The authors thank the reviewer for the constructive comments. The comments are shown in regular fonts, ***our responses are in bold italic, blue fonts***. *Changes made in the manuscript are printed in italic, underlined, blue fonts.* ***Our*** *line references* ***refer to the updated manuscript with track-changes.***

In addition to the answers below, we implemented a few minor editorial changes that are marked in the track changes.

Reviewer #1

The paper has in my opinion been significantly improved. But I also think the general approach taken by the authors is still not clearly described, and that improvement must still be done in the presentation of the paper.

My main comments are as follows. Again, I stress that I have no expertise whatsoever on the occurrence of landslides, and my comments bear exclusively on the methodological aspects and the presentation of the paper.

We thank the reviewer for the detailed feedback on the methods that help to improve the readability of the paper and understanding of our research even further.

The line numbers below refer to file nhess-2021-360-ATC1.pdf, *i.e.* the file containing the Author's Tracked Changes.

1. Although I think I have now basically understood what the authors have done, I still think the paper would be difficult to understand for a non-expert reader. For instance, I not found that Figure 1, which is meant to describe the general methodological approach of the paper, is really useful.

Thank you for this feedback. We have now revised the figure to only show the essential aspects of the approach followed in this study:

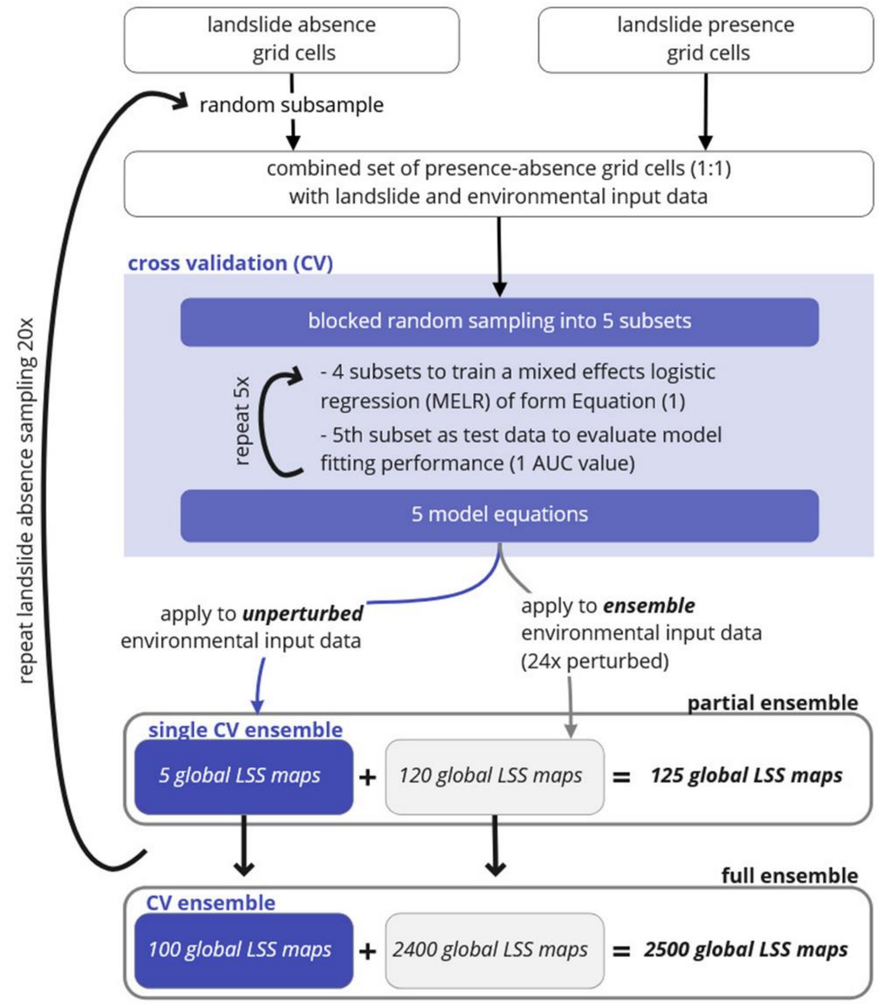

Figure 1 *Schematic of methodology used in this study to derive ensembles of global landslide susceptibility (LSS) maps. 'Ensemble' refers to a collection of LSS maps. In the course of this study, we refer to different subsets of the* ***full ensemble*** *(LSS_{2500}), namely the ensemble from one single blocked random CV application (**single CV ensemble**, LSS_5), when adding input perturbations to it (**partial ensemble**, LSS_{125}) or when repeating the underlying landslide absence subsampling (**CV ensemble**, LSS_{100}). Subscript numbers indicate the size of the LSS ensemble. Model fitting performance is evaluated during the process of CV by calculating the area under the Receiver Operating Characteristic curve (AUC) for each model equation.*

