Major revisions cover letter re: Operational ecoforecasting for coral reefs using artificial intelligence and integrated near real-time environmental data

From: BMS 2022-0012 Decision Letter

Should you decide to revise the manuscript for further consideration, your revision should address the specific points made by each reviewer. With the revision you will need to submit a detailed (i.e., point-by-point including original reviewer comments) cover letter that fully explains how you addressed the reviewers' comments. Should you disagree with a reviewer on any substantial point, please include the reason for your disagreement.

Dear Editor:

Please consider for resubmission our manuscript entitled, “Operational ecoforecasting for coral reefs using artificial intelligence and integrated near real-time environmental data”. We would like to extend our thanks to the editor for the opportunity to resubmit our manuscript, as well as both reviewers for their helpful and insightful comments. We have addressed each comment individually in our responses bellow as well as in the updated text of our manuscript.

Signed,
Madison Soden, Lew Gramer, and James Hendee

07/2022 Bulletin of Marine Science Revisions

Reviewer #2: This manuscript describes an ecoforecasting tool developed to provide near-real-time warnings of potential coral bleaching events. The motivation and big picture are clearly described and results for a relatively small test set of data presented. However the manuscript does not include a detailed discussion of the methods necessary to evaluate or replicate the work, including selection of models, calibration of models, and extensibility of models to other places/applications. Some specific suggestions for improvement are provided below.

Introduction:

- A review of other environmental ecoforecasting approaches would be useful to supplement the discussion which is mainly focused on the list of available variables/signals. For instance, in what ways (quantitatively or qualitatively) do existing approaches fall short?

> The reviewer has rightly suggested the need for a more complete review of previous ecological forecast efforts for coral bleaching. We have expanded the text in our Introduction as follows, in order to address this excellent point:
> "Previous efforts to remotely monitor coral reefs and alert managers about the potential incidence of coral bleaching have been based on evaluations of satellite observed sea surface temperatures (SSTs; Liu et al. 2014). These alerts are spatial in character and based solely on temperatures at the surface of the ocean. Such an approach is not always optimized for effective, near real-time feedback for stakeholders, dockmasters, researchers and other users in the field. The tool described in the present study is designed to provide quick, usable feedback to enable users to adapt to generalized, potentially hazardous biological conditions on a specific spatial and temporal scale. Coral Reef Watch (Liu et al. 2014) is the most similar ecoforecasting system operating on the Florida Reef Tract; it is an early warning system based on current satellite SST and climatological conditions. It is a robust and helpful tool but the majority of its data products are created and reviewed on a monthly time scale which is not effective for the type of environmental damage mitigation and responsiveness that we are seeking to enable/motivate. "

- A discussion of the frequency of data required would be helpful to frame the need for the developed system, i.e., are these data needed minute-by-minute? Hour-by-hour? Day-by-day? (This would also give the reader context for later issues, such as the implications for dealing with misaligned time-stamps in data, especially from disparate data sources.)

The reviewer makes an excellent point. We have added the following to our Introduction to provide context for the reader in understanding the utility of the hourly in situ observations for our purpose here:
"Physical environmental extremes that affect marine ecosystems are directly impacted by the dominant scales of atmospheric and radiative forcing and ocean response. The extremely fine-scale, high-relief bathymetry found on barrier and fringing coral reef ecosystems results in very high-frequency physical environmental variability; hour-by-hour observations of both sea temperature and wind speed provide valuable context for forecasting the extremes that individual reefs may encounter, and therefore for improving the accuracy and extensibility of bleaching ecoforecasts."

Line 77: It would be helpful if the authors could provide some justification for the choice of fuzzy logic to underpin this system, as a wide range of computational methods can ingest disparate data streams and assign relative importances or identify dependencies in relationships. It would seem that representing and propagating uncertainty in a quantitative way is also important, but this as well can be achieved with other approaches such as Gaussian models, neural networks, random forests, etc.

Reviewer 2 makes a valid point questioning the choice to use heuristic methodologies as the bases of our monitoring software. This was a choice primarily based on pragmatism for creating a useful, efficient monitoring tool rather than a highly specific model. In the future, we would like to expand upon our software to include additional, more quantitative options to propagate uncertainty and represent ecological conditions. We have included the following text below Line 77 to expand upon this further: "There is not a high enough temporal or spatial resolution of validation data, or coral bleaching monitoring, to accurately train a neural network. Gaussian models and random forest models tend to be very time consuming to run in a near real time capacity particularly if you are looking at data from several different sites individually. The fuzzy logic generation is site specific and this tool is intended for use in monitoring large regions, hopefully with a large amount of monitoring locations. Ultimately these techniques were ill suited to our particular goals in developing a flexible, extensible tool to be used in near real-time monitoring situations."

Methods:

Line 138: It would be helpful to understand generation of the "fuzzy intensity" value in more specific detail. From the description it sounds like the historical data distribution is used to model variability and this is used to essentially add (fuzzy) uncertainty to the measured value? This is never fully explained in the manuscript.

