We thank the reviewers for their helpful and insightful comments that have greatly improved the paper and ensured that we clarified our language about the predict and posted flags.

**RC1**

- 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 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….*

- 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.*

- 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).

- 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.

- 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.

- 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.

- 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.*

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

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

- L18 – should be 'lifeguard(s)'

  Corrected.

- L22 – should be 'machine learning is used'

  Corrected

- L24 – should be 'wrong colour flag'

  Corrected

- 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.*

- L30 – should be 'surf hazard was associated'

  Corrected

- L33 – should be 'lifeguards'

Corrected

- 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).

- L42 – should be 'and are capable of'

    Corrected

- 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.

- 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)."

- L61 – should be 'flag colour'

    Corrected

- L69 – should be 'nearshore bars'

    Corrected

- 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.*

- 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*."

- 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.

- 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.

- 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.*

- 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."*

- 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*

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

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

- L138 – should be ‘flag colour’

  Corrected

- L141 – should be ‘number of rescues’

  Corrected

- 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.

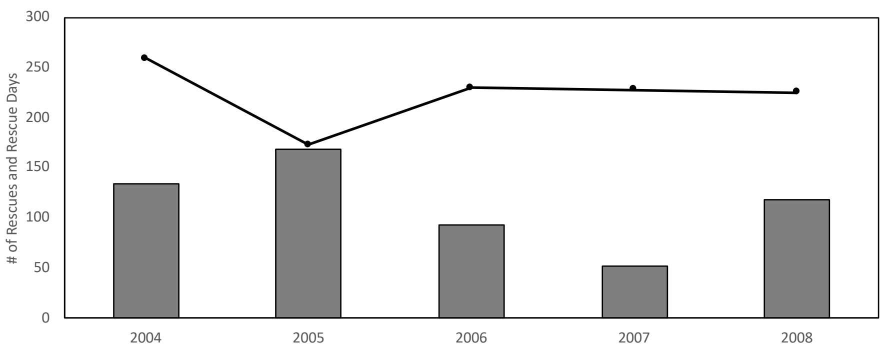


**Figure 2.** Interannual variation in number of rescues and rescue days at Pensacola Beach between 2004 and 2008.

- 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)….*

- 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.*

- 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.

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

Corrected here and throughout manuscript

- 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

- 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).*

- 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.

- 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…"*

- L213 – should be spelled 'annually'

  Corrected

- 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.*

- L249 – instead of 'is appropriate', should be 'as being appropriate'

  Corrected

- L251 – should be 'wave breaking conditions'

  Corrected.

- 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.

**<u>RC2</u>**

- 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.*

- 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.

- 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*

- 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…."*

- 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.

- 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."*

- 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.

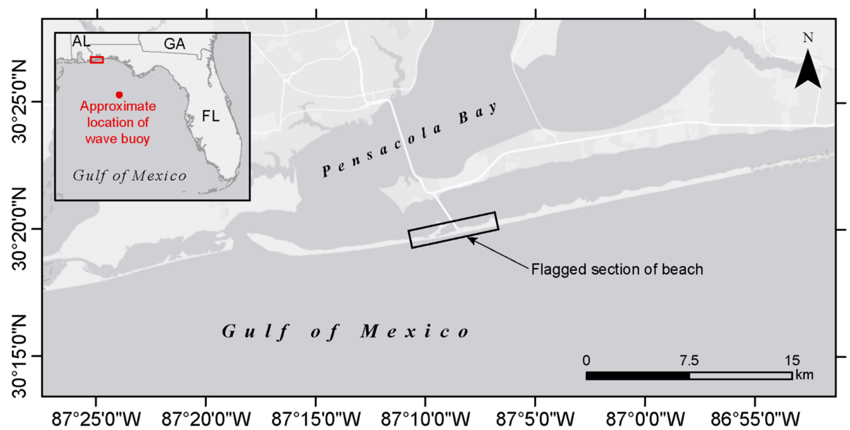


**Figure 1.** Map of study site showing location of flagged section of beach and approximate location of the wave buoy used in the analysis and for regional rip forecasts. Also shown is the presence of transverse-bar and rip morphology of the innermost bar and the variable nature of the outermost bar 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.

- 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.

- Line 115: where is this number coming from? "The innermost bar varies alongshore at a scale of ~1000 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.

- 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)."*

- 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):

