

Review of "Evaluation of a model for mapping intense pluvial runoff hazard using proxy data of runoff-related impacts. Application to the railway context." by Braud et al.

The paper focuses on the use and evaluation of the Indicator of Intense Pluvial Runoff (IRIP) method. The method is used to provide susceptibility maps for a railway line in Northern France. The paper is well written and organized. The topic is also of high interest. However, one major issue in the methodology may undermine the results of the entire work. Moreover, the authors should be very careful in not using the same figures and same wording used in their previous works.

Major comments:

- 1) This is the critical issue: In Step 4 is written that "If an area classified at risk has a specific supervision measure or mitigation structures have been built, it is moved from 'false alarm' to 'hit' as the implementation of mitigation measures means that the area was indeed at risk, but that no impact was recorded due to the efficiency of the mitigation measure." Step 4 brings a significant bias in the evaluation, which can be seen as a unidirectional attempt to improve model performance. If authors would like to use an approach taking into account mitigation measures, they should also do the opposite: "If an area not classified at risk has a measure or structure, and no impact was recorded, it should be moved from 'Correct negative' to 'miss'", or, at least, according to section 2.4, if the area was in the "bad weather tour". Since this may worsen performance, maybe the authors should either create a different model for areas with mitigation measures, or entirely remove them from the evaluation. The performance boost that results in the last column of Table 5 is due to the inappropriate method mentioned above. This aspect is crucial for the entire paper, because the "Results" section concludes with this sentence:
"The results in the last column present very encouraging values, highlighting the added value of the IRIP maps, and of the vulnerability and mitigation measures characterization, for the evaluation of the IRIP model."
But the last column is affected by this issue, and this casts a shadow on the entire paper.
- 2) Figure 1 is the same as Figure 4 already published in Lagadec et al., 2018. I understand that you are using the same method. But if a figure has been already published, this should be mentioned in the paper. Moreover, the description of the IRIP method on page 3 uses exactly the same words used in Lagadec et al., 2018.
Reference:
Lagadec, L.-R., Moulin, L., Braud, I., Chazelle, B., and Breil, P.: A surface runoff mapping method for optimizing risk assessment on railways, *Safety Science*, 110, 253-267, <https://doi.org/10.1016/j.ssci.2018.05.014>, 2018.
- 3) Page 3 lines 36-40: Please specify the classification method used. These lines are very unspecific and the conclusions from these lines are not supported.
- 4) Page 5 lines 15-16: this statement is incorrect. The results of the chi-square do not demonstrate that the relationship is highly significant, but that it possible to reject the

null hypothesis of independence, because it is unlikely that the null hypothesis of independence is true.

- 5) Page 10 lines 2-8: the demonstration (or assumption?) that each section had the chance to experience a rare event is obscure to me.

Minor comments:

- 1) Page 3 line 39: reduces->reduced
- 2) Page 6 line 15: either...or
- 3) Table 2: please specify also in the table the number of d.o.f for the chi-square test.