We thank Reviewer #1 for their constructive comments and support for publication. Below we indicate how these comments will be addressed.

General comments

This paper addresses the occurrence of rip current drownings and shore-break related injuries, and seeks to develop predictive and diagnostic models in the form of Bayesian Networks to model the occurrence of such incidents under different environmental conditions. The paper addresses a scientific question of relevance to NHESS, and uses a relatively novel way to model the occurrence of bathing incidents. In general, the paper is well written and structured, and the analysis of the data is well executed. The presented models seem to provide some skill in predicting the occurrence of the incidents, however, the skill varies considerably depending on which skill metric they present. Of most concern, they present a confusion matrix (Fig. 6b) which indicates that the model provides a very high rate of false negatives and only correctly predicts 4 out of the 56 observed injuries. If this table is correct, then the models are worse than useless as the vast majority of the time that injuries occur would be predicted to be safe. Please check this table and if the results are correct, the models need re-designing to a point where very few false negatives occur, otherwise the poor performance of the model undermines all the other findings.

Specific comments

- **Line 6: "hidden hazard and exposure variables". It is not clear to the reader what is meant by 'hidden variables' at this stage, so I would suggest elaborating on this in the abstract if it is important to the methodology, or not mention 'hidden' in the abstract at all.**

We will clarify.

- **Lines 7-8: "Validation (prediction) results slightly improve predictions of SZIs with a poor to fair skill based on a combination of different metrics." It is not clear what is meant by this sentence; validation is used to prove the goodness of fit of a prediction, while calibration is used to improve the goodness of fit, so it is unclear how the validation results could have improved the model performance. Perhaps you mean some form of calibration improved the performance? Please re-word this sentence to make its meaning clearer.**

We will rephrase

- **Lines 9-10: "Shore-break related injuries appear more predictable than rip current drowning incidents as the shore-break BN systematically performed better than the rip current BN." This logic is not quite right – what this result shows is that your model did not perform as well for rips as it does for shore-break injuries, but unfortunately that does not mean that rips are necessarily less predictable (although that may be true). i.e. a different type of model may find rips easier to predict than shore-break injuries. I suggest changing the sentence to something like: "Shore-break related injuries appear**

more predictable than rip current drowning incidents using the selected predictors within a BN, as the shore-break BN systematically performed better than the rip current BN"

We will change.

- **Lines 18-20: "Rapid change in tide elevation during days with large tidal range was also found to result in more drowning incidents, presumably because it favors the rapid onset of rip current activity and can therefore surprise unsuspecting bathers." Is it not also possible that stronger tidal currents (not rip currents) could play a role here? This is certainly observed in parts of the UK where beaches are found near to estuaries/inlets or where tidal flows are very strong. In most cases, tidal currents are very weak compared to rip current flows however, in which case please elaborate on why you discount tidal currents so easily.**

The reviewer is right. According to a similar comment made by Reviewer #2, the presence of tidal currents at some beaches is now discussed but given that it is only hypothesis, it will be removed from the abstract.

- **Line 36: "generally lower tide levels" I would suggest including a citation to Scott et al (2014) 'Controls on macrotidal rip current circulation and hazard', which addressed this topic in some detail**

We will include the citation.

- **Line 50: "A related challenge based on current research is filtering the effect of how water users choices are influenced by environmental conditions (e.g. wave height Hs)." I would suggest also looking at Stokes et al (2017) 'Application of multiple linear regression and Bayesian belief network approaches to model life risk to beach users in the UK' who observed certain parameters were related to higher water user exposure (even beach morphology), and in some cases also higher hazard**

We will cite and discuss Stokes et al. (2017).

- **Line 53: "Finally, the respective contributions of hazard and exposure to the overall life risk for shore-break waves and rip current are virtually unknown." I would suggest explaining the three components (exposure, hazard, and life risk) briefly early on in the introduction, and how they relate to one another. For instance, see explanation in Stokes et al (2017). This is not widely understood I think, and therefore warrants brief explanation.**

We agree. This concept will be explained in the first paragraph of the introduction section.

- **Line 56: "mass rescue days" this term may not be well understood outside of the rip current community, so I would add a brief definition in parenthesis here.**

Given that it is not further used in the paper, reference to mass rescue day will be removed.

- **Line 131: "resulting in a maximum tide elevation error of 0.3 m at all sites" – how was**

this error determined?