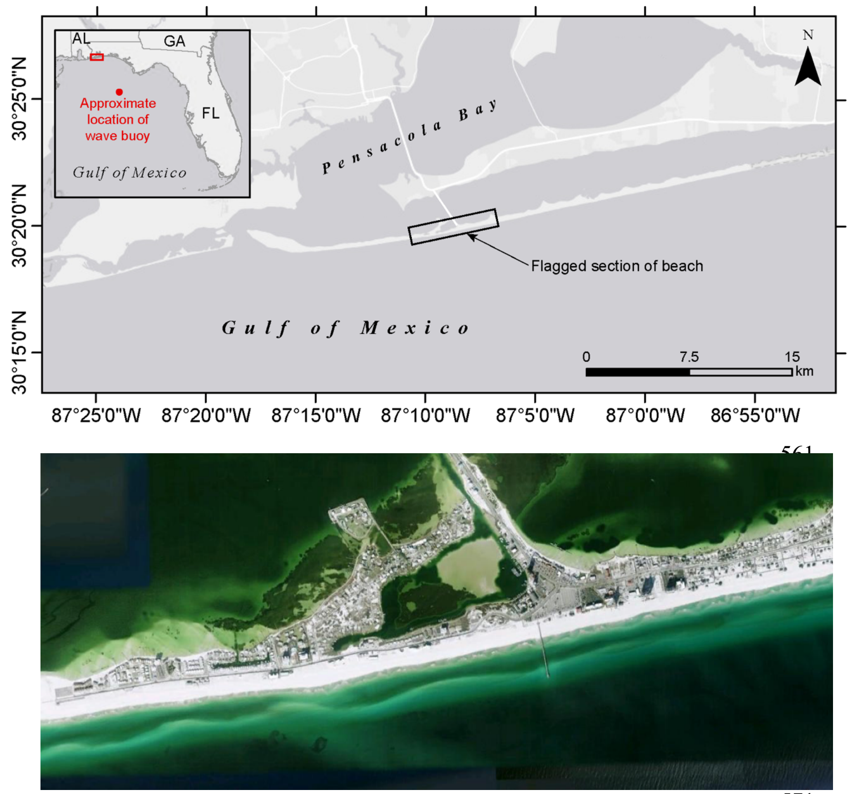


**Figure 1.** Map of study site showing location of flagged section of beach and approximate location of the wave buoy used in the analysis and for regional rip forecasts. Also shown is the presence of transverse-bar and rip morphology of the innermost bar and the variable nature of the outermost bar 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.

- 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……"*

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

That is correct.

- 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:

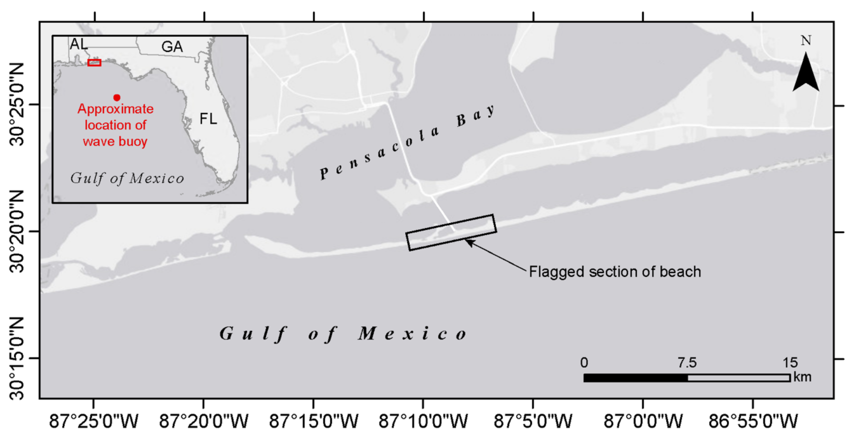

**Figure 1.** Map of study site showing location of flagged section of beach and approximate location of the wave buoy used in the analysis and for regional rip forecasts. Also shown is the presence of transverse-bar and rip morphology of the innermost bar and the variable nature of the outermost bar 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.

- 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).*

- 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.*

- 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.*

- 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.*

- 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."*

- 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."*

- 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.

**Machine Learning Analysis of Lifeguard Flag Decisions and Recorded Rescues**

Chris Houser, Jacob Lehner, Nathan Cherry, Phil Wernette

Faculty of Science,
University of Windsor,
401 Sunset Avenue,
Windsor, Ontario, Canada
N9C 2J9

Corresponding Author: chouser@uwindsor.ca

**Abstract**

Rips currents and other surf hazards are an emerging public health issue globally. Lifeguards, warning flags and signs are important and to varying degrees they are effective strategies to minimize risk to beach users. In the United States and other jurisdictions around the world, lifeguards use coloured flags (green, yellow and red) to indicate whether the danger posed by the surf and rip hazard is low, moderate, or high respectively. The choice of flag depends on the lifeguard(s) monitoring the changing surf conditions along the beach and over the course of the day using both regional surf forecasts and careful observation. There is a potential that the chosen flag is not consistent with the beach user perception of the risk, which may increase the potential for rescues or drownings. In this study, machine learning is used to determine the potential for error in the flags used at Pensacola Beach, and the impact of that error on the number of rescues. 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). Days when there is a difference between the predicted and posted flag colour represent only 17% of all rescue days but those days are associated with ~60% of all rescues between 2004 and 2008. Further analysis reveals that the largest number of rescue days and total number of rescues is associated with days where the flag deployed over-estimated the surf and hazard risk, such as a red or yellow flag flying when the model predicted a green flag would be more appropriate based on the wind and wave forcing alone. While it is possible that the lifeguards were overly cautious it is argued that they most likely 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. Results suggest that machine learning techniques have the potential to support lifeguards and thereby reduce the number of rescues and drownings.

