I must first apologize to the authors and to the editor for this late review. As an author, I have myself experienced how irritating it can be to wait for an overdue review. I can only mention heavy load of work as an excuse for my delayed response.

Not being myself an expert on landslide causes and occurrences, I am not in a position to comment on the parts of the paper that specifically deal with those aspects, nor actually to evaluate the paper in comparison with what has already been done on estimation of landslide susceptibility and uncertainty. But I have comments on the methodological aspects of the paper, which may be useful for all readers.

I must say I have had difficulties to clearly understand what the authors have done, as concerns both their methodological approach and the validation of the results they have obtained. I will limit myself to what are the most important points I want to stress.

1. My first question has actually to do with landslides. The authors focus their study to *hydrologically triggered landslides* (l. 68). This means that they ignore, for instance, landslides triggered by earthquakes. What is the reason for that restriction ? How can the distinction be made between different kinds of landslides, once they have occurred ? And does the Global Landslide Catalog report only hydrologically triggered landslides ? These questions may look naïve to specialists, but some appropriate information (and references) may be useful to outsiders.

2. The first purpose of the paper is to derive *LSS model equations* (ll. 66-67). The authors do not actually show any equation that is explicitly identified as such. The model equations must be equations of form (1), where the quantity P(Y = 1) is what is called LSS elsewhere. Equations of form (1), which are defined as a form of logistic regression, are used on appropriate training sets for determining, through MELR and Cross Validation, values of the parameters α and β_i (i=1,…,n). This raises a number of questions.

a. What is the rationale for the logistic form of equation (1) ? What are the advantages of that specific approach ? It seems to me to be a rather arbitrary choice. In their conclusion, the authors mention *the choice of the statistical model* (l. 365) as one possible source of uncertainty in the whole estimation process. Do they refer there to Eq. (1) ? The authors give references concerning logistic regression, but some basic explanation would be useful.

b. In MELR, what is the criterion for quality ? Given a tentative set of values (α, β_i, i=1,…,n), by which measure is the corresponding fit to P(Y = 1) evaluated ? A simple quadratic fit, or what ?

c. I understand that the values P(Y = 1) in the training sets are taken in the data set built in subsection 2.1 from the GLC catalog, so that these values are restricted to 0 and 1 (absence or presence of landslides). It would *a priori* seem more appropriate to consider a quantity such as the frequency of occurrence of landslides (that would not be impossible from GLC since the latter mentions more than one landslide for a number of individual grid cells). That may not be practically possible, but it would in my mind be appropriate to mention, and preferably briefly discuss, that alternative approach.

3. *a*. Concerning the quality of the uncertainty of their LSS estimates, the authors use as diagnostic the *Receiver Operation Characteristic* (*ROC*) and the associated *area under the ROC curve* (*AUC*). They mention AUC values for individual members of ensembles, *i.e*. LSS maps (ll. 214-217, l. 275 and Fig. 7), as well as for global ensembles (ensembles of maps). The latter are all right for me, but I do not understand what AUC values for individual maps can be. ROC curves (https://en.wikipedia.org/wiki/Receiver_operating_characteristic) are parameterized by a threshold *T*, each point on the curve corresponding to a value of *T*. The corresponding coordinates are relative to the circumstances when the value of a given parameter is larger than *T* (in the context of the present paper, that parameter must be LSS). Unless all grid cells are lumped together, which does not seem to be reasonable, it does not make sense to consider the situation LSS > *T* on a single map. That makes sense only on an ensemble of maps, with grid cells being considered independently of each other. I may of course be mistaken as to what the authors have exactly done, but clarification is necessary.

b. And the reference to *one fully deterministic reference MELR equation (based on neither CV nor input perturbations)* (ll. 220-221) is confusing. Does it mean you have performed the validation on other outputs than the ones obtained from CV ? I have a similar question about the *one deterministic MELR equation* mentioned on ll. 355-356.

4. The authors write (ll. 357-359) *The finding of Kalnay et al. (2006) (show) that the introduction of ensembles increases the accuracy of the prediction does not hold for our LSS modelling. This is probably due to the non-linear characteristics of logistic regression and LSS being static*. I understand the authors mean that the accuracy of the mean of the output ensemble is higher than the accuracy of an individual deterministic estimate (at least statistically). From what I understand, non-linearity cannot be the problem here. Consider a process $F(x)$ where there is uncertainty of the input x. Let $\{x_i\}$ a sample of independent realizations of the probability distribution for x. The ensemble $\{F(x_i)\}$ is a sample of independent realizations of the probability distribution for $F(x)$. As such, the mean of that ensemble is the best estimate of $F(x)$, at least in a least-square sense. That is true whether the process F is linear or not.

5. The authors write in their conclusion (ll. 373-374) … *predictor variable perturbations results in a reliable assessment of the associated total prediction uncertainty*. It is of course more difficult to assess the uncertainty on an estimate than to obtain the estimate itself. But the authors' statement seems to be a bit of an exaggeration. The AUC values given in the paper do show some reliability in the assessment of the uncertainty, but no more. Actually, the amplitude of the predictor variable perturbations has been evaluated on the same data set as the LSS values. The whole process is therefore subject to some form of inbreeding, the impact of which is difficult to assess. And the authors write themselves *A comparison of $\sigma_{LSS2500}$ with independent global estimates is currently not possible for lack of uncertainty estimates* (ll. 340-341). I suggest the authors soften down their concluding statement.

I would have also comments on editing aspects of the paper, but I think they are of lesser importance at this stage.