This is a (conservative) empirical estimation given that we have no other tide gauge along the coast but only, this will be clarified.

- **Line 136: Insolation is an uncommon parameter in coastal hazards research (I think) and perhaps warrants a brief definition, along with explanation of how the parameter was recorded or predicted.**

We will define insolation (sunshine hourly duration).

- **Figure 4, upper panel: the symbol for the second legend entry cannot be seen**

We will fix this

- **Line 211: you mention model runs, but it is not clear what you mean by 'run' is a run one iteration through the k-fold data divisions? Please clarify this. Also, it is not clear whether the k-folds evaluation was used purely to test the BN or to actually infer the best probabilities within the model; can you clarify the wording in this paragraph to make it more explicit how the model was calibrated/validated.**

Thank you for this comment. A run is indeed one iteration through the k-fold data division. The K-folds evaluation was used purely to test the BN. Best probabilities within the model were fitted differently at each run. We will clarify.

- **4: you state that a value of sk = 1 means perfect prediction, but the equation multiplies by 100% indicating that the values would be from 0-100% with sk = 100% being perfect prediction. Please check this again.**

Thank you for pointing this. 100% will be removed from equation (4)

- **Figure 6 – Major Correction needed. Is this a result or an example? If it is a result, there are a large number of false negatives, which are the worst possible outcome for a predictive model of this kind. This figure makes it look like the model performed very poorly, and therefore the implication of the results in this table needs to be discussed in far more detail than you provide. For instance, out of the 56 observed injuries in the confusion matrix, only 4 of them were correctly identified by the model – is this correct? If so, I would personally throw the model out, but I suspect this is not the case, based on your other skill metrics. Please check this matrix again, and if the results are correct, you need to re-model your incidents in a different way that reduces the occurrence of false negatives.**

This matrix was first presented as an example to provide additional understanding of the ROC curve, but according to comments made by both reviewers it was a misleading piece of information. The confusion matrix depends on the threshold used and cannot be useful as is. Indeed, such a threshold is arbitrary and may vary depending on the goal of the model (do we want a good negative predictive value or a good positive predictive value?). Given that this figure might lead to more confusion than clarification, it was removed in the revised manuscript. All the necessary information for the classification performance is provided by the ROC curve.

- **Line 252: you define variable M, but this variable does not appear in Eqs. 6-8. Is this missing from one of the equations?**

M will be included in Eq(8)

- **Section 4.1: you explain more clearly here how the calibration/validation was performed than in the methods section. I would suggest a slight rewording to make it very clear how it was performed – instead of "For calibration, the BN was used to make predictions of an injury based on the input variables of the training cases" it might be clearer to say "For calibration, the BN was trained to make predictions of an injury based on the input variables of 90% of the training cases."**

Change will be done.

- **Line 258: when you mention complexity are you only referring to number of bins in the discretisation of each variable, or the number of parameters included in the BN? Please clarify this.**

We refer to the level of definition, i.e. number of bins, this will be clarified in the revised manuscript.

- **Lines 292-296: this is an interesting finding, but I think needs to be re-worded slightly as on first reading this did not make sense to me. Now I have read it a few times I think I understand what you are trying to say is that when the hazard parent variables are combined within the BN via the latent hazard variable, they make up the equal largest contribution to life risk, but when those variables are considered separately within the BN and instead linked directly to life risk, they individually contribute relatively little to the overall life risk. Therefore those parameters are required to combine/occur simultaneously in order for the model to explain rip current hazard. I would therefore also avoid stating in lines 296-297 that the exposure variables are the most sensitive parameters for the rip current BN; this is misleading as this is only the case when the latent variables are not included (which is not the final BN you presented earlier).**

We agree with the reviewer that it was misleading, we will reworded so that exposure and hazard exposure are discriminated.