**Keywords:** *rip current, surf zone, beach safety, beach hazard*

**Deleted:** does not accurately reflect
**Deleted:** the potential risk
**Deleted:** A
**Deleted:** suggests
**Deleted:** that the wrong flag was flown on ~35% of days
**Deleted:** , and that
**Deleted:** those differences account for only
**Deleted:** and
**Deleted:** the total number of
**Deleted:** would suggest
**Deleted:** Regardless whether this is a result of the
**Deleted:** being
**Deleted:** or the
**Deleted:** and surf hazard is associated with weak rips
**Deleted:** ,
**Deleted:** are
**Deleted:** the
**Deleted:** it isn't

## Introduction

Rip currents are the main hazard to recreational swimmers and bathers, and, in recent years, have been recognized as a 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). Rips are strong, seaward-directed currents that can develop on beaches characterized by wave breaking within the surf zone (Castelle et al., 2016), and are capable of transporting swimmers a significant distance away from the shoreline into deeper waters. Weak swimmers or those who try and fight the current can become stressed and experience panic (Brander et al., 2011; Drozdzewski et al., 2015) leading to increased adrenaline, an elevated heart rate and blood pressure, and rapid and shallow breathing. On recreational beaches in Australia and the United States, rips have been identified as the main cause of drownings and are believed to be responsible for nearly 80% of all rescues (Brighton et al., 2013; Brewster et al., 2019). It is estimated that the annual number of rip current drownings exceeds the number of fatalities caused by hurricanes, forest fires, and floods in Australia, the United States (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 evidence suggests that public 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).

**Deleted:** SLSA, 2017; USLA, 2017

**Deleted:** , and Costa Rica

**Deleted:** ;

**Deleted:** , but

**Deleted:** in press

Many beaches have warning signs at primary access points to warn beach users of the rip hazard, but recent studies suggest that signs may not be effective (e.g. Matthews et al., 2014; Brannstrom et al. 2015). Many beaches also use a combination of beach flags to either designate the location of supervised and safe swimming areas (e.g. Australia and the United Kingdom), or areas and times to avoid entering the water (e.g. Costa Rica and the US). Unfortunately, not every country uses the same flagging convention and there are regional variations that can lead to confusion amongst beach users. The United States and Canada use green, yellow, and red coloured flags to indicate whether the danger posed by the surf and rip hazard is low, moderate, or high, respectively (ILSF, 2004). A beach manager or lifeguard decides on the surf hazard and the flag colour to fly based on a combination of daily updates on rip conditions provided by local lifeguards as well as a rip forecast from the US National Weather Service (NWS). Most rip forecasts are based on a simple correlation between the number of rip-related rescues and meteorological and oceanographic conditions on that day (Lushine, 1991a, b; Lascody, 1998; Engle, 2002; Dusek and Seim, 2013; Kumar and Prasad, 2014; Scott et al., 2014; Moulton et al., 2017). These forecasts do not account for the surf zone morphology, which may be conducive to the development of rips on days when wave breaking is relatively weak. Even under 'green flag' days, the presence of shore-attached nearshore bars (called a transverse bar and rip morphology; Wright and Short, 1984) can force a current of ~0.5 m s$^{-1}$ that can pose a threat to weak swimmers (Houser et al, 2013).

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. To be effective, the flag system requires lifeguards to continuously assess surf conditions and monitor swimmers and bathers, and ultimately intervene if someone does not heed the warning implied by a yellow or red flag indicating moderate and high ('do not enter the water') hazard levels respectively. Recent evidence suggests that many beach users do not adhere to warnings if their own experience (whether accurate or not) or behavior of others on the beach, contradicts the hazard, as indicated by the warning flag (Houser et al., 2017; Menard et al., 2018). Beachgoers may lose trust in authority (i.e. the lifeguards) if a forecast is perceived, wrongly or rightly, to be inaccurate (Espluga et al., 2009). If the forecast is for dangerous surf conditions and a yellow or red flag is placed on the beach when 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. It can be difficult to change (or 'reset') public perception about the accuracy of the flag system as soon as a discrepancy is perceived, and subsequent visits and experiences may confirm the biases of the beach user (Houser et al., 2018). It is a situation analogous to the boy who cries "*wolf*" (Wachinger et al., 2013).

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. A decision tree, a form of machine learning, is used to predict the posted flag colour using lifeguard observations in combination with wind and wave forcing. The modelled flag colour, based solely on wave and wind forcing, can be compared

**Deleted:** The presence of a rip when the forecast predicts that the hazard potential is low, can put beach users at risk when a lifeguard is not present and able to intervene/rescue.

**Deleted:** the warning flag

**Deleted:** has

**Deleted:** een

**Deleted:** and

**Deleted:** overly-cautious

**Formatted:** Font: Italic

**Formatted:** Font color: Auto

**Deleted:** when daily data is available for flag colour, wind and wave forcing, as well as the number of rescues performed by lifeguards 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.

**Deleted:** posted flag colour

**Deleted:** discrepancy

**Deleted:** between the posted and predicted flag colours, which is, in turn, compared to the number o

**Deleted:** discrepancy

**Deleted:**

## Study Site

The analysis was completed at Pensacola Beach, Florida (Figure 1), where there is available records of daily flag colours, wind and wave forcing, and lifeguard-performed rescues between 2004 and 2008. The beaches of the Florida Panhandle have been described ''*as the worst in the nation for beach drowning*'' (The Tuscaloosa News, 2002), based on the presence of semi-permanent rips along the length of the island (Houser et al., 2011; Barrett and Houser, 2012). These rips can be active and pose a threat to swimmers when conditions may appear to be safe for swimming (Houser et al., 2013). During the period of the study (2004-2008), the Santa Rosa Island Authority maintained a flagging system to alert beach users about the heavy surf and rip hazard based on the NWS rip forecast. The highest flag colour for that day was recorded by the Island Authority, along with the number of prevents, assists, and rescues. The Island Authority reserve the rescue definition for those persons in extreme difficulty who, in the opinion of the lifeguard, would have drowned without assistance.