1a. The general methodological approach is described in Section 3, and particularly in subsection 3.2, entitled *Cross validation (CV) and input perturbations for reliable uncertainty estimation*. The subsection begins with introduction of the Brier score (Eq. 2). From what I understand, the Brier score has nothing to do with either cross validation or input perturbations per se, but only with the assessment of the input perturbations. It should be introduced at a later stage, and certainly after cross validation has been introduced.

Thank you for this feedback. To increase readability, we moved the introduction of the Brier score to later in this section and section 3.2. starts as follows:

"In this study, the predicted total ensemble uncertainty results from the combination of CV techniques and input ensemble perturbations." Lines 218-219

1b. In any case, a 'predicted average' LSS^{bar} (sorry, no upper bar on my text editor) is introduced on the occasion of the Brier score. It is not clearly said which kind of average that is, nor on which kind of prediction it is obtained. By the time the Brier score is introduced, it should have already been said that the output of the entire estimation process essentially consists of two *LSS* maps, viz. LSS_{100} and LSS_{2500}, and that LSS^{bar} will normally be, for each grid cell *i*, the average of *LSS* over one of those two maps.

Thank you for this feedback. $\overline{LSS}$ is the average of all LSS values (100 or 2500) per grid cell, and the Brier Score is the sum over all grid cells of the difference between $\overline{LSS}$ and the corresponding observation (landslide presence = 1, landslide absence = 0). We added "per grid cell" to the introduction of the averages to make its definition clearer:

"[…] we obtain a total of 100 LSS maps […] that allow for calculations of an ensemble average LSS ($\overline{\boldsymbol{LSS}}_{\mathbf{100}}$), as well as a standard deviation ($\boldsymbol{\sigma}_{\boldsymbol{LSS}_{\mathbf{100}}}$) per grid cell." Lines 233-234

"In combination with the 5 model equations and 20 repetitions for the CV ensemble, this results in a total amount of 2500 LSS maps […] with corresponding average ($\overline{\boldsymbol{LSS}}_{\mathbf{2500}}$) and standard deviation ($\boldsymbol{\sigma}_{\boldsymbol{LSS}_{\mathbf{2500}}}$) per grid cell." Lines 238-240

With the feedback from 1.b, we have moved the introduction of the Brier Score to after these averages are introduced.

1c. Then, just after Eq. (2) (l. 217), reference is made to an undefined ensemble variance σ^2 *LSS* which has nothing to do with the Brier score, and to an undefined '*actual' uncertainty*.

With the new order, the variance is now introduced after the ensemble standard deviations ($\sigma_{LSS_{100}}$ and $\sigma_{LSS_{2500}}$), so that it should be understandable. We moreover now clearly introduce what the phrase 'actual' uncertainty refers to:

"The aim is to design an LSS model setup so that the predicted total ensemble uncertainty, quantified by the ensemble variance or spread (σ^2_{LSS}) matches the discrepancy between predictions and observations which we refer to as the 'actual' uncertainty. A measure of this actual uncertainty is the Brier Score (BS) (Wilks, 2011) which compares the predicted average LSS ($\overline{\boldsymbol{LSS}}$) against landslide observations from the GLC (o) at different grid cells i (i = 1; …;N): […]

This actual uncertainty by design includes model and input error ($\overline{\boldsymbol{LSS}}$), but also error in the reference data (o), and spatial representativeness error." Lines 241-247

1d. It is only later, in the course of rather intricate explanations, that the two maps LSS_{100} and LSS_{2500} are introduced (ll. 237-238 for LSS_{100}, and ll. 244-245 for LSS_{2500}).

All that is only to stress how confusing the paper can be for a reader who is an outsider. I suggest that, as is a common practice in scientific literature, the authors end their Introduction with brief description of the text that will follow, with what will be the content of each Section.

We added a paragraph providing an overview at the end of the introduction, and additionally at the beginning of Section 3. We hope that this helps to avoid confusion of the reader.