We thank the review for pointing out the need for additional clarification. Conceptually speaking, (fuzzy) uncertainty is 'added' to the original intensity value but that is not the literal process we go through to create fuzzy uncertainty. We have added the following text after line 138 to address the authors concerns, and provide more of an example:

“These fact factories are generated, first, by calculating these corresponding percentile brackets for each variable/station combination: [00.62, 02.27, 06.68, 15.87, 30.85, 69.15, 84.13, 93.32, 97.72, 99.38]. Secondly, the fact-factory tables are further refined based on bleaching observations, and the false negatives/false positives alerted to over the training period. For example, say that after reviewing the alerts generated by the training data, multiple false positive, high temperature alerts are generated for the SANF1. After reviewing the station and the averaged temperature integer values, it may be found that a 95.32 - 97.72 percentile range is better suited than a 93.32 - 97.72 range for generating the corresponding ‘very-high’ heuristic value.”

Lines 151-157: How were these multipliers determined? Why 2 and 2.5 (as opposed to for instance 3 and 5, or any other two values)?

We thank the reviewer for pointing out that the rationale for using these multipliers was not well explained. We have addressed this by providing the following edited text in the MS.:

This numeric measure, the Stimulus/Response Index (S/RI), is calculated for each day as a count of three times the number of 3-hour periods for each contributing variable that matches the corresponding ecological forecast criteria for that variable (Hendee et al. 2009). During periods when particular physical variables have extreme enough values (stress, or stimulus) to suggest a qualitatively greater ecosystem response, the S/RI associated with those variables is multiplied by a factor of 2 before being added to the daily total; where data suggest a particularly severe response, the S/RI is multiplied by a factor of 2.5 to reflect the rather rare but significant contributing environmental factor. For a more thorough explanation of the reasoning behind the S/RI concept, the appointing of points per production rule fired, and the multipliers, please see Hendee et al 2009.

Line 161: These examples are helpful for illumination, however the reader does not have context for what the concern factors are for each variable. Could the authors add a table with a list of each variable and what would ultimately define "conditions of concern" for this variable? Skipping forward to Line 255 - it was not clear at Line 161 that these conditions were not global but were instead derived site-by-site. Can the authors speak to the generalizability of the model if site-specific conditions have to be identified everywhere?

We thank the reviewer for their insightful questions. We address the need for additional information on the “conditions of concern” in our updated text, with the following paragraph and reference:

“Due to the nature of heuristic programming, the S/RI is subjective and used to indicate the cumulative severity of the time and intensity of multiple variables that contribute to a rule firing. The context under which we assign different levels of severity to environmental values in constructing fuzzy intensities, rules, and our S/RI are also discussed in an earlier work (Hendee et al. 2008). That earlier methodological work also provides a set of tables and a figure to clarify this approach." .

Line 166: What criterion or decision process was used to set the "training" versus "validation" periods?

We thank the reviewer for pointing out that we did not clarify our rationale for choosing the training period. We have added the following text to the MS. to provide this clarification:
"The record of physical measurements covering multiple sub-region of the Keys began in 1991. The years 1997 and 1998 were years with severe bleaching at many sites. These thus represented important "positive signals" to include in the training period. Bleaching was again observed to be relatively severe in 2005 and later years in the in situ record: for this reason, and to demonstrate the broader applicability of the method, we choose 2005-2017 as our validation period."

Line 167-168: Please provide more information on the granularity of data that were available in the literature. For instance did any provide specific days or weeks or were all sources only months or years in which a bleaching event was experienced?

We thank the reviewer for pointing out the need to specify more information on bleaching observations. We have added the following text to the MS. to address this:
"All sources for bleaching observations were based on diver reports collected from a community of professional scientists and dive operators, utilizing Reef Check, SECREMP (2020), and AGRRA (2022) monitoring protocols. The granularity of these data was site- and species-specific and reports were gathered at monthly frequency."

Line 189: Combined with comment above, although it makes sense to choose a common data resolution, it is not clear that the data for an entire year during which a bleaching event occurs will correlate with a bleaching event - and further the authors motivate the paper with the need to have near-real time data to get information to scientists and managers on a days (< week) time scale. Is there any way to identify within years where bleaching events occurred a more granular level of when actual events occurred so that the system could be trained on data that actually precede or relate to the events? What would this timescale be and can the authors justify that choice? Depending on result of this analysis, it would be helpful to see an update to Figure 2 to provide data more granularly than annual.

We thank the reviewer for pointing out that better time and space granularity of validation data would be preferable. As we address in the new text below, such validation data are generally not available for the sort of ecosystem response we address with this system. Therefore, we adopted a pragmatic modeling approach in the design of the system described in this MS. As we now describe in the text:

"The choice to use expert systems was based on pragmatism--there is not a high enough temporal or spatial resolution of validation data, or coral bleaching monitoring, to accurately train a neural network. Gaussian models and random forest models tend to be very time consuming to run in a near real time capacity particularly if you are looking at data from several different sites individually. The fuzzy logic generation is site specific and this tool is intended for use in monitoring large regions, hopefully with a large amount of monitoring locations. Ultimately these techniques were ill suited to our particular goals in developing a flexible, extensible tool to be used in near real-time monitoring situations."