**Deleted:** for

**Deleted:**

**Deleted:** in an

Rescues, assists, and prevents are recorded regardless of whether they are conducted in a 'guarded' area, a designated swimming area where there are typically many beach users (Casino Beach, Fort Pickens Gate Beach, and Park East), or along the ~13 kms of unguarded beach where lifeguards conduct regular patrols and respond to emergency calls. As shown by Barrett and Houser (2013), there are rip current hotspots with semi-permanent alongshore variation in the nearshore morphology due to a ridge and swale bathymetry on the 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

**Deleted:** inner-shelf

the deeper swales (Barrett and Houser, 2012; see Figure 1). 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).

Santa Rosa Island experienced widespread erosion and washover during Hurricane Ivan in September 2004. The storm reinforced the alongshore variation in the nearshore bar morphology and forced the bars farther offshore. As described in Houser et al. (2015), the nearshore bars migrated landward and recovered to the beachface for 3 years following the storm. During this period, the inner-bar morphology transitioned from a rhythmic bar and beach morphology to a transverse bar and rip morphology before ultimately attaching to the beachface in May 2008 (Houser and Barrett, 2010). This changing bar morphology is a primary control on the presence of rip channels, with the greatest density of rips present in 2005 as the inner-most bar first started to develop a transverse bar and rip morphology (Houser et al., 2011).

**Deleted:** September,

## Methodology

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 direction, and the wind data included speed and direction. Local water level data was acquired from a station at the Port of Pensacola just north of the study site. A decision tree analysis was used to determine what combination of wave and wind forcing was associated with the flag posted by the Santa Rosa Island Authority on that day. After training on the available dataset, the model produces a decision tree that can be used for future decisions about what flag colour should be posted, although further training would be required to validate the model and operationalize. The modelled (*i.e.* predicted) flag colour is then compared to the posted flag colour for all days to determine if there is a relationship between the flag colour and the number of rescues. The

**Deleted:** s

**Deleted:** near the study region

**Deleted:** 42039 and 42040

comparison is also used to determine if there is a specific combination of wind and wave forcing on the days when the modelled flag colour and the posted flag colour do not align.

A decision tree model was developed using the Chi-square Automatic Interaction Detector (CHAID) technique developed by Kass (1980). 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 to explain the results in a given dependent variable. To develop a decision tree, the first step is declaring the root node, this corresponds to the target variable that will be predicted throughout the model. Then, the independent variable that provides the most information about the target values is identified. The root node is then split on this independent variable into statistically significant different subgroups using the F-test. These subgroups are then split using the predictor variables that provide the most information about them. CHAID analysis continues this process until terminal nodes are reached and no splits are statistically significant. 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).

**Deleted:** The

**Deleted:** .

## Results

The decision tree model was trained on the 1125 days with complete data between 2004 and 2008. Over this same period there were 145 days with rescues. 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. The number of rescues was at a minimum in 2007, while the number of rescue days was at a minimum in 2006 (Figure 3). The number of rescues decreased linearly between 2005 and 2007 as the nearshore bar morphology continued to recover following Hurricane Ivan and welded to the beachface consistent with previous observations at the site (Houser et al., 2011). 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

**Deleted:** during which there was

morphology conducive to the development of rip currents despite winds and waves typical of green flag conditions.

The decision tree analysis suggests that the posted flag colour was not predicted by the model on 35% of days between 2004 and 2008 (n=396). There was a total of 342 rescues over 66 days when the model predicted a different flag than was posted representing over 60% of all rescues (Table 1). By comparison, 40% of all rescues (n=224) occurred over 79 days when the predicted and posted flags were the same. Chi-square analysis suggests that the number of rescue days is significantly greater at the 95% confidence level when the predicted and posted flags are different ($\chi^2$=7.77, $\rho$~0.005). This supports the hypothesis that there are a greater number of rescues performed on days when there is a discrepancy between the predicted and posted flag colour.

**Deleted:** primary hypothesis that

**Deleted:** will be

**Table 1.** Results of Chi-square analysis of posted and predicted flag colour versus rescue and no rescue days at Pensacola Beach, Florida between 2004 and 2008.

| | Rescue Days | No Rescue Days | |
|---|---|---|---|
| **Posted=Predicted** | 79 | 650 | $\chi^2$=7.77, $\rho$~0.005 |
| **Posted≠Predicted** | 66 | 330 | |

Chi-square analysis was also used to determine if the number of rescue days depends on whether the model predicts a flag of greater or lesser hazard compared to the posted flag (Table 2). Results suggest that the number of rescue days is greater when the model predicts hazardous surf (i.e. red or yellow flag), but the posted flag was either yellow or green ($\chi^2$=18.11, $\rho$~0.0001). The number of rescue days was over-represented when the posted flag colour was red or yellow, but the model predicted that the flag should have been yellow or green, respectively, suggesting that posting what a beach user may perceive as an overly cautious flag can present a danger. These 47 days were associated with 268 of the total 566 rescues between 2004 and 2008, or ~7.2 rescues per day when the island authority posted a more cautious flag then was predicted by the model. 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).