"Section 2 introduces the landslide (presence, absence) and environmental data used to create ensemble LSS maps. The LSS model construction based on MELR is introduced in Sect. 3, along with the methods of CV and input predictor variable perturbations for uncertainty assessment, and methods to evaluate the results. Section 4 presents the resulting LSS model structure and selected predictor variables, and the ensemble LSS evaluation for different input perturbations. Section 5 discusses various aspects of the results. The paper closes with a summarizing conclusion." Lines 86-90

“This section introduces the methods used in this study for model construction and evaluation. Section 3.1 introduces the general principles of logistic regression used to derive global LSS estimates, before elaborating the predictor variable selection process and the implementation of average road network density as a random effect. Section 3.2 introduces methods for uncertainty assessment. First, cross validation is introduced with a detailed explanation of the blocked random sampling. Second, the methods of input ensemble perturbations are briefly explained (details are elaborated in Appendix A2). LSS results based on the first approach alone are referred to as ‘CV ensemble’ or LSS_{100}. Results based on both CV and input ensemble perturbations are referred to as ‘full ensemble’ or LSS_{2500}. Section 3.3 introduces the methods and data used for the evaluation of ensemble average LSS and the impact of the extended uncertainty assessment through input perturbations.” Lines 159-166

And, for another example, the authors write (ll. 236-237) *This results in 5 different model equations …* Simply writing *model equations of form (1) …* would make it much easier for the reader.

Thank you for this suggestion. We added it to the manuscript:

“This results in 5 different model equations of form Equation 1 and corresponding LSS maps.” Lines 231-232

2. Ll. 137-138, *We […] sample from the absence grid cells …* Random sampling, or what ?

Then, later on (l. 237), *By repeating the absence sampling 20 times ….* Is that a new sampling of the same kind as in ll. 137-138, or something else ? Please clarify.

Yes, the subset of landslide absence grid cells is randomly sampled and this random subsampling is repeated. We added :

“We therefore randomly sample from the absence grid cells […]” Lines 132-133

“By repeating the random absence grid cell sub-sampling 20 times, we obtain a total of 100 LSS maps […]”Lines 232-233

3. L. 255, *… the logistic regression […] is asymptotic*. What do you mean by *asymptotic* (you use the same word on a number of other occasions, for instance ll. 433-435) ?

By “asymptotic” we want to stress that the logistic regression function is bound to the interval (0,1) following an S-shape: For intermediate predictor variable values, P(Y=1) behaves quasi-linearly. At the upper or lower edge of the predictor variable spaces, however, variations in the exponent ($\alpha + \sum_{i=1}^{n} \beta_i \cdot x_i$) do not propagate into P(Y=1). We added a sentence to clarify this at the first mention of the phrase “asymptotic” in Section 2.3:

“Note that these perturbations in x_i do not linearly propagate into the LSS estimates, because the logistic regression (see Equation 1) relates x_i to LSS via an S-shape curve, with quasi-linear behaviour at the center (i.e. intermediate x_i values) and asymptotic behaviour towards the upper or lower limit (i.e. for very low or high x_i values). Locations of largest perturbation do thus not necessarily coincide with large resulting ensemble uncertainty.” Lines 254-257

4. Figures 2 (right) and 7. The exact meaning of boxes and vertical lines (total spread ?) does not seem to be mentioned.

Figures 2 (right) and 7 used Tukey’s definition of boxplots, where boxes describe the inter-quartile range (IQR), i.e. from Quantile 25 (Q25) to Quantile 75 (Q75), and whiskers extend from the minimum within Q25-1.5*IQR to the maximum within Q75+1.5*IQR. Everything outside this range is defined as outlier (not shown in Figures 2 and 7). To avoid unnecessarily complicated explanation in the paper, we updated Figures 2 and 7 with whiskers now extending from the overall minimum to maximum and

indicate this definition in the captions as well as upon the first mention of box plots in the text (see below). We also updated Figure 8c-d where whiskers formerly extended from the 5th to 95th quantile to have all boxplots shown in the paper of the same definition. The conclusions from the data remain unchanged by these changes in the visualization.

