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.

Anonymous Referee #1

Received and published: 9 July 2019

General Comments This is a fascinating and important topic that attempts to try and qualify the accuracy of posted flag colours on Pensacola Beach in relation to the incidence of lifeguard rescues, something that has not yet been attempted in the literature (to my knowledge) and is clearly relevant to any beach using the same green, yellow, and red flag system (or similar). Such a study should be published, if for no other reason than to bring attention to this important topic – as the outcomes may affect loss of human life.

However, I do have some problems with the way the study is pitched that I think can be corrected with some modifications, and clarification and perhaps a more two-sided view. Some of the following comments are perhaps easier to deal with than others and are not presented in any particular order.

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. 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.

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.

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 ab-

solutely 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.

Or am I missing something here? There seem to be several cases of arguments that counter-act each other? Maybe shortening the results in order to provide greater clarity would help in this regard. There seems to be some repetition and over-explanation at times. I keep reading the results and getting confused.

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 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.

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.

Some specific and technical comments are provided below (some which relate to the general comments above): - Abstract L16 – perhaps specify '. . .risk to whom'

L18 – should be 'lifeguard(s)'

L22 – should be 'machine learning is used'

L24 – should be 'wrong colour flag'

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

L30 – should be 'surf hazard was associated'

L33 – should be 'lifeguards'

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

L42 – should be 'and are capable of'

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

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

studies

L61 – should be 'flag colour'

L69 – should be 'nearshore bars'

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

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?

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

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

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

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)

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?

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

L138 – should be 'flag colour'

L141 – should be 'number of rescues'

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

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'

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.

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).

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

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

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

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?

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).

L213 – should be spelled 'annually'

L237-239 – this sentence does not read properly

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

L251 – should be 'wave breaking conditions'

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