**Deleted:** a

**Deleted:** overly-cautious

**Deleted:** was

**Deleted:** overly-cautious in their flag choice

**Table 2.** Results of Chi-square analysis of posted and predicted flag colour versus rescue and no rescue days at Pensacola Beach, Florida between 2004 and 2008.

| | Rescue Days | No Rescue Days | |
|---|---|---|---|
| **Posted>Predicted** | 47 | 171 | $\chi^2$=18.11, $\rho$~0.0001 |
| **Posted<Predicted** | 19 | 159 | |
| **Posted=Predicted** | 79 | 650 | |

The greatest number of rescues were performed on days when the posted flag was yellow (moderate hazard, moderate surf and/or currents), but the model predicted a green flag (low hazard, relatively calm surf and/or currents) based on the wind and wave forcing. 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. In comparison, there were only 12 rescues on 3 of 20 days when the posted flag was red (high hazard, strong surf and/or currents) and the model predicted flag colour was green. Finally, there were 25 rescues preformed on 7 of 30 days when a red flag was posted, and the model predicted a yellow flag was appropriate. The number of rescues and rescue days when the posted flag was more cautious than predicted by the model were at a maximum in 2005 and linearly decreased to a minimum in 2007 as the bar morphology recovered from Hurricane Ivan.

While there were fewer than expected rescue days when the posted flag was green or yellow and the model predicted a yellow or red flag, rescues were still performed on those days. There was a total of 66 rescues on 13 of 80 days when the posted flag was yellow, but the model predicted a red flag should be posted (Table 3). Only 7 rescues were performed on 5 of the 83 days when the posted flag was green and the model predicted a yellow flag, with even fewer rescues performed on days when the posted flag was green, but should have been red. The number of rescues and rescue days when the posted flag was lower than the predicted flag decreased from 2004 to 2007, with a statistically significant outlier in 2008. The large number of rescues in 2008 is the result of 2 days with 13 rescues each (April 19 and September 14), when a yellow flag was being flown, but the model predicted a red flag was more appropriate. This suggests that the difference between posted and predicted flag colours can vary inter-annually with changes in the nearshore morphology and/or changes in the individual who makes the flag decision.

**Deleted:** A

**Deleted:** yellow

**Deleted:** was

**Deleted:** .

**Deleted:** posted

**Deleted:** was appropriate

**Deleted:** annualy

**Deleted:** ¶

**Table 3.** Number of days and rescues (in brackets) based on the combination of posted and predicted flag colours.

| | | Predicted Flag | | |
|---|---|---|---|---|
| | | **G** | **Y** | **R** |
| **Posted Flag** | **G** | 475 (48) | 83 (7) | 15 (1) |
| | **Y** | 168 (231) | 154 (125) | 80 (66) |
| | **R** | 20 (12) | 30 (25) | 100 (51) |

## Discussion

Results of the present study suggest that over 60% of all rescues at Pensacola Beach, Florida between 2004 and 2008 occurred on days when the posted hazard flag was different from the flag colour predicted by a decision tree model. The posted flag colour was not predicted by the model on 35% of days between 2004 and 2008 (n=396), with one or more rescues occurring on 66 of those days (~17%). 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. This is not to suggest that Santa Island Authority made a mistake in their flag choice. 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.

**Deleted:** The model was trained using average wind and wave forcing at buoys offshore and the flag colour selected by the Santa Rosa Island Authority over the entire study period.…

**Deleted:** they

**Deleted:** disproportionately large number

**Deleted:** rs is

**Deleted:** of the innermost nearshore bar, which

**Deleted:** is not

**Deleted:** a model and forecast

**Deleted:** . The decisions made by the

**Deleted:** and

**Deleted:** lifeguards are not only dependent on the wind and wave forecast, but also their assessment of the morphology and the potential for rip development

**Deleted:** These d

**Deleted:** model-predicted and manager-

**Deleted:** Increasing the volume of available data into the future, through continuous collection, can broaden the information provided to the model, contributing to model evolution is better able to account for subtle distinctions while remaining computationally efficient. Furthermor

**Deleted:** e, i

**Deleted:** intuition associated with

**Deleted:**

The model predictions and most forecasts are based solely on wind and wave forcing (Lushine, 1991a, b; Lascody, 1998; Engle, 2002; Dusek and Seim 2013; Arun Kumar and Prasad, 2014; Scott et al., 2014; Moulton et al., 2017). Noticeably absent from the current model is surf zone morphology, which ultimately determines whether a rip can develop under those conditions or not. The beach manager and lifeguard can observe the nearshore morphology and assess the potential for rip development, which would lead to them putting out a yellow or red flag when the model would predict a green or yellow flag as being appropriate. While beach managers and lifeguards are being prudent, their assessment may not conform to those of the beach user who decides on whether the water is safe or not based on wave breaking conditions (Caldwell et al., 2013; Brannstrom et al., 2013; 2015). 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 green or yellow flag and 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 (Brannstrom et al., 2015) based on experience at the site or elsewhere (see Houser et al., 2018). Whether this causes beach users to lose confidence in the lifeguards and other authorities managing the beach is an important question for future research.