"The right panel shows boxplots of the β-values for each predictor variable (see Equation 1). Whiskers extend from minimum to maximum and boxes from 25th to 75th quantile, with the median indicated in between." Lines 295-296

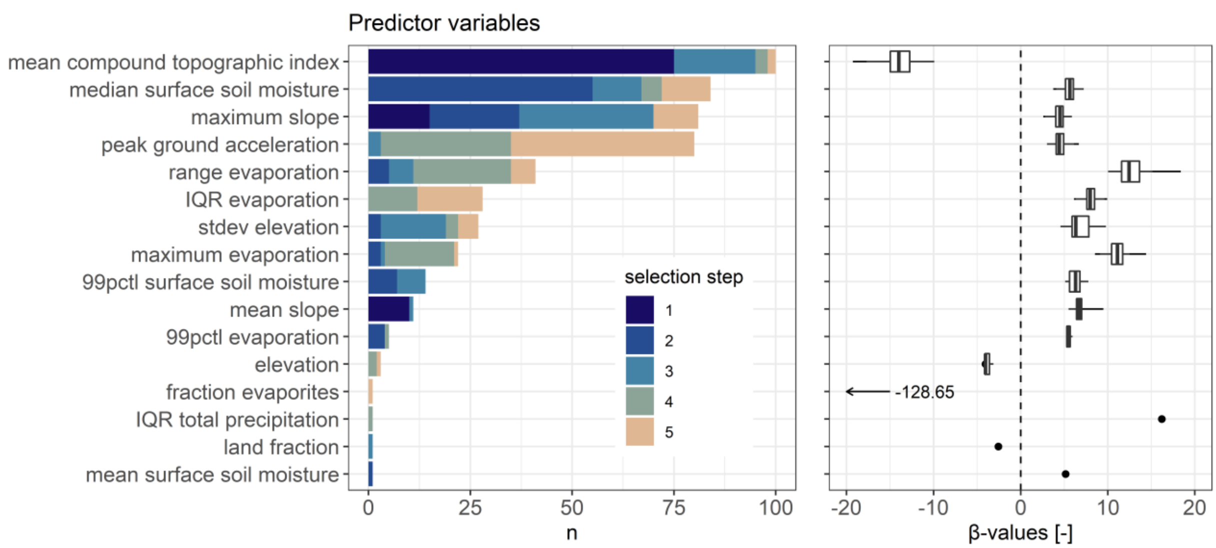

Figure 2 (Left) Frequency of selected predictor variables and (Right) corresponding β-values. [...] Whiskers extend from minimum to maximum β-values. [...]

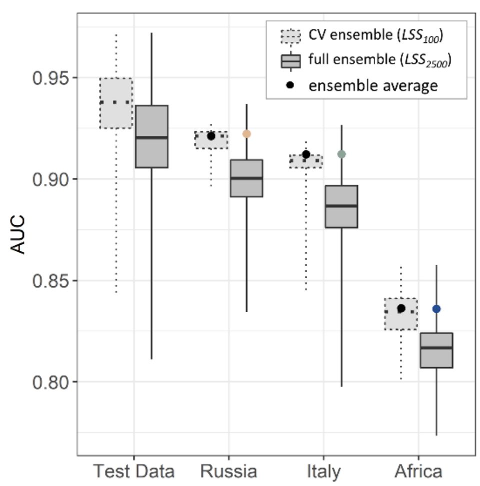

Figure 3 Distribution of AUC for model fitting performance (test data) and model prediction performance [...]. Boxplots are shown for CV ensemble members (LSS_{100}) and full ensemble members (LSS_{2500}, including CV ensemble members), with whiskers extending from minimum to maximum AUC. [...]

