Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2019-142-AC1, 2019

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Machine Learning Analysis of Lifeguard Flag Decisions and Recorded Rescues" *by* Chris Houser et al.

Chris Houser et al.

chris.houser@uwindsor.ca

Received and published: 7 September 2019

RC1

âĂć There seems to be two very strong assumptions made in the paper: i) that the decision tree analysis is infallible; and ii) that the morphology of the inner nearshore bar is the most critical factor relating discrepancies between posted and predicated flag colours and rescues. In the case of the former, it may be that I do not fully understand the methodology, but no model is 100% correct without some sort of ground truthing. I would temper some of the statements/findings in this regard.

This is a fair criticism of the paper and our language may have provided an emotional tone to the description of the model, and we focused strictly on the morphology of the

Printer-friendly version

Discussion paper

inner bar.

There is a potential that the chosen flag is not consistent with the beach user perception of the risk, which may

Results of a decision tree analysis indicate that the colour flag chosen by the lifeguards was different from what the model predicted for 35% of days between 2004 and 2008 (n=396/1125).

when the model predicted a green flag would be more appropriate based on the wind and wave forcing. It is possible that the lifeguards were overly cautious, or they identified a rip forced by a transverse-bar and rip morphology common at the study site. Regardless, the results suggest that beach users may be discounting lifeguard warnings if the flag colour is not consistent with how they perceive the surf hazard or the regional forecast.

the difference between the posted and predicted flag colour could be associated with the lifeguards noting that the nearshore had a transverse bar and rip morphology, which is common at this location. The morphology of the nearshore and other variables that could influence whether a beach user will enter the water or not (e.g. weather, number of beach users or presence of seaweed) are not captured by the current model, which is based on wind and wave forcing alone. The model developed in this study is similar to rip forecasts produced by the US National Weather Service (NWS), and does not include local variables known to the beach manager based on experience and years of careful observation. Discrepancies between the predicted and posted flag colours provide a basis for future model development and expansion. Incorporating more data into the model will it to evolve and better capture the variables that influence the colour of flag chosen by the lifeguards, while ensuring that the model remains computationally efficient. Introducing additional variables, such as nearshore morphology, to the model has the potential to better capture a lifeguard or beach manager's understanding of what constitutes dangerous surf conditions at their beach.

Variables such as the nearshore morphology and the potential for rip development is not included in traditional forecasts or the model developed in this paper, and most beach users use a simple assessment of wave breaking to determine if the water is safe. Even though a lifeguard will post the appropriate flag. . ..

âĂć In the case of the latter, while morphology is indeed critical to rip current formation/presence and rescues, there are many other variables not considered or mentioned in this paper that are also clearly important, such as the weather (sunny, overcast, temperature), the number of beach users, the presence of seaweed and any other factor that may contribute to beach users entering or not entering the water. I therefore think that the rather strong emphasis that nearshore morphology is the critical factor should also be tempered, particularly as some of it is conjecture. The other variables should at least be mentioned as factors to be considered for further extensions of this study, and also for rip forecasts themselves.

While we did provide some additional text (as requested by this reviewer) in the discussion to describe how beach users make decisions, it is the decision of the lifeguard that is of greatest importance in this study and at this location, the morphology of the nearshore is of greatest importance.

The morphology of the nearshore and other variables that could influence whether a beach user will enter the water or not (e.g. weather, number of beach users or presence of seaweed) are not captured by the current model, which is based on wind and wave forcing alone.

âĂć I think there also needs to be a bit more explanation for the chosen 2004-2008 period. There's nothing wrong with that, but were certain data not collected or available after 2008? I would also describe the actual location of the wave buoys – how far offshore were they and at what water depth? Are wave conditions at the buoy likely to be consistent with wave conditions in the nearshore? I would have a location of study diagram indicating their location and also have a picture of a section of the beach showing ‘typical’ rip current conditions along the beach.

We have made the description of the limited data period more explicit in the introduction:

This study examines the consistency of flag warnings at Pensacola Beach, Florida between 2004 and 2008 when daily data is available for flag colour, wind and wave forcing, as well as the daily number of rescues performed by lifeguards.

Please see responses below for more information about the buoy and its location relative to the study site, which is also presented on a map (Figure 1).