A large number of rescues occurred when the posted flag was yellow, but the model predicted the wind and wave forcing warranted a green flag. Rightly or wrongly, the beach user will observe that wave breaking is limited and assume that conditions must be safe. As shown by Caldwell et al. (2013) and Brannstrom et al. (2013) most beach users along the Gulf Coast of the United States assume that the calm flat water of a rip is safer than adjacent areas where the waves are breaking. The lifeguard, however, may observe a bar morphology that is conducive to the development of rips and post a yellow flag to warn about the potential for rips, despite the weak wind and wave forcing. As observed by Houser and Barrett (2012), rips with speeds of ~0.5 m/s can develop on 'green flag' days because of the transverse bar and rip morphology that is present in the inner-nearshore. 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

**Deleted:** is

**Deleted:** conditions

**Deleted:** was overly-conservative (*e.g.*

**Deleted:** was

**Deleted:** posted when the

**Deleted:** )

**Deleted:** erodes

**Deleted:** '

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. Going to the beach is a reward-based activity, and many people commit significant personal and financial investment to be at the beach (Houser et al., 2018). If they believe that the lifeguard is 'wrong' they will ignore the warning and remain committed to entering the water. The longer and more times that their perceptions are inconsistent with the experience and knowledge of the lifeguard, the more trust in authority is lost - a beach that is perceived to be safe based on experience will always be safe despite warnings to the contrary (Menard et al., 2018). This is an example of confirmation bias, in which an opinion quickly becomes entrenched and subsequent evidence is used to either bolster the belief or is rapidly discarded. How this can be addressed to reduce the number of rescues is an important focus for future research on rips and other hazards in general.

**Deleted:** .

**Deleted:** eroded

The results of this study also highlight the limitations of regional rip forecasts that are used in the United States and elsewhere around the world. A forecast based solely on the wind and wave forcing does not account for the nearshore morphology, which determines the potential for rip development. This raises one of the most important considerations for future modeling efforts based on machine learning techniques - the model will only be accurate if the bar morphology and conceptual knowledge of the lifeguard is included as input variables. Getting the beach user to observe and heed that forecast and warning, however, will remain a challenge.

**Deleted:** the

## Conclusions

Lifeguards and beach managers decide on warnings and flag colours based on careful monitoring of the changing surf conditions along the beach and over the course of the day using both regional surf forecasts and direct observation. A decision tree analysis predicts a flag colour different to the one flown on ~35% of days between 2004 and 2008 (n=396/1125), and that those differences account for only 17% of all rescue days and ~60% of the total number of rescues. The posting of a yellow flag when the model would predict a green flag based solely on the wind and wave forcing was found to be responsible for the largest number of rescues over the study period. Variables such as the nearshore morphology and the potential for rip development is not included

**Deleted:** The

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 based on direct observation of the bar morphology and experience, the beach user, like simple models based solely on meteorological data, may not believe that warning and still enter the water. This suggests that reducing the number of rip and surf rescues will require that we are able to address confirmation bias on the part of the beach user, which can cause them to lose their confidence in the lifeguards.

**Deleted:** erode their confidence

**Acknowledgements**

This study was partly funded through a NSERC Discovery Grant to CH.

## References


Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H.. A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides*, *11*(6), 1063-1078, 2014.

**Formatted:** Indent: Left: 0 cm, Hanging: 1.27 cm, Line spacing: Double

Arozarena, I., Houser, C., Echeverria, A.G., and Brannstrom, C.: The rip current hazard in Costa Rica, Natural Hazards, 77(2), 753-768, 2015.

**Formatted:** Line spacing: Double

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

Bielders, C.L., Ramelot, C. and Persoons, E. Farmer perception of runoff and erosion and extent of flooding in the silt-loam belt of the Belgian Walloon Region. *Environmental Science & Policy*, *6*(1), 85-93, 2003.

Brander, R.W., Bradstreet, A., Sherker, S., and MacMahan, J.: Responses of swimmers caught in rip currents: Perspectives on mitigating the global rip current hazard, International Journal of Aquatic Research and Education, 5(4), 11, 2011.

Brander, R.W., Dominey-Howes, D., Champion, C., Del Vecchio, O., and Brighton, B.: Brief Communication: A new perspective on the Australian rip current hazard, Natural Hazards and Earth System Sciences, 13(6), 1687, 2013.

Brannstrom, C., Trimble, S., Santos, A., Brown, H.L., and Houser, C.: Perception of the rip current hazard on Galveston Island and North Padre Island, Texas, USA, Natural Hazards. 72(2), 1123-1138, 2014.

Brannstrom, C., Brown, H.L., Houser, C., Trimble, S., and Santos, A.: “You can't see them from sitting here”: Evaluating beach user understanding of a rip current warning sign, Applied Geography, 56, 61-70, 2015.

Brewster, B.C., Gould, R.E. and Brander, R.W. Estimations of rip current rescues and drowning in the United States. *Natural Hazards and Earth System Sciences*, *19*(2), 389-397, 2019.