In this MS., however, we wished to examine bleaching reports at the highest frequency available in the long term record. We therefore calculated summed S/RIs over each bleaching season, to simulate the verification of daily alerts to managers through these monthly observations. It was our intention with the paragraph centered on line 189, to describe this approach. We have edited this paragraph to more clearly outline the rationale for our approach.

Line 211-212: The process described here is not exactly clear but may potentially be addressing the point above in some way. From the way this sentence is written it sounds like only the outlier values were used but the calculation was done across the entire evaluation period, i.e., that the outlier values were copied onto all other days in the period being evaluated? If that is not the intended interpretation of the sentence, it would help if the authors could provide more detail on this calculation and justification for this approach.

We thank the reviewer for correctly pointing out that this paragraph was unclear. We have modified it as follows to clarify what our actual approach was: "The concept of a Stimulus/Response Index was first implemented and discussed in Hendee et al 2009, and has been implemented in NEIS. We examined the averages of hourly environmental data for the entire physical record from each of four Florida C-MAN stations, including the 3-hour rolling average sea temperature and wind speed, 3-day average wind speed, and 30-day average sea temperature. During certain summers, we found extreme values in the averaged sea temperature (highs), wind speed (lows), or both at each lighthouse that were outliers within the multiyear record for that site (Fig. 2). From these outliers, we calculated a total S/RI for each day of the evaluation period during which the outliers occurred, and summed the S/RI values over each year (Fig. 3)."

Line 170-171: Please provide more insight into how these limits were tightened or loosened - was this done manually (brute force search)? Was an algorithmic approach employed to identify the combination of conditions that predicted the observed bleaching events?

We thank the reviewer for pointing out that more detail on this methodology was needed. We have added that detail to our Methods, as described in the commentary noted for line 138 (original manuscript) above.

Line 233: This paragraph has different text formatting.

We thank the reviewer for catching this mistake. We have updated this paragraphs formatting.

Figure 1: Suggest making this a single panel wide and 8 panels tall as the trigger dots are very difficult to discern. Addition of a legend would be helpful.

We thank the reviewer for pointing out that Figure 2 (which was previously Figure 1) was difficult to interpret. We have added annotation ellipses to Figure 2, to highlight periods during which we identified triggers for bleaching within the original in situ physical data. We hope this annotation is helpful to the reviewer and ultimately to our readers in understanding our methods and results.

Discussion: It would be informative for the authors to outline what would be needed in terms of systems (more sensors? how many? more observations? how many?) to enable use of the tool they developed at the timescales envisioned in the intro/motivation (<week for alerts to scientists) - or if this is already enabled from the system, it would be helpful to understand more about how the authors convert data averaged or summed over entire years to the appropriate values to be used in day-to-day forecasting.

We thank the reviewer for this cogent suggestion. The goal of the study described in the MS. was to take the highest available spatiotemporal resolution of data for both the physical environment (C-MAN data) and the ecological response (reported coral bleaching), and to match those to produce alerts for managers, scientists, and the public which were as timely as possible given those data. The MS. is intended to describe how well we achieved that goal; however, in response to this comment, we have also added some information to the Discussion on a novel implementation of this system that is currently being used to monitor potential turbidity impacts in another region of the Florida reef tract. This additional text includes the following paragraph, which we hope addresses the reviewer's suggestion:

"Although testing of this Python-based newer version of the original CLIPS-based system (Hendee 1998) precluded the use of such maintenance blogs, we feel timely use of feedback from the field--whether via blogs or other chronicling of the changes to the instrumentation and/or environment--is absolutely essential to a successful deployment of instrumental arrays with NEIS as the ecological forecasting component (e.g., see Fletcher et al 2022).

Such a system is likely to be integrated into future adaptations of NEIS, for example, in monitoring turbidity and sedimentation at reef sites impacted by ongoing human activities in Florida. What must be kept in mind is that correct ecoforecasts are only as good as, a) the precision of the instruments, which means regular cleaning, maintenance, and if necessary, replacement of them, and, b) timely feedback from the field as to whether or not (or to what degree) the ecoforecasts were correct, thus permitting the necessary fine-tuning of the thresholds within the rule-based system.

Based on our successes with open-source components and the methods presented in this study, we are collaborating with partners in NOAA operational (non-research) line offices and other agencies to expand this system to daily operations that cover coral reefs and other sensitive marine ecosystems in disparate locations. We expect the system to ultimately find applications in nowcasting harmful algal blooms, upwelling, and enhanced turbidity related to human activities. However, so long as, a) the basic environmental influences of a particular phenomenon are known, b) the instruments are available to measure those environmental variables, and, c) there is in place a reliable feedback system of maintenance recording and field observations of the phenomenon in question, the system described herein should find applicability in a multitude of ecological forecasting events."