



EGUsphere, referee comment RC1  
<https://doi.org/10.5194/egusphere-2022-950-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on egusphere-2022-950**

Katy Burrows (Referee)

---

Referee comment on "Machine-learning-based nowcasting of the Vögelsberg deep-seated landslide: why predicting slow deformation is not so easy" by Adriaan L. van Natijne et al., EGU sphere, <https://doi.org/10.5194/egusphere-2022-950-RC1>, 2022

---

I have enjoyed reviewing this paper "Machine learning nowcasting of the Vogelsberg deep-seated landslide: why predicting slow deformation is not so easy", in which the authors attempt to nowcast deformation (and particularly accelerations) of a slow-moving landslide with a complicated triggering mechanism. Machine learning is well-suited to this problem and the authors have chosen to use a neural network technique (LSTM) which is a good choice for modelling the deformation of a coherent landslide that evolves through time (as opposed to fast landslide occurrence can be nowcasted using more simple regression techniques e.g. the fuzzy logic model used in LHASA rainfall-triggered landslide product or the logistic regression model used by the USGS to nowcast earthquake-triggered landslides).

The study is well-designed and carried out, I cannot identify any problems with the methodology (although I am not an expert in neural networks). The authors had mixed success in creating a nowcast model, but this paper represents a useful starting point for future development of such techniques and the current limitations are clearly discussed. It would clearly be a problem if someone writing a review on this topic in a few years could only find examples of landslides where nowcasting had been easy, so this kind of paper covering difficult case studies is important. The paper is quite long, but this is because of the difficulties encountered in the modelling (it is easier to describe a straightforward positive result briefly) so I think the length is justified. Overall the manuscript is clear and well-written, I have just a few comments and suggested technical corrections as follows:

### **Comments**

**Line 87** "Both will be introduced briefly" In this paper, you use the continuous type, I would specify that here to make the paper easier to follow

**Line 88-94 "2.1 Classification methods" / "Continuous models"** Could you label both sections as "Classification methods" "Continuous methods" or "Classification models" "Continuous models" so they match better?

**Line 157-158 "Furthermore the amplitude of the filtered signal lags behind the original deformation signal"** Why is this? Is it a side-effect of the filtering or have you done it deliberately?

**Line 209** Is API calculated from the ERA5 or GPM datasets? Does it make any difference which one you use?

**Figure 6** The shaded areas for lines 1-4 begin before the deformation measurements. Is this a mistake?

**Line 253** "do not lower the training loss" do you mean the loss function i.e. the MSE?

**Section 5, line 274-292** In the end, your model does not contain any snowmelt input, although you expected this to be relevant for the landslide. Could this be why your model only predicts the training data well in Summer and Autumn? Does a model including one of your snowmelt inputs (V3 or V4) predict Spring and Winter better (even if it's worse over all 4 seasons combined?)

Another thing that could adversely affect your model could be that small spatial scale rainfall events might not be captured by your satellite rainfall products, which have quite a coarse resolution (although unfortunately there would not really be any solution to this)

**Figure 7** Like my comment for figure 6, I wonder why the shaded area starts before your deformation dataset. Is this the "warm-up time" you describe in your figure caption? Or is that the part starting from early June 2016 where you have deformation data but no prediction? Maybe you could label this warm-up time on the time series with a box or shaded section?

**Section 6.1.1 Lines 360-365** Your R2 value (0.31) seems low, but interpreting a single R2 value is difficult. I'm not sure it's useful to include this metric when you have nothing to compare it to.

**Section 6.1.2** I think I would have put this subsection in your methods section as it contains similar information to Sections 4.1 and 4.2

**Lines 409-413** If you separate the two benchmarks in the model, would this result in them no longer being connected in space? (So one could accelerate independently of the other). I would have thought that since one part of a landslide moving is likely to destabilise another part, separating them would be a disadvantage.

Actually, I think I have misunderstood what you mean to say in these lines, can you find another way to write this?

**Lines 480-486** I would specify Sentinel-1 since the temporal resolution of SAR satellites varies

Also, it is not clear here whether you are suggesting the use of InSAR as an input variable for the model, or if you are suggesting that maybe for other landslides where you don't have such detailed deformation data, deformation time series derived from InSAR could be used to train a similar model.

**Lines 522** For landcover changes as an input, won't you run into the same problem of temporal resolution as you found in the SAR data? And if your landcover product was derived from e.g. Sentinel-2, it could actually be worse because of cloud cover.

**Lines 543-547** Here, with the EGMS product, it is based on Sentinel-1 data so you would only have a 12-day temporal resolution (Especially following the failure of Sentinel-1B), which would result in the same temporal resolution problem you discussed in Section 6.3.1

**Line 567** "Is not catastrophic" here, you mean the landslide does not undergo catastrophic failure. I think that in the conclusions, it might be better to actually spell this out (in case of people who are not completely familiar with the terminology reading the paper fast and skipping to the conclusions). So I would say "and does not undergo catastrophic failure" or something similar

**Table C1** This is a comprehensive table, but since not all the studies were included in the original table from Van Natijne (2020) there is no definition of some of the acronyms e.g. GRNNS

I would also put something like "See Table C1 for a summary of past work on the topic" in your first paragraph of Section 2 to direct the reader to the literature. When I first read this section, I was wondering where all the references were.

## Technical corrections

**Line 9** "such machine learning system" should be "such a machine learning system" or "such machine learning systems"

**Line 10** "Primary issue is" should be "The primary issue is"

**Line 21** "Such system" should be "Such systems"

**Line 33** "gradual, non-catastrophic, deformations" should be "gradual, non-catastrophic deformations" (No need for a comma after the last adjective in a list of adjectives, but I think in NHESS this kind of thing is corrected in typesetting)

**Line 36** "may only be used" it would be more correct here to say "can only be used"

**Line 43** "Unlike to catastrophic" should be "Unlike catastrophic"

**Line 44-45** "analysis of the monitoring data of deep-seated landslides are" should be "analysis of the monitoring data of deep-seated landslides is", the verb is associated with "analysis" not "data"

**Line 46** "relation" should be "relationship"

**Line 51** "Such data-driven model" should be "Such data-driven models"

**Line 60** "available, remotely-sensed, data" should be "available, remotely-sensed data"

**Line 76-78** "For example, by including recent observations... evaporation" this sentence is incomplete. You could make a complete sentence by joining it to the sentence before with a comma

**Line 96** "instant" it would be better to say "immediate"

**Line 124** "too few training data is provided" should be "too few training data are provided"

**Line 263** "the system is not capable to describe" should be "the system is not capable of describing"

**Line 405** "The southern, inhabited, part of the slope" should be "The southern, inhabited part of the slope"

**Line 406** "In contrast, the benchmark on the northern part of the slope, shows" should be "In contrast, the benchmark on the northern part of the slope shows" (No comma before shows)

**Line 551** "capable, model" should be "capable model"