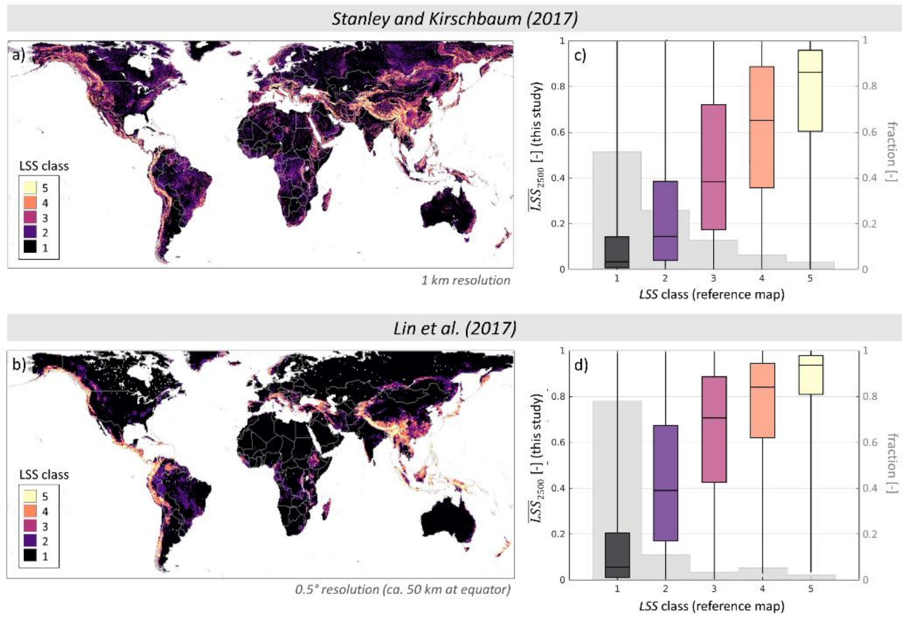

Figure 4 Comparison of $\overline{\boldsymbol{LSS}}_{2500}$ against existing global categorical LSS maps […]. Boxplots show distributions of $\overline{\boldsymbol{LSS}}_{2500}$ values extracted from the nearest 36-km grid cell for each (c) 1-km and (d) 0.5° grid cell in the reference map per LSS class. Whiskers extend from minimum to maximum $\overline{\boldsymbol{LSS}}_{2500}$. […]

5. Ll. 195-196, *The 6 α-values are assumed to come from a zero-mean normal distribution*. What does that mean ? The text that follows says that the quantity RND has not been used as a predictor variable, but does not really explain how the parameter α has been defined.

Thank you for this question. The parameter α, being the intercept in the exponent of the logistic regression function (1), is obtained during the model fitting process. For traditional logistic regression (fixed effect), this would be one parameter value for the whole training data set, fit alongside the β-factors by maximum likelihood estimation. For the mixed effects logistic regression, the best fitting α_j values are obtained separately for the j=6 RND groups, while ensuring that the β-values remain the same for the whole training data set, i.e. across all RND groups. In the model fitting algorithm glmer in R, the variation of these α_j-values around an "average" intercept or α, is defined to be from a zero-mean normal distribution (see Zuur et al. 2009).

We acknowledge that the phrasing was unfortunate, and altered the manuscript as follows:

"The mixed effects approach will then result in one global logistic regression equation that has the same β-factors for all grid cells, but 6 different α-values according to each grid cell's RND class. For model fitting purposes it is assumed that these 6 α-values come from a normal distribution (Zuur, 2009)." Lines 194-197

6. Ll. 177-178, *A one unit change in the predictor variable x_i results in a multiplicative change in the odds of landslide presence by exp(β_i).* Well, only for small P(Y=1)

Thank you for this remark. The multiplicative change we refer to is in the odds, i.e. the ratio of P(Y=1) and P(Y≠1) = 1-P(Y=1), and not in P(Y=1) itself. More specifically, the multiplicative change is obtained

by comparing the odds for x_i+1 to those of x_i. This so-called odds ratio (Zuur et al. 2009) simplifies to $exp(\beta_i)$, which is the multiplicative change we refer to.

$$odds = \frac{P(Y=1)}{1-P(Y=1)} = exp(\alpha + \sum_{i=1}^{n} \beta_i \cdot x_i)$$

$$odds\ ratio = \frac{odds(x_i+1)}{odds(x_i)} = exp(\beta_i)$$

While not linearly connected, an increase in the odds goes along with an increase in LSS (as illustrated on the figures below). We added a sentence explaining the concept of the odds and the connection with LSS:

"A one unit change in the predictor variable x_i results in a multiplicative change by exp(β_i) in the odds of landslide presence, defined as the ratio of P(Y = 1)/(1-P(Y = 1)) = $exp(\alpha + \sum_{i=1}^{n} \beta_i \cdot x_i)$. An increase in the odds of landslide presence is associated with a (non-linear) increase in LSS. Positive (negative) β-values hence indicate an increase (decrease) in LSS with an increase in the predictor variable." Lines 176-179