Brighton, B., Sherker, S., Brander, R., Thompson, M. and Bradstreet, A. Rip current related drowning deaths and rescues in Australia 2004–2011. *Natural hazards and earth system sciences*, *13*(4), pp.1069-1075, 2013.

Caldwell, N., Houser, C., and Meyer-Arendt, K.: Ability of beach users to identify rip currents at Pensacola Beach, Florida, Natural Hazards, 68(2), 1041-1056, 2013.

Castelle, B., Scott, T., Brander, R., and McCarroll, R.J.: Rip current types, circulation and hazard, Earth-Science Reviews, 163, 1-21, 2016.

Claudino-Sales, V., Wang, P. and Horwitz, M.H. Factors controlling the survival of coastal dunes during multiple hurricane impacts in 2004 and 2005: Santa Rosa barrier island, Florida. *Geomorphology*, *95*(3-4), 295-315, 2008.

Claudino-Sales, V., Wang, P. and Horwitz, M.H. Effect of Hurricane Ivan on coastal dunes of Santa Rosa Barrier Island, Florida: characterized on the basis of pre-and poststorm LIDAR surveys. *Journal of Coastal Research*, 470-484, 2010.

Drozdzewski, D., Shaw, W., Dominey-Howes, D., Brander, R., Walton, T., Gero, A., and Edwick, B.: Surveying rip current survivors: preliminary insights into the experiences of being caught in rip currents, Natural Hazards and Earth System Sciences, 12(4), 1201, 2012.

Dusek, G., and Seim, H.: A probabilistic rip current forecast model, Journal of Coastal Research 29(4), 909-925, 2013.

Engle, J., Macmahan, J., Thieke, R. J., Hanes, D. M., and Dean, R.G.: Formulation of a Rip Current Predictive Index Using Rescue Data. In *Proc. National Conf. on Beach Preservation Technology, FSBPA*, 2012.

Espluga, J., Farré, J., Gonzalo, J., Horlick-Jones, T., Prades, A., Oltra, C., and Navajas, J.: Do the people exposed to a technological risk always want more information about it? Some observations on cases of rejection. Safety, Reliability and Risk Analysis. Londres: CRC Press. Taylor & Francis, 1301-1308, 2009.

Fallon, K.M., Lai, Q. and Leatherman, S.P. Rip current literacy of beachgoers at Miami Beach, Florida. *Natural hazards*, *90*(2), 601-621, 2018.

Gallop, S.L., Woodward, E., Brander, R.W. and Pitman, S.J. Perceptions of rip current myths from the central south coast of England. *Ocean & coastal management*, *119*, 14-20, 2016.

Houser, C., Barrett, G., and Labude, D.: Alongshore variation in the rip current hazard at Pensacola Beach, Florida, Natural Hazards, 57(2), 501-523, 2011.

Houser, C., Arnott, R., Ulzhöfer, S., and Barrett, G.: Nearshore Circulation over Transverse Bar and Rip Morphology with Oblique Wave Forcing, Earth Surface Processes and Landforms, 38, 1269-1279, 2013.

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

Houser, C., Wernette, P., Rentschler, T., Jones, H., Hammond, B. and Trimble, S.: Post-storm beach and dune recovery: implications for barrier island resilience. Geomorphology, 234, 54-63, 2015.

**Deleted:** resilience, Geomorphology

Houser, C., and Barrett, G.: Divergent behavior of the swash zone in response to different foreshore slopes and nearshore states, Marine Geology, 271, 106-118, 2010.

Houser, C., Murphy, T., and Labude, D.: Alongshore correspondence of beach users and rip channels at Pensacola Beach, Florida, Natural Hazards, 78(3), 2175-2183, 2015.

Houser, C., Trimble, S., Brander, R., Brewster, B.C., Dusek, G., Jones, D., and Kuhn, J.: Public perceptions of a rip current hazard education program: "Break the Grip of the Rip!", Natural hazards and earth system sciences, 17(7), 1003, 2017.

Houser, C., Vlodarchyk, B. and Wernette, P. public interest in rip currents relative to other natural hazards: evidence from Google search data. *Natural Hazards*, 1-11, 2019.

Kass, G.V.: An exploratory technique for investigating large quantities of categorical data, Applied Statistics, 29(2), 119-127, 1980.

Kumar, S.A., and Prasad, K.V.S.R.: Rip current-related fatalities in India: a new predictive risk scale for forecasting rip currents, Natural hazards, 70(1), 313-335, 2014.

Lascody, R.L.: East central Florida rip current program, National Weather Digest, 22(2), 25-30, 1998.

Lushine, J.B.: A study of rip current drownings and related weather factors, In Natl. Wea. Dig, 1991.

Lushine, J.B.: Rip Currents: Human Impact and Forecastibility, In Coastal Zone'91, 3558-3569. ASCE, 1991.

Matthews, B., Andronaco, R., and Adams, A.: Warning signs at beaches: Do they work? Safety science, 62, 312-318, 2014

**Deleted:** 91, 3558

**Deleted:** work?.

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

Moulton, M., Dusek, G., Elgar, S., and Raubenheimer, B.: Comparison of Rip Current Hazard Likelihood Forecasts with Observed Rip Current Speeds, Weather and Forecasting, 32(4), 1659-1666, 2017.

