I have revised the manuscript "Estimating global landslide susceptibility and its uncertainty through ensemble modelling" by Anne Felsberg and co-authors. The procedure is very interesting and it would be interesting to apply it to a smaller area where you can control better the training and the validation landslide and thematic data. The manuscript is well written and well organized but there are few major thinks nor really convincing.

1. It is possible to select the hydrologically triggered landslides from the Global Landslide Catalogue?
2. Is not clear how you have used the road network density
3. In the analysis you have mentioned model uncertainty and input uncertainty. I think the uncertainty associated to a 36-km spatial resolution grid is so large that the entire analysis is not relevant. As you mention at line 51 "*coarser input data might be less representative for local events, such as landslides"*. If data are not representative the entire modelling is not representative.
4. Line 20 *LSS maps derived from environmental conditions are a fundamental tool for informing local population, city planners and decision makers both on the immanent landslide likelihood, but also about secondary effects such as major sediment sources* → this is true but very difficult to be applicable at the resolution of your analysis.
5. (Line 60) The total uncertainty is estimated by comparing the predicted average LSS against the observed presence and absence of landslides → in your case the presence/absence of landslide is related to a too coarse grid resolution.
6. Due to their generalizing nature, LSS models are however prone to uncertainty. → true but the uncertainty it is also highly related to the thematic data/landslide distribution/model used for the assessment. In your case the uncertainty associated to data and landslide distribution is more relevant than the entire modelling.
7. (line 89) When you aggregate landslide data in a landslide location, do you check that your aggregation is reliable?
8. (Line 90) Multiple landslides within the same 36-km grid cell are aggregated into one 'landslide location --> The environmental condition selected in a 36 km grid cell can be completely inappropriate and not relevant to explain the landslide.
9. (Line 111) *Absence grid cells are hence selected from grid cells 7 to 15 around a landslide occurrence* → How you can be sure that the selected conditions are not prone to landslides?
10. To compute the compound topographic index, you need the specific catchment area and the slope. How do you measure then in a 36 km grid?
11. Why do you consider the peak ground acceleration if you want to evaluate hydrologically triggered landslides?
12. (Line 161) → The mixed effects approach allows us to include a so-called 'random effect', here the random intercept α, for which we use the average road network density stratified into 6 groups (divided by the global quintile thresholds) → not clear
13. (Line 185) We group the grid cells into a total of 100 blocks according to climatological conditions within 10 predefined regions (roughly two per continent), independent of landslide absence or presence → a) this means that each block has about 75 pixel? b) What is the rational to select roughly two climatological conditions per continent? If this is not a consistent selection is not representative.
14. (Line 183) One subset consists of 20 randomly sampled 'blocks', i.e. small groups, of the 7514 grid cells selected for model creation. We group the grid cells into a total of 100 blocks according to climatological conditions within 10 predefined regions (roughly two

per continent), independent of landslide absence or presence. Within these regions, we mimic typical climatological zonations (for example that of Köppen) through k-means clustering (Lloyd, 1982) of 30-year average soil surface temperature and rainfall (see Table 1), dividing each region into 10 blocks. → Not clear the relation between the blocks and the 5 subset.

15. (Line 224) Aggregated data vs observations → Do you really think is reliable to aggregate original observations? In Italy for example you have aggregated 5438 observations in 309 points. I think the two data are completely different and infact you get very low ROC curve.
16. (Line 236) *Values of the intercept, which is part of all models, vary with road network density as part of the MELR and mostly have and average close to zero (not shown)* → can you explain better this statement?
17. Fig. 3 → how much all this complex analysis improves/enhances at worldwide scale a simple regression model applied to obtain an LLS?
18. (Line 265) *The LSS2500 map hence performs very well over Russia and Africa, while showing some difficulties to capture the patterns for Italy* → This is the situation where you have modelled 309 points aggregated from 5438 observation. Are you sure that the aggregated points are representative of the failure distribution around the world?
19. The procedure is quite complex and the real meaning between the LSS_{2500} and LSS_{100} is not very easy to understand