- **Lines 311-314: these lines at first seem to contradict themselves. I would clarify that by increasing the wave height variable, the latent hazard variable increased and life risk therefore also increased. Because the exposure latent variable is not linked to wave height (nor is exposure observed), it doesn't decrease in your BN when you increase wave height, but you can infer that exposure lowers as wave height increases because wave height is intermediate when highest life risk is predicted. Also, is it not possible that life risk is lower at high wave heights purely because they happen less often in summer, and therefore there are less observations at large wave heights? Please explain in the text whether BNs account for the background distribution in the forcing conditions (e.g. in the way that Scott et al (2014) account for the background distributions).**

These lines will be reworded accordingly. It is true that high wave conditions occur less often in summer. However, when we compared the BN updated for an injury, probabilities of larger wave height bins were lower compared to prior probability of larger wave height. The prior probability BN is based on all daily hours during summers, whether there is an injury or not. Consequently, the comparison of prior and updated BN does take into account all summer conditions

- **Lines 315-318: I think these lines are misleading as they suggest that your IFS variable changes with the tide, when in fact IFS was always observed between 1-3 m amsl. IFS changes in this context are therefore due to gradual morphological changes in beach slope, rather than reflecting the slope of the beach face at different stages of the tidal cycle. Can you please check your logic here, and consider re-wording or clarifying these lines please?**

According to your comment and one of the other reviewer, this part will be removed from the revised manuscript**.**

- **Lines 320-321: It is slightly misleading to compare shore break injuries to rip current 'injuries' in the way you do here – you are actually reporting the rate of rip current drownings (fatal or non-fatal), which is a much more severe 'injury' than many of those in the shore break injuries data and probably occurs less often as a result. They are not easily comparable, as there is a whole spectrum of injuries you could get from a shorebreak, from a sprained ankle to drowning, whereas with rip currents a person either survives (and probably nothing reported) or drowns (equivalent to the most severe shore break injury). Essentially, the rip data is almost binary while the shorebreak data represents a continuum of injury severity. Therefore, that sentence should be reworded to say that a shore break injury is more likely than a rip current drowning (fatal or non-fatal).**

The sentence will be reworded accordingly.

- **In section 4.2 you discuss how the driving parameters affect the likelihood of a shorebreak or rip current related incident, but you do this by assuming in the BN that there is a an incident occurring and looking at which forcing variables increase/decrease. I suggest it would also be useful to explore how the percentage likelihood of an incident changes when you change each forcing variable one at a time (e.g. from max to min). This is somewhat addressed in section 4.1.1, but this doesn't provide insight into how the percentage of incidents changes with each variable, only how the variance of the incidents changes.**

Implicitly we account for this by comparing the sensitivity of each variable to the injury variable. This is in terms of % variance reduction, which is a measure for sensitivity. This gives an idea of the magnitude of influence each individual variable has on the injury variable (specifically on precision of prediction) and which combination of variable interactions are useful to explore. In the scenario analysis section 4.2, we highlighted the most noticeable interactions, indeed by using an injury as evidence. Changing one variable at a time gives, in general, little response to the probability of an injury. This, we think, makes sense, because it is particularly the *interaction* between the evidence of *many* variables, which raises the probability. The evidence of a high tide by itself does not dramatically raise the probability of a shorebreak injury, high tide is not the only explanatory variable given how we constructed the BN. If hypothetically we create a BN with just tide, wave height and sun time, the tide might have a more significant effect on the probability of an injury, but it will not be accurate. Accordingly we did not run additional analysis.

- **Lines 334-338: You have set rip hazard to 100% and explored the change in shoreline sinuosity and wave angle, but rip hazard in your BN is a latent variable that doesn't actually relate to an increase in incidents in the real world. Can you provide more justification on how this latent variable can tell us something about rip hazard/incidents please? Shouldn't you instead vary these forcing variables and observe the change in probability of a rip incident? Or, are you simply interested to see how sinuosity changes as wave angle changes? Please clarify this paragraph.**

Yes, here we were essentially interested in the variable interaction, here between sinuosity and wave angle. This is will be better frames in the revised manuscript.

- **Line 344: while Stokes et al (2017) did model hazard and exposure variables, they were not hidden variables in their study, as I believe they had observations of number of people in the water (exposure) and incidents divided by the number of people in the water (hazard).**

The reviewer is right, reference to Stokes et al. (2017) will be removed her**e.**

- **Line 359: I would suggest adding a citation to Scott et al. (2014) 'Controls on macrotidal rip current circulation and hazard' as this paper forms the basis of rip predictions used throughout the UK for the lifeguards there, and uses the approach you mention.**

Reference will be added

- **Line 380: "dn does not affect the hazard posed" – it would be worth pointing out that tidal currents (not rip currents), which are driven by dn, are typically of very low velocity compared to rip currents, which is why you can state that dn is not a driver of hazard directly, while n itself is a driver of hazard as it determines whether the surf-zone is located around, and therefore interacts with, the bar-rip features.**

This part of the discussion will be reworded along the lines provided by the reviewer.