âĂć I also found some of the reasoning of the posted vs predicted flag colours and rescues to be a bit confusing, although this might just be me. The authors suggest that the largest number of rescue days/rescues was associated with posted yellow/red flag conditions when the decision tree analysis suggested a green flag would be more appropriate. They suggest that this represents an over-estimation of the surf and hazard risk by the lifeguards (being overly cautious). However, maybe the flag level was absolutely appropriate – dangerous conditions lead to more rescues, not because the lifeguards got the flags wrong, but because beach users were discounting (or were ignorant of) the flags and surf conditions – which the authors note. Defining yellow flag conditions seem to be the main problem as they are associated with most rescues. If a green flag were flying on these days, I don't see how the number of rescues would be any different. In fact, they could lead to more rescues as beach users may assume that conditions were safer and would be more likely to enter the water.

Please see our responses to the line-by-line comments below to see examples of how we have altered the language to make it clear that the ‘overly cautious’ is in the eyes of the beach user if they see a yellow or red flag, but believe that conditions are green or yellow respectively.

âĂć The only way I can see that the central hypothesis of the paper would be correct is if a green flag was flying when a yellow or red flag should have been posted. This is supported by the results on L179-180. Perhaps a central hypothesis is not needed. The paper would be just as valuable if the differences between posted and predicted flag colours was described with a discussion of the real-world implications (which the authors do a good job of in the Discussion). Taking out the hypotheses would eliminate some of the confusion, I think.

The hypotheses, which are now explicitly stated, are important as they are based on the perception of lifeguard accuracy described in the introduction. We have made sure that the results are strictly a presentation of the data without interpretation, which we reserved for the discussion section.

Specifically, it is hypothesized that a greater number of rescues will occur on days when the model underestimated the hazard level compared to the lifeguard who made their decision based on local observations including the presence of semi-permanent rip channels. In this scenario, the public may believe that the lifeguard is being overly cautious leading to people entering the water.

More clarifications are provided through the remainder of this reviewer response for line-specific comments.

âĂć The abstract states that the decision tree analysis suggests that the wrong flag was flown on 35% of days. The term 'wrong' seems overly harsh and does not take into account that the fact that lifeguards were actually there to observe surf conditions. Having said that, there is a considerable amount of subjectivity involved in choosing the flag colour, some of which would be related to human factors of the lifeguards themselves. But to say it's 'wrong' is assuming that the decision tree analysis is always right, which I disagree with.

We have removed all references to the word wrong and replaced by difference between predicted and posted flags. In fact, the phrasing throughout the article makes it clear that the local lifeguard decisions are probably more accurate than a model prediction based solely on wind, wave and water level. Please see the responses to the line-by-line comments below for examples of how this has been changed.

âĂć There is also an important point that should be discussed. The green flags mean a 'low' level of surf and rip current hazard danger, but green is generally universally accepted as 'safe'. This study has clearly shown that rescues can occur during both posted and predicted green flag conditions. An argument could be made that ocean conditions, particularly in the presence of breaking wave activity, should never be flagged as 'green' because, as the authors state, strong rip currents can form under green flag conditions. Other studies (e.g. Scott et al., 2014) have also linked the occurrence of rescues with seemingly 'fine weather' conditions (or something similar). However, this raises important, if not controversial, questions about the validity of the existing flag system and the impact of this, via confirmation bias, on beach users' perceptions if flags were always yellow or red.

This would suggest that posting a green flag should never be permitted when wind and swell waves are breaking over the bar, even if the regional forecast suggests a low-level hazard that day. As shown by Scott et al. (2014), rescues are still possible with seemingly 'fine weather' conditions when a green flag would be predicted by the model or in regional forecasts. Even in the presence of small swell wave, breaking can be induced as water levels fall with the tide (Castelle et al. 2016).

It is difficult for beach users to spot a rip or assess the potential for rip development, and they may assume that the lifeguard is being overly cautious if they perceive fine-weather conditions and the lifeguard posts a yellow or red flag.

âĂć Abstract L16 – perhaps specify '. . .risk to whom'

Update to: "...effective strategies to minimize risk to beach users."

âĂć L18 – should be 'lifeguard(s)'

Corrected.

Printer-friendly version

Discussion paper

âĂć L22 – should be ‘machine learning is used’

Corrected

âĂć L24 – should be ‘wrong colour flag’

Corrected

âĂć L25 – I find this statement a bit confusing – can it be clarified?

This sentence has been updated to: 17% of all rescue days accounting for ∼60% of the total number of rescues.