Scott, T., Masselink, G., Austin, M.J., and Russell, P.: Controls on macrotidal rip current circulation and hazard, Geomorphology, 214, 198-215, 2014.

Silva-Cavalcanti, J.S., Costa, M.F. and Pereira, P.S. Rip currents signaling and users behaviour at an overcrowded urban beach. *Ocean & coastal management*, *155*, 90-97, 2018.

Surf Life Saving Australia (SLSA): National Coastal Safety Report 2017. Surf Life Saving Australia, Sydney, 2017.

Tehrany, M.S., Pradhan, B. and Jebur, M.N. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stochastic environmental research and risk assessment*, *29*(4), 1149-1165, 2015.

Trimble, S., and Houser, C.: Seawalls and Signage: How Beach Access Management Affects Rip Current Safety. In Beach Management Tools-Concepts, Methodologies and Case Studies, 497-524, Springer, Cham, 2017.

Vlodarchyk, B., Olivito, A. and Houser, C. Spatial and Temporal Variation of Surf Drownings in the Great Lakes: 2010–17. *Journal of Coastal Research*, *35*(4), 794-804, 2019.

Wachinger, G., Renn, O., Begg, C., Kuhlicke, C.: The risk perception paradox—implications for governance and communication of natural hazards, Risk analysis, 33(6), 1049-1065, 2013.

Wang, P. and Horwitz, M.H. Erosional and depositional characteristics of regional overwash deposits caused by multiple hurricanes. *Sedimentology*, *54*(3), 545-564, 2007.

Williamson, A., Hatfield, J., Sherker, S., Brander, R. and Hayen, A. A comparison of attitudes and knowledge of beach safety in Australia for beachgoers, rural residents and international tourists. *Australian and New Zealand Journal of Public Health*, *36*(4), 385-391, 2012.

Woodward, E., Beaumont, E., Russell, P., Wooler, A. and Macleod, R. Analysis of rip current incidents and victim demographics in the UK. *Journal of Coastal Research*, *65*(sp1), 850-856, 2013.

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

## **Figures**

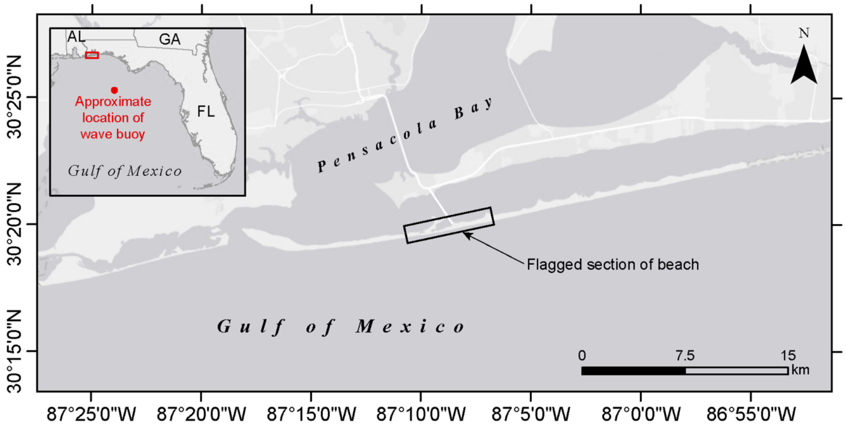


**Figure 1.** Map of study site showing location of flagged section of beach and approximate location of the wave buoy used in the analysis and for regional rip forecasts.

**Figure 2.** Satellite image of the flagged section of beach in 2004 (before Hurricane Ivan) showing 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 not necessarily representative of the nearshore morphology throughout the remainder of the study.

**Formatted:** Font: Bold

**Formatted:** Indent: Left: 0 cm, First line: 0 cm, Line spacing: single

**Formatted:** Indent: Left: 0 cm, First line: 0 cm

**Formatted:** Font: Bold

**Formatted:** Indent: Left: 0 cm, First line: 0 cm, Line spacing: single

**Formatted:** Indent: Left: 0 cm, First line: 0 cm

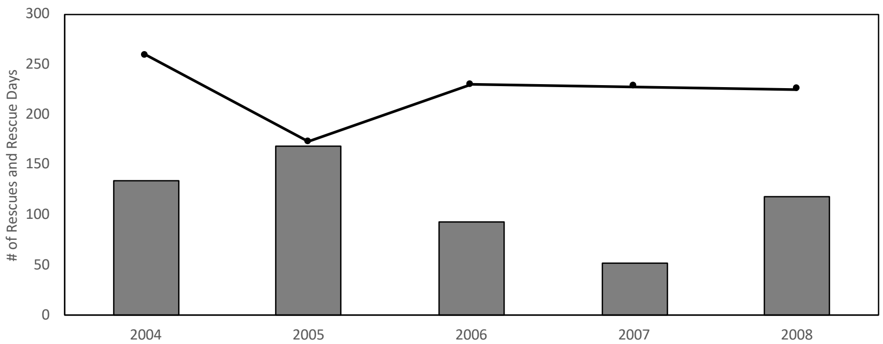


**Figure 3.** Interannual variation in number of rescues and rescue days at Pensacola Beach between 2004 and 2008.

**Formatted:** Justified, Indent: Left: 0 cm, First line: 0 cm, Line spacing: single

**Formatted:** Font: Bold