We thank Reviewer #2 for their constructive comments and support for publication. Below we indicate how these comments will be addressed.

Authors analyse two beach safety-related hazards in the Atlantic French coast and propose a BN-based model to describe/predict them as a function of a hazard and an exposure component, each one comprising relevant variables characterising environmental (hydrodynamic and climatic/weather) conditions. The topic is in line with one of the targets of NHESS and, in this sense, the manuscript can be of interest for many NHESS readers. In what follows, some observations/comments/suggestions are given.

General comments

[1] The proposed model includes different environmental variables to characterize the hazard and exposure components. With respect to the exposure component, the implicit hypothesis is that the meteorological conditions together the time of occurrence should indicate the beachgoers affluence. Although it seems a reasonable assumption, it would be recommendable to perform an independent validation because the errors/uncertainty in this "submodel" (weather conditions controlling number of beachgoers) should affect the predictability skill of the overall model (dependence of the probability of SZI on number of beachgoers). In a previous paper (Castelle et al 2019), the authors concluded the need of better quantify exposure to better explain SZI. Is there any curve of affluence for the studied beaches which can be used to "validate" the conceptual model of exposure?

We agree with the reviewer that beachgoer affluence, and particularly exposure (affluence of beachgoers in the surf), should be validated. Unfortunately, at the time of writing this response we still have no affluence / exposure data. We have a proposal currently under review which one of the aims will be to estimate beach affluence, including number of people in th bathing zone at the time of each accident in the future injury report forms that are filled by the lifeguards. In line with Castelle et al. (2019) we will further indicate that research effort must be made into beachgoer affluence and exposure to hazard in the surf.

[2] Line 137. Please define "tidal gradient".

Tidal gradient was defined as the time derivative of tidal elevation, this will be specified in the revised manuscript when introducing the variables.

[3] Line 148. Please specify the number of the "Figure".

We will fix this

[4] Figures 3 and 4 can easily be combined in just one.

We think that, for readability and to better discriminate shore-break and rip-current related injuries, it is better to keep two different figures.

[5] Lines 196-198. One of the implications of discretizing injuries in binary form (and duplicating cases when exceeding 1) is to artificially augment data to be used in the BN. May this affect the real representativeness of environmental conditions affecting SZI? Also, this may also artificially increase the BN predictive performance.

The objective is indeed to artificially increase the amount of data to be used in the BN. It is similar to a weight function. It is a common way to improve prediction model performance (BN or other models). It will not falsely increase performance as long as the model is tested/validated against different dataset (i.e. a replicated data are not splitted between fit and validation datasets).

[6] Figure 6.b. The confusion matrix indicates a poor performance of the model to predict injuries. I assume that this matrix corresponds to one of the tested fold runs (in fact, one with the poorest performances). However, as it is presented it seems that this is the overall performance of the model, which if so, it indicates the absence of reasonable predictive model. Please put the matrix in the right context.

The Reviewer is right. According to a comment by the other reviewer, this panel has been removed as it provides misleading information.

[7] You propose three different metrics to measure the BN predictive performance. Do we really need to use all of them to identify the best number of bins to be used in the model? If so, how their results must be jointly interpreted (integrated)?

See response on RC1. Metrics are complementary. Skill is less suitable for binary predictors and does not take into account the confidence of the model, whereas LLR does this. AUC/ROC takes the ratio of false positives and true positives, which is a natural for a binary metric.

[8] Line 252. M is defined, but where do you use it?

We will include M in Eq(8) which was missing

[9] Line 278-279. So, finally, how many bins are selected for each BN. Looking to the final BNs (figures 10 & 12), you have selected 6 for shore-break and 7 for rip-current. However, the variation in prediction for both when changing from 6 to 7 is almost the same. Why didn´t you use the same number of bins in both BNs? This should give more consistency to the analysis since input variables which are almost the same will be discretized in the same way.

Given, that the AUC/ROC does perform significantly better for the 7 bins, we chose 7 bins. Even though for Skill/ Summed LLR variations are small. Choosing the same number of bins for both BNs would result in at least one sub-optimal model, which we avoided.

[10] Line 258. Are you using "complexity" to refer to the number of bins used. I would not say this is complexity but "level of definition" (or similar). I understand "complexity" by the number of components/variables used in the model.

The reviewer is right, this sentence will be reworded accordingly.

[11] Line 289. You mention that beach profiles were taken only at one location. Unless that the beach is alongshore uniform, this may significantly affect to the role of IFS within the BN. Since you are using sinuosity (to predict rip SZI) as a measure of beach departure from alongshore uniformity, I presume that the slope will change along the beach. Do you have an estimation of the range of variation of IFS? Maybe this is the reason for the very low contribution of IFS to the variance.

The writing was misleading here, thank you for pointing out this issue. As indicated in Section 2.2 and shown in Figure 3a, in fact four profiles distributed along the coast were used to account for the alongshore variable beach slope. This will be further clarified in the revised manuscript.