âĂć L30 – should be ‘surf hazard was associated’

Corrected

âĂć L33 – should be ‘lifeguards’

Corrected

âĂć L40 – I think the first statement should have some references in relation to the specific recognition of rips as a global public health issue

We have added references to rip current drownings and rescues from India, the UK, Costa Rica, Australia, the Great Lakes and the United States:

. . ... serious global public health issue (Brighton et al., 2013; Woodward et al., 2013; Kumar and Prasad et al., 2014; Arozarena et al., 2015; Brewster et al., 2019; Vlodarchyk et al., 2019).

âĂć L42 – should be ‘and are capable of’

Corrected

âĂć L47 – there are better references for this. . .Brighton et al., 2013 for the Australian context (although SLSA, 2017 can remain. . ..if not updated to their national coastal safety report for2018) and Brewster et al. 2019 for the US context

Updated to: for nearly 80% of all rescues (Brighton et al., 2013; Brewster et al., 2019).

Brewster, B.C., Gould, R.E. and Brander, R.W., 2019. Estimations of rip current rescues and drowning in the United States. Natural Hazards and Earth System Sciences, 19(2), pp.389-397. Brighton, B., Sherker, S., Brander, R., Thompson, M. and Bradstreet, A., 2013. Rip current related drowning deaths and rescues in Australia 2004–2011. Natural hazards and earth system sciences, 13(4), pp.1069-1075.

âĂć L51 – there are other papers that could be referenced in addition to the Brannstrom studies

"....knowledge of this hazard is limited (Brander et al., 2011; Williamson et al., 2011; Brannstrom et al., 2014; 2015; Gallop et al., 2016; Fallon et al., 2018; Menard et al., 2018; Silva-Cavalcanti et al., 2018; Trimble and Houser, 2018) and that few people are interested in rip currents compared to other hazards (Houser et al., 2019)."

âĂć L61 – should be 'flag colour'

Corrected

âĂć L69 – should be 'nearshore bars'

Corrected

âĂć L71 – this statement is a bit confusing at it refers to beach users on beaches with either no lifeguards or who may be a long distance away. So presumably if there are no lifeguards, there are no flags? Needs a little bit of clarification

This has been changed to: Rip currents can still be present even if a regional forecast predicts that the hazard potential is low based on wind and wave conditions. Beach users can be at risk if the flag colour is based solely on the regional forecast.

âĂć L74 – not sure I understand the bit about lifeguards intervening if the beach users do not heed the warning flag. The green, yellow, red mean low, moderate, high hazard, but are the latter also associated with the message of 'do not enter the water'? Is that implicit?

We have clarified this sentence:

"…does not heed the warning implied by a yellow or red flag indicating moderate and high ('do not enter the water') hazard levels respectively."

âĂć L80 – they may perceive conditions to be relatively calm, but reinforcing this point is if they enter the water under yellow/red flag conditions and do not experience any difficulties

We have clarified and reinforced this statement:

…conditions appear to the beach user to be relatively calm, the beach user may discount or ignore the forecast now and, in the future, if they enter the water and do not experience any difficulties. Trust and confidence in the authority figures can be eroded if they believe that the lifeguards are being overly cautious.

âĂć L81 – I would re-word to say 'may be eroded' and 'they may believe'. . .don't know for sure unless this is backed up with a reference to study indicating these perceptions are evidence-based

We have qualified this statement:

…conditions appear to the beach user to be relatively calm, the beach user may discount or ignore the forecast now and, in the future, if they enter the water and do not experience any difficulties. Trust and confidence in the authority figures can be eroded if they believe that the lifeguards are being overly cautious.

âĂć L90 – I guess there is an inherent assumption here that the modelled flag colour is always correct? Is that the case?

We have added some language here to show that the model is relatively limited compared to the posted flag. In this respect, it is assumed that the lifeguard is correct based on their local observation:

The modelled flag colour, based solely on wave and wind forcing, can be compared to the flag colour posted by the lifeguards on a particular day to identify days when there is a difference and how that influences the number of rescues performed on that day. It is hypothesized that there will be a greater number of rescues performed on days when there is a difference between the predicted and posted flag colour. Specifically, it is hypothesized that a greater number of rescues will occur on days when the model underestimated the hazard level compared to the lifeguard who made their decision based on local observations including the presence of semi-permanent rip channels. In this scenario, the public may believe that the lifeguard is being overly cautious leading to people entering the water.

