**Q 1**

**Please summarize the main findings of the study.**

**Reviewer 1** | 09 Aug 2020 | 03:33

**#1**

This paper has used four machine learning model to predict the Karenia Brevis blooms using river discharge, nutrient concentration, wind, and temperature.  
The prediction accuracy was assessed for all the tested models and the better ones were applied to estimate the environmental driver factors.

**Q 2**

**Please highlight the limitations and strengths.**

**Reviewer 1** | 09 Aug 2020 | 03:33

**#1**

This is an interesting topic to investigate the probability of predicting Karenia brevis blooms.  
However, I am concerned on the prediction accuracy. The overall F1-score is only ~0.5, which is no better than random guess. In one previous study (cited in the paper, Hill et al., 2019), the accuracy can be ~91%. Therefore, I think the authors needs to justify why this prediction is meaningful.

*Reply:*

*Hill et al. (2019) utilizes satellite inferences of chlorophyll a concentrations to detect Karenia brevis blooms. Chl a acts as a proxy of algal activity and thus the blooming seasonal HAB species such as K. brevis, so it is unsurprising this detection method achieves a high accuracy. In this paper we adopt a different approach and predict the blooms from environmental conditions, e.g. river flow, nutrient loading, and wind speed/direction. We added to the explanation of the difference between our model and Hill et al. (2019). Please see lines 508-516.*

**Q 3**

**Please comment on the methods, results and data interpretation. If there are any objective errors, or if the conclusions are not supported, you should detail your concerns.**

**Reviewer 1** | 09 Aug 2020 | 03:33

**#1**

Reviewer 1: Overall, the prediction accuracy from all the four methods seems too low. The F score is only 0.5. From a simple experiment, for the unbalanced dataset, if we predict that all the years are blooms, we can have 3/7 precision and 3/3 recall, which results in a F1 score of ~0.6.

*Reply: In the unbalanced classification problem with 318 HAB events and 755 non-HAB events, there would be an accuracy A=0.3, recall R=1, precision = 0.3 and F1=0.46 if all events are predicted to be blooms. The problem with this prediction is the low accuracy and large amount of false positives. The accuracy obtained from the four machine learning algorithms is 0.79 from SVM, 0.62 from RVM, 0.52 from NB and 0.74 from ANN, according to the nested k-fold cross validation, respectively.*

*The three metrics, accuracy, recall and precision, have clear and intuitive interpretations. The fourth metric F1 is a derived mathematical quantity from recall and precision and is harder to interpret. As shown in the above example, F1=0.46 if all events are predicted to have HABs. Few would consider this as a good model because it would give too many false alarms to coastal managers. However, F1 is not low because the recall value is artificially inflated (R=1). For this reason, we have decided to remove F1 from the text. Please see lines 354-358 in the revised manuscript.*

If we assume that the samples are balanced (50% of the samples are blooms and 50% are non-blooms), then if we predict that all the years are blooms, the precision is 0.5 and the recall is 0.5 and the F1 score is 0.5.

*Reply: The HAB and non-HAB samples are not balanced. This is an unbalanced classification problem.*

This is already much higher than the best prediction results using the RVM and NB model (which are only 0.4 or 0.5). So therefore, I am unsure if the prediction is useful. The authors should provide explanations on this.

*Reply: We would like to point out the higher accuracy from the machine learning algorithms and relatively high recall from RVM (0.73) and NB (0.85). In our view, the most important metric for bloom prediction is recall, the percentage of HAB events that are predicted correctly. These predictions can warn coastal resource managers who have to close shellfish beds and alert beachgoers of potential exposure to aerosolized toxins. The recall is 0.73 for RVM and 0.85 for NB, according to the nested k-fold cross-validation with random oversampling of the minority class (see column 1 in Table 1).*

Given the low accuracy, I was at first confused why