[12] Line 293. I think that figures referring % variance of Exposure and Hazard are wrong (figure 9c). Exposure is the larger than hazard.

Thank you for pointing this. This has now been fixed.

[13] Figure 9. Can you use the same scale in the X-axis for both hazards for an easier visual comparison? Are figures 9b and 9d just a zoom of figures 9a & 9c after removing the first two contributions (hidden variables)? If yes, please indicate, as it is written it suggests they were obtained independently.

We stress that the variance reduction is model dependent and should not be interpreted as an absolute metric to do inter-model comparison. Therefore, for better clarity of the figure we kept the same scale as in the former manuscript.

[14] Lines 294-301. When you compare the contribution of hazard latent variables to the variance for shore-break and rip-current hazards, you obtain a larger contribution in the case of rip than for shore-break. May this indicate that the built model (selected variables) for the shore-break is worse than the one for rips? (e.g. the above-mentioned potential effect of neglecting variations in IFS -comment [10]-)

This should not be interpreted as a inter-model comparable performance metric, but as a model dependent sensitivity to the target variable (injury). Consequently, we can only conclude that hazard is less sensitive to injury then exposure is for the shorebreak BN and vice versa for the rip BN.

Line 297. If we compare the %variance of exposure variables (hour, temp, I) in both hazards (fig 9b & 9d), they are very similar (different ordering but same order of magnitude).

We agree, this will be pointed out in the revised manuscript.

Line 300. Formally, you are not including wave energy as a variable in your BN. Consider that wave energy will involve a non-linear combination of H and T and the observed contribution of these individual variables may significantly vary when combined to characterize wave energy.

The reviewer is right, this will be rewritten.

[15] Line 311. Where the update for larger wave heights (and the resulting change in % of shore-break injuries) can be seen?

It was not shown in order to reduce the number of figures, this will be clarified.

[16] Lines 315-318. I would say here that the most likely IFS during a shore-break hazard is a steep slope whatever the tidal elevation is. Then, if you want you can highlight secondary differences in IFS, but in any case you are just concentrating in one single class for intermediate IFS (27.5 to 35), but when you refer to intermediate you could also include (20 to 27.5) and then, the probability at both tide levels would be almost the same. Furthermore, trying to draw any conclusions about IFS at low tide does not seem to make much sense, since IFS is measured above mean sea level.

We agree. This paragraph and Figure 10 will be removed.

[17] Lines 321-323. What are larger tidal gradients? If we consider the three central bins as representative of medium-low gradients, they are concentrating a similar % of occurrence. In fact, if we compare the probability distribution is almost similar to the prior one (Fig 11a).

The reviewer is right, differences are slight and mostly for the extreme tidal gradients, this wil be clarified.

[18] Lines 326-328. It is not clear from Figure 12b that more sinuous shorelines show increased probability of injuries. The updated distribution of Sin is quite similar to the prior distribution. Are the small changes detected large enough to support your conclusion?

We agree that the difference is small, we will slightly reword this part.

[19] Lines 329-332. I disagree that your analysis is really reflecting that the peak of rip injuries (13h – 15 h) is much earlier than the one for shore-break one (14 to 16.33). Both bins overlap, which may be associated with the comparison of different bins resulting from using 6 classes in shore-break and 7 classes in rip for a same variable (Time). According to Castelle et al (2019) "For low TR, daily minimum tide elevation, which is when channel rip activity is maximised, tends to occur during the patrolled hours in the mid-to-late afternoon (Fig. 12b) when beach attendance (exposure) is maximised". This does not seem to fully support your conclusion.

For the first point made, we agree that “much earlier” was too strong a statement, it will be reworded. We think that the comparison with Castelle et al. (2019) is not relevant here as it deals with low TR only, while here all tide ranges are considered.

[20] Legend of Figure 13. Please change the legend to something similar to the one used in Fig 11 (e.g. Scenario with low sinuosity resulting in a higher probability of shorenormal wave direction). As it stands, it seems that you build the scenario by fixing booth sinuosity and direction.

Figure caption will be improved.

[21] Lines 347-348. I agree with this comment. It would be interesting to assess the profile of people injured by shore-break and rips to identify potential factors affecting their relative exposure.

Thank for this comment. It is, together with beachgoer attendance, the topic of a research proposal with social scientists which is currently under review.

[22] Lines 359-360. See also the combination of video images and numerical modelling to help managing beach safety (Jiménez et al. 2007. Beach recreation planning using video-derived coastal state indicators. Coastal Engineering, 54, 507-521).

Thank you, reference to this site will be included in the revised manuscript.

[23] Section 5.2. Please adjust comments on the role of different environmental factors according to your response to previous comments [e.g. 16, 17, 18, 19]

This section will be rewritten in depth according to these comments and those by the other reviewer.