âĂć L92 – this is a good hypotheses, but perhaps it should be specific to a particular type of discrepancy. For example, if the flags are red, but the modelled flag colour shows conditions to be yellow or green and vice-verse(n.b. this does come later in the results)

We have clarified the direction of the difference and our belief about the impact on rescues:

"...there is a difference between the predicted and posted flag colour. Specifically, it is hypothesized that a greater number of rescues will occur on days when the model underestimated the hazard level compared to the lifeguard who made their decision based on local observations including the presence of semi-permanent rip channels. In this scenario, the public may believe that the lifeguard is being overly cautious leading to people entering the water."

âĂć L99 – might want to specify the period this data is available for. . .is it just 2004-2008 or ongoing beyond that Methodology L133 – is there a way to describe the actual location of the buoys, at least in terms of distance offshore and water depth?

We have described the dates that the data is available to complete this study:

The analysis was completed for Pensacola Beach, Florida where there is available records of daily flag colours, wind and wave forcing, and lifeguard-performed rescues between 2004 and 2008.

We also added more information about the buoy later in the paper:

Offshore wave conditions and wind forcing function are based on long-term meteorological and oceanographic records from an offshore wave buoy located ∼100 km southeast of the study area (buoy 42039; Figure 1). Between 2004 and 2008, this was the closest buoy to Pensacola Beach and had been previously used to estimate the incident wave field (Wang and Horwitz, 2007; Claudino-Sales et al., 2008; 2010; Houser et al., 2011) and was the basis for the rip hazard at Pensacola Beach until a new buoy was placed closer to the beach in 2009. The available wave data from buoy 42039 included offshore significant wave height, significant wave period, and

âĂć L134 – is this significant wave height and period? Or mean?

This has been updated to: ....significant wave height, significant wave period, and direction

âĂć L138 – should be ‘flag colour’

Corrected

âĂć L141 – should be ‘number of rescues’

Corrected

âĂć L157 – perhaps a Table could be added to show the number of rescues/rescue days per year

Rather than a table, we have added a graph showing the interannual variation in rescues and rescue days.

Figure 2. Interannual variation in number of rescues and rescue days at Pensacola

Beach between 2004 and 2008.

âĂć L159 – I think the term 'rescue days' should be formally defined, perhaps in the Methods to say something like 'a rescue day is defined as any day that had at least one rescue performed'

We have added a definition:

The annual number of rescues and rescue days (ie. days with one or more rescues)....

âĂć L161 – the assumption here is that all the rescues were somehow related to nearshore morphologic conditions, but presumably other factors would influence rescue numbers such as weather (beachgoing weather), waves, beach user numbers etc.

We have qualified this statement and provided an explanation for this assumption:

It is important to note that the CHAID Analysis does not incorporate nearshore morphology as an independent variable because changes in nearshore morphology were not tracked daily over the study period. In this respect, differences between the posted and predicted flag colour may reflect lifeguard observations of nearshore morphology conducive to the development of rip currents despite winds and waves typical of green flag conditions.

âĂć L170 – while it is true that this supports the primary hypotheses, I think it's a bit misleading. Much more relevant are the results in 177-187 and Table 2. My suggestion would be just to focus on these (meaning that Table 1 is not needed).

While this is true, we believe it is important way to start introducing the results of the data analysis to show at the first level that the number of rescues is larger than expected when the posted flag and predicted flag are different. We wouldn't go to the next level of analysis (> or < than) if this weren't true.

âĂć L180 – comma before but needed (throughout the manuscript as well)

Corrected here and throughout manuscript âĂć L183 – should be 'an overly. . .' and '. . .. . ..47 days were associated with 268..' – should explain briefly why an overly cautious flag can present a danger in the context of this paper

Corrected

âĂć L187 – shouldn't this statement also be backed up by rescue numbers?

Rescue numbers have been added:

In comparison, the number of rescues (n=298) was under-represented on days when the posted flag suggested conditions were not as hazardous (n=74) as the model or were identical to the model (n=224).

âĂć L193 – L197 – two statements essentially say the same thing – merge into one – so this essentially says that the modelled flag colour would have been incorrect?

We have kept both statements but changed the sentence structure slightly. Considering that this is the main finding of the study, we use the first sentence as a general/descriptive introduction and provide the specific data in the second sentence:

Specifically, a total of 231 rescues were performed on 37 of the 168 days when the posted flag was yellow, and the model predicted that the flag colour should be green.

At this point in the paper (the results) it is not appropriate to say that the model was incorrect. We have left this interpretation to the discussion section.