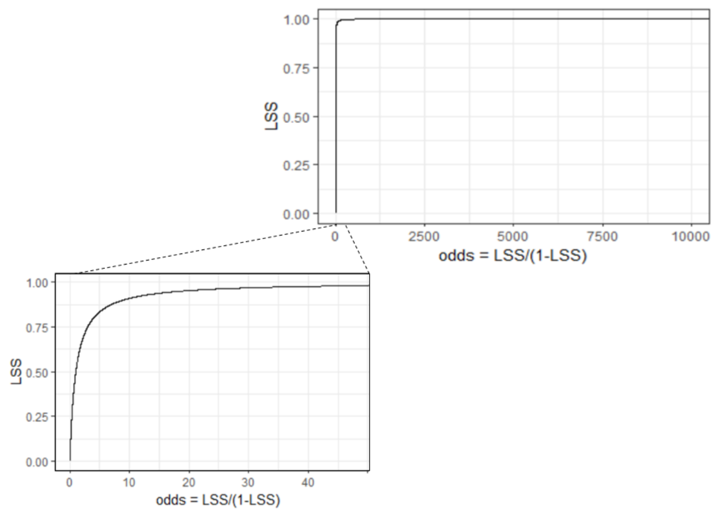

Figure 5 Illustration of the connection between LSS and the odds

7. Figure 4b does not seem to be commented upon. There is no point in including a figure in a paper if it not for saying what conclusion, however succinct, must be drawn from it.

Thank you for this remark. We updated the discussion of Figure 4 as follows (see more in response to comment #10 below):

"Figure 4a and b show that the LSS uncertainty is a function of the average LSS values[…]" Line 329-330

8. Ll. 261-262, *true positive rate , false positive rate.* Explain

We added an explanation of these two in the manuscript:

"For the ROC, the true positive rate of one LSS map is displayed against its false positive rate for different possible thresholds in the continuous probability (here: LSS) that is predicted. The true positive rate is the proportion of correctly predicted landslide presence grid cells when applying said threshold ('true positives') of all observed landslide presence grid cells (Wilks, 2011). The false positive rate is the

proportion of erroneously predicted landslide presence grid cells ('false positives') of all observed landslide absence grid cells." Lines 261-265

9. Ll. 324 and 409, *Sinai peninsula*, actually the Sinai peninsula is a very small region at a global scale. I suspect you mean *Arabian peninsula* Same lines, *Sahara → a large part of Africa*

Thank you for this remark. We updated these in the manuscript.

"Very flat areas or planes, such as central northern Canada, Siberia, the Tibetan plateau, the Arabian peninsula, large parts of Africa (especially the Sahara) as well as central Australia have very low $\overline{\boldsymbol{LSS}}_{\mathbf{2500}}$." Lines 313-315

"At the same time, $\overline{\boldsymbol{LSS}}_{\mathbf{2500}}$ shows much less variation than the map by Stanley and Kirschbaum (2017) within large deserts (Sahara, Arabian peninsula and central Australia)."Lines 396-397

10. Ll. 347-348, … *the ensemble averages […] are similar*, …. That is actually visible from the bottom panel 4c. Correct accordingly.

We agree and revised the discussion of Figure 4 in the manuscript as follows:

"Figure 4a and b show that the LSS uncertainty is a function of the average LSS values and that $\boldsymbol{\sigma}_{\boldsymbol{LSS}_{\mathbf{2500}}}$ is typically higher than $\boldsymbol{\sigma}_{\boldsymbol{LSS}_{\mathbf{100}}}$. Figure 4d shows that the differences between $\boldsymbol{\sigma}_{\boldsymbol{LSS}_{\mathbf{2500}}}$ and $\boldsymbol{\sigma}_{\boldsymbol{LSS}_{\mathbf{100}}}$ are smallest for the very high and low $\boldsymbol{\sigma}_{\boldsymbol{LSS}_{\mathbf{100}}}$. However, Figure 4c shows that the ensemble averages $\overline{\boldsymbol{LSS}}_{\mathbf{2500}}$ and $\overline{\boldsymbol{LSS}}_{\mathbf{100}}$ are similar, as expected from the additional zero-mean predictor variable perturbation. The values of $\overline{\boldsymbol{LSS}}_{\mathbf{2500}}$ are slightly smaller than those of $\overline{\boldsymbol{LSS}}_{\mathbf{100}}$, except for very small $\overline{\boldsymbol{LSS}}$ (<0.1). " Lines 329-335

11. L. 433, LSS_{2500}- $\sigma_{LSS2500}$ (not LSS_{2500}- $_{\sigma LSS2500}$)

Thank you for spotting this. This is an unintended artifact of the pdf rendering, and to avoid this, we rephrase the text as follows:

"The reasons for this relationship between $\overline{\boldsymbol{LSS}}_{\mathbf{2500}}$ and $\boldsymbol{\sigma}_{\boldsymbol{LSS}_{\mathbf{2500}}}$ are twofold: […]" Lines 419