of the software is available. LSAT PM is subject to continuous further development".

Line 20 <u>Rephrase</u> "Landslides occur ..." instead of "Gravitational mass movements or, more general, landslides occur ...".

Line 27-28 <u>Add text</u> Specify why is more challenging to address the temporal domain (scarcity of data, spatial heterogeneity …)

Line 30-33 <u>Review References</u> Reichenbach et al 2018 refers only to statistically driven methods. I suggest to add some updated references for physically based methods.

Line 32 <u>Rephrase</u> "However, based on the …"

Line 37-39 <u>Rephrase</u> "In the case of LSAs, the classifier's task is to distinguish whether a specific countable element in a study area (e.g., raster pixel or unique condition unit) is a landslide or non-landslide, based on available features." instead of "The classifier's task is, in a narrower sense, to distinguish based on available features whether a specific countable element in a study area (e.g., raster pixel or unique condition unit) is a landslide or non-landslide."

Line 63 <u>Add text</u> "… standalone and easy-to-use…" instead of "… standalone, easy-to-use…"

Line 64-66 <u>Rephrase</u> "We want to highlight such methods' capabilities and limitations, making them more transparent and providing a convenient framework for model evaluation and uncertainty assessment" instead of "Providing a convenient framework for model evaluation and uncertainty assessment, we want to highlight such methods' capabilities and limitations, making them more transparent."

Tab 1 <u>Correct</u> Logistic Regression (LR) instead of (ANN)

Line 115 <u>Rephrase</u> "Separately from LSAT PM package, a dataset to test the functionalities of the software is available in ..." instead of "A dataset to test the functionalities of the software is available separately from ..."

Line 121 <u>Rephrase</u> "test study area" instead of "test data area"

Line 177-179 <u>Rephrase</u> "Thus, it is …" I'm not sure if I understood this sentence. Please try to explain in better.

Line *190* deception? Maybe perception is better?

Line 213 "as" instead of "for"

Line 236 remove comma

Line 271-274 remove the last two sentences.

Line 279 <u>Rephrase</u> "The model evaluation is based on the …" instead of "The model evaluation is basing on the"

References has to be revised in format and content

Polemio and Petrucci, 2000 is not in the reference list

Petley, D. (2012) Global patterns of loss of life from landslides. Geology 40: 927–930, https://doi.org/10.1130/G33217.1. is not in the right format.

Balzer et al., 2020 is not in the reference list

Makkawi et al., 2006 is not in the reference list

Lombardo & Mai, 2018 - Lee and Evangelista chose between AND or &

…

Software user's manual

I tested the software in both windows and linux machines. The installation in windows was done by the *.exe installation without problems. However, as it is common, in linux (Ubuntu 20.04.3) I had to face some little issues. I describe them in the following just if you decide to include some specific guides in the manual user.

Even if I had already `pyhton3`, I had to install `python3-venv` and `python3-pip`

```
sudo apt install python3-venv
sudo apt install python3-pip
```

I had to give execution rights to the *activate* file

```
sudo chmod ugo+x venv/bin/activate
```

Then I was able to run the sequence in the user's manual

```
python3 -m venv venv
```

```
  ./venv/bin/activate    # instead of .\venv\Scripts\activate
  python3 -m pip install -r requirements.txt
    ERROR: launchpadlib 1.10.13 requires testresources

  python3 -m pip install testresources
  python3 -m pip install -r requirements.txt


  python startMenu_main.py
```

I already had *gdal* installed but in case, I believe something like the following command would be helpful as well
```
  sudo apt install gdal-bin python-gdal python3-gdal
```

There is no way to access the Documentation button. Why not to include a simple pdf file in the directory?