- **Line 390: it is not clear how you came to this conclusion. Can you please clarify this finding earlier on in the paper. I did not come to the same conclusion by reading lines 334-338, which is what I think you are referring to.**

We agree, according to the modifications following the previous comes, this will be reworded.

- **The role the latent variables play in your BN warrants further explanation. Unlike previous studies that have used hazard and exposure in beach life risk studies (for example, Stokes et al., 2017 who you cite) these variables in your BN are not observed variables. Therefore, please explain further how the BN estimates an increase in exposure or hazard based only on the forcing variables and the resulting change in number of incidents.**

The latent variable uses correlations among child nodes to determine the relation between the predictor variable (injury). The latent /hidden node should be interpreted as the (discretized) probability distribution of these correlations amongst child nodes in relation to the predictor variable (injury). This is frequently used when there is an unobserved cause. The latent variables in our BN are the representation of our beliefs translated into probabilities. The goal of this is to show how important these variables are and how they change, given a results (SZI) and/or an environmental parameter.

- **Your model performs poorly when measured with LLR and to some degree 'skill', yet seems to perform very well against AUC. It is therefore hard to place confidence that the model is useful, when some metrics say it performs worse than prior probability. Can you comment on which skill metric is the most appropriate to measure your BN against as it is hard otherwise to conclude on its usefulness.**

Skill and summed LLR are the lesser metrics, because skill does not take into account the confidence of the model in its prediction. The summed LLR is sensitive to highly negative 'anomalous cases' that are hard to predict, giving a distorted view of the model performance. The %LLR>0 nuances this distorted view. The AUC/ROC does take into account the confidence of the model to some extent with a cut-off value, and is therefore better for binary metrics.

Technical corrections

All the small technical corrections below will be done.

- **Abstract line 6: "a hidden hazard and exposure variables" should read "hidden hazard and exposure variables"**
- **Introduction line 57: "…the number of drowning incidents occur during warm sunny days…" should read "…the number of drowning incidents increases disproportionately during warm sunny days…"**
- **Introduction line 63: "the benefits of a BN approach to identify of the characteristics of high risk beaches from a large data set." Should read "the benefits of a BN approach to identify the characteristics of high risk beaches from a large data set."**
- **Line 76: "The Gironde coast is located in southwest of France" should read "The Gironde coast is located in the southwest of France" or "The Gironde coast is located in southwest France"**
- **Line 139: "beach upper beach slope" should read "upper beach slope"**
- **Line 324: reword to 'rip current related drownings are slightly more likely to occur when tides are low' or similar, as cause and effect are the wrong way round in your sentence.**
- **Line 325: "such drowning incidents occur for increased incident wave energy" should read "such drowning incidents occur during increased incident wave energy"**
- **Line 341: please re-word this sentence as it does not make sense.**
- **Line 342: "This allowed to use different beach" should read "Thisf allowed the use of different beach"**
- **Line 345: It is not clear why the word 'importantly' is used in this sentence/line.**
- **Line 355: "Summer beach profiles acquired along the different beaches should help**

improving IFS estimation and, in turn, prediction of rip-current drowning incidents." I think this is incorrect, as you don't relate IFS to rip incidents, only S.

- **Line 361: "the risk of drowning the Gironde" should read "the risk of drowning along the Gironde coast"**
- **Line 363: "could not be retrieved to either" should read "could not be attributed to either"**
- **Line 394: "including a wisely pre-defined hidden hazard variable" this sounds a bit self-congratulating! I would remove the word 'wisely'**
- **Line 399: "with decreased exposure for Hs > 2.5 m, large surf, and thus heavy shore-break waves at the shoreline, discourage the beachgoers to enter the water near high tide." reword to "the predicted decrease in exposure for Hs > 2.5 m, representing heavy shore-break waves at the shoreline, is thought to discourage beachgoers from entering the water near high tide"**
- **Line 404: "more controlled by the exposure than by hazard" reword to "more controlled by exposure related variables than by hazard related variables"**
- **End of line 411: 'improve' rather than 'improves'**
- **Line 420 "skill but providing much less diagnostic Tellier et al. (in revision)." Should read "skill but providing much less diagnostic capability (Tellier et al., in revision)."**