âĂć L197 – does the low number of rescues on posted red flag days suggest that the red flags (and lifeguards) are doing their job? Deterring people from entering the water? Table 3 – to me this says that the lower number of rescues on red flag days is due to the red flags/lifeguards doing their job and/or beachgoers clearly recognising that conditions are not good for swimming (or the weather is inclement).

This is correct and we have included this interpretation later in the discussion:

“Most beach users assume that larger breaking waves are more dangerous, and many will not enter the water if they (and the model) believe that it is a ‘red’ flag condition. This may partially explain why there were fewer than expected rescues on days when the posted flag colour was overly conservative (e.g. green or yellow flag was posted when the model predicted a yellow or red flag, respectively). Independent of the flag or warning signs, beach users appear to be making personal decisions about the surf and rip hazard. . .”

âĂć L213 – should be spelled ‘annually’

Corrected

âĂć L237-239 – this sentence does not read properly

This sentence has been changed to: The continuous collection of input data will allow the model to evolve and recognize subtle distinctions in wind and wave conditions that influence flag colour, while ensuring that the model remains computationally efficient.

âĂć L249 – instead of ‘is appropriate’, should be ‘as being appropriate’

Corrected

âĂć L251 – should be ‘wave breaking conditions’

Corrected.

âĂć L277 – should reference the study on confirmation bias in relation to rip currents by Menard et al. (2018)

This reference has been added to the text and to the reference list.

Ménard, A.D., Houser, C., Brander, R.W., Trimble, S. and Scaman, A., 2018. The psychology of beach users: importance of confirmation bias, action, and intention to improving rip current safety. Natural Hazards, 94(2), pp.953-973.

RC2

âĂć Line 25: Which differences? It is not clear

This sentence has been updated to: wrong flag colour was flown on ∼35% of days between 2004 and 2008 (n=396/1125). Days with the wrong flag colour represent only 17% of all rescue days but those days are associated with ∼60% of all rescues between 2004 and 2008.

âĂć Line 29: this seems strange to me; flag deployed over- estimating the risk and more rescues or drownings present? It seems the beach user does not obey the flag command or, if the sea condition for the user didn't match the warning, then the flag warning was correct! Maybe I'll understand later on, but I can understand how an overestimation of risk leads to more rescues.

This is the interesting outcome of the study- the model underpredicting the hazard based solely on wind and wave data alone.

âĂć Line 32: So the largest number of rescues is due to people don't believing the criteria of lifeguards when choosing the colour of the flag?

We added some clarifying language here:

It is possible that the lifeguards were overly cautious, or they identified a rip forced by a transverse-bar and rip morphology common at the study site. Regardless, the results suggest that beach users may be discounting lifeguard warnings if the flag colour is not consistent with how they perceive the surf hazard or the regional forecast

âĂć Line 57: In Costa Rica just few beaches do so

We have clarified this statement: "....(Brander et al., 2013; NWS, 2017), while rip-related drownings on a relatively small number of beaches in Costa Rica account for a disproportionately large number of violent deaths in the country (Arozarena et al., 2015). However, recent...."

âĂć Line 69: ". . .(called a transverse bar and rip morphology). . ." I suggest to write the reference for these classification of beaches which would be Wright and Short (1984).

Reference added:

Wright, L.D. and Short, A.D., 1984. Morphodynamic variability of surf zones and beaches: a synthesis. Marine geology, 56(1-4), pp.93-118.

âĂć Line 94: When the difference overestimates and underestimates the risk, or only in one of these cases?

We have clarified the hypothesis:

“. . ..difference between the predicted and posted flag colour. Specifically, it is hypothesized that a greater number of rescues will occur on days when the model underestimated the hazard level compared to the lifeguard who made their decision based on local observations including the presence of semi-permanent rip channels. In this scenario, the public may believe that the lifeguard is being overly cautious leading to people entering the water.”

âĂć Line 98: it would be convenient and illustrative the inclusion of a view from above of Pensacola beach. Google map shows a large rip current system along the beach.

Figure 1. Map of study site showing location of flagged section of beach and approximate location of the wave buoy used in the analysis and for regional rip forecasts. Also shown is the presence of transverse-bar and rip morphology of the innermost bar and the variable nature of the outermost bar for the flagged section of beach. The aerial image is from summer 2004 (before Hurricane Ivan) and is not necessarily representative of the nearshore morphology throughout the remainder of the study.

âĂć Line 100: the “worst” or the best for beach drowning?

The quote is to the worst, meaning that it has the greatest number of drownings.

âĂć Line 115: where is this number coming from? “The innermost bar varies alongshore at a scale of âĹij1000 m, consistent with the ridge and swale bathymetry, and tends to exhibit a transverse bar and rip morphology immediately landward of the deeper swales." I would show a map of the study site, pointing the main access points and other important features. In addition, a bathymetry contour map would be really appreciated. This would be useful for the reader to really comprehend the beach morphology.

References have been added:

Barrett, G. and Houser, C., 2012. Identifying hotspots of rip current activity using wavelet analysis at Pensacola Beach, Florida. Physical Geography, 33(1), pp.32-49.

Houser, C., Hapke, C. and Hamilton, S., 2008. Controls on coastal dune morphology, shoreline erosion and barrier island response to extreme storms. Geomorphology, 100(3-4), pp.223-240.

âĂć Line 120-128: I think that some pictures or bathymetric/topographic plots showing the evolution of the beach during the period described in this paragraph would really help the reader.

It is not possible to show adequately show the ridge and swale bathymetry and the nearshore morphology on the site map, and this level of detail alongshore does not match the rescue data which has no spatial information. We have, however, provided references to the inner shelf bathymetry and impact on nearshore morphology in the text:

"....inner shelf. The innermost bar varies alongshore at a scale of ~1000 m, consistent with the ridge and swale bathymetry (Houser et al., 2008), and tends to exhibit a transverse bar and rip morphology immediately landward of the deeper swales (Barrett and Houser, 2012). Historically, most drownings and rescues on this popular beach have occurred at these rip hotspots because they correspond to the main access points along the island (Houser et al., 2015; Trimble and Houser, 2018)."

âĂć Line 134: it would be nice to have on a map the location of these buoys

The buoy used in the study has been included in a new Figure (1):

Figure 1. Map of study site showing location of flagged section of beach and approximate location of the wave buoy used in the analysis and for regional rip forecasts. Also shown is the presence of transverse-bar and rip morphology of the innermost bar and the variable nature of the outermost bar for the flagged section of beach. The aerial image is from summer 2004 (before Hurricane Ivan) and is not necessarily representative of the nearshore morphology throughout the remainder of the study.

âĂć Line 130-153: Which exactly are the offshore wave conditions and wind ′ forcing functions used in the model?

The independent variables included in the model are now explicitly referenced in the model description:

“. . ..The goal of CHAID analysis is to build a model that helps explain how independent variables (wind speed, wave height, wave period, wave direction, wind direction and water level) can be merged. . .. . .”

âĂć Is the available data (wave height, period, direction) the same as the data used in the model?

That is correct.

âĂć Which exactly are the wave buoys ′ located near the study area? How far are exactly from the shore? How well correlated are the offshore wave parameters from the buoys to the nearshore wave climate?

We have revised the text to describe the location of the buoy to the field site, but also made note that this was the buoy used in the rip forecasts by the NWS during the study period, and has been used in several other studies to describe the wave field incident to Pensacola Beach:

“. . ..∼100 km southeast of the study area (buoy 42039; Figure 1). Between 2004 and

2008, this was the closest buoy to Pensacola Beach and had been previously used to estimate the incident wave field (Wang and Horwitz, 2007; Claudino-Sales et al., 2008; 2010; Houser et al., 2011) and was the basis for the rip hazard at Pensacola Beach until a new buoy was placed closer to the beach in 2009. The. . ..”

We have also included a figure to show the location of the buoy:

Figure 1. Map of study site showing location of flagged section of beach and approximate location of the wave buoy used in the analysis and for regional rip forecasts. Also shown is the presence of transverse-bar and rip morphology of the innermost bar and the variable nature of the outermost bar for the flagged section of beach. The aerial image is from summer 2004 (before Hurricane Ivan) and is not necessarily representative of the nearshore morphology throughout the remainder of the study.

âĂć I would enrich the description of the CHAID technique with references showing cases where this statistical tool has been applied.

We have included some examples of how CHAID has previously been used in natural hazard research with a focus on those that include perception and decision-making:

Previous use of CHAID analysis in hazard studies include landslide prediction (e.g. Althuwaynee et al., 2014), farmer perception of flooding hazard (Bielders et al., 2003; Tehrany et al., 2015), and property owner perception and decision making along an eroding coast (Smith et al., 2017).

âĂć Line 143: Which are exactly the variables the model uses? Only wave and wind forcing? It is not clear. Does the model use variables relate to nearshore morphology? If not, why does the model identifies situations related to morphology not detected by lifesavers? Or maybe is the lifesavers which identifies those situations and not the model? Those things are not clear here and in the discussion section.

The independent variables included in the model are now explicitly referenced in the model description:

"....The goal of CHAID analysis is to build a model that helps explain how independent variables (wind speed, wave height, wave period, wave direction, wind direction and water level) can be merged......"

We have also provided further clarification in the first part of the results section:

It is important to note that the CHAID Analysis does not incorporate nearshore morphology as an independent variable because changes in nearshore morphology were not tracked daily over the study period. In this respect, differences between the posted and predicted flag colour may reflect lifeguard observations of nearshore morphology conducive to the development of rip currents despite winds and waves typical of green flag conditions.

âĂć Lines 159 "The annual number of rescues and rescue 159 days varied by year with a peak in both the total number of rescues and the number of rescue days" It would be good to better define the differences between number of rescues and the number of rescue days. It would be also necessary to properly define rescue day.

We have provided a definition of rescue day in the sentence:

The annual number of rescues and rescue days (ie. days with one or more rescues) varied by year, with a peak in both the total number of rescues and the number of rescue days in 2005.

âĂć L 227-229 "While rescues did not occur on a vast majority of the days when the posted and predicted flag were different, they accounted for a disproportionately large number of the rescues." Perhaps the term "disproportionately large number" is exaggerated as the number it refers to is just the 60% of the rescues.

We have replaced "disproportionately large number" with "majority of":

While rescues did not occur on a vast majority of the days when the posted and predicted flag colours were different, days when the predicted and posted flag colours were different accounted for a majority of the rescues.

âĂć L 230-232 “Rather, the results suggest that the difference between the posted and predicted flag colors is associated with the morphology of the innermost nearshore bar which is not captured by a model and forecast based on wind and wave forcing alone.” This is a very strong statement as it assumes that the decision made by the beach manager are 100% correct and thus the model is “bad” because it does not account for all the information that the manager have like the beach morphology. However, how accurate the beach managers can really discern beach morphology? Is there any statistics available such as successful rates of discerning beach morphology by lifeguards?

We qualified the focus on nearshore morphology in this sentence and the remainder of the paragraph:

“Rather, the results suggest that the difference between the posted and predicted flag colour could be associated with the lifeguards noting that the nearshore had a transverse bar and rip morphology, which is common at this location. The morphology of the nearshore and other variables that could influence whether a beach user will enter the water or not (e.g. weather, number of beach users or presence of seaweed) are not captured by the current model, which is based on wind and wave forcing alone. The model developed in this study is similar to rip forecasts produced by the US National Weather Service (NWS), and does not include local variables known to the beach manager based on experience and years of careful observation. Discrepancies between the predicted and posted flag colours provide a basis for future model development and expansion. Incorporating more data into the model will it to evolve and better capture the variables that influence the colour of flag chosen by the lifeguards, while ensuring that the model remains computationally efficient. Introducing additional variables, such as nearshore morphology, to the model has the potential to better capture a lifeguard or beach manager’s understanding of what constitutes dangerous surf conditions at their beach. At the same time, it is also important to examine the accuracy of beach managers and lifeguards in assessing the nearshore morphology and potential for rip development."

âĂć L 242 "to the model has the potential to better capture a lifeguard or beach manager's intuition associated with dangerous surf conditions." Again, it is assumed that the lifeguard "intuition" is beyond failure.

This is correct. We have changed the sentence to focus on understanding: ". . . .the potential to better capture a lifeguard or beach manager's understanding of what constitutes dangerous surf conditions at their beach."

âĂć L 258, 276, 283 In these lines phrases such as "erode confidence" are "thrust is eroded" are used. I would suggest to rewriting these phrases and replacing "erode" by other words like "lost" for example.

These have been changed to:

Whether this causes beach users to lose confidence in the lifeguards and other authorities managing the beach is an important question for future research.

the more trust in authority is lost - a beach

beach user, which can cause them to lose their confidence in the lifeguards.

Please also note the supplement to this comment:
https://www.nat-hazards-earth-syst-sci-discuss.net/nhess-2019-142/nhess-2019-142-AC1-supplement.pdf

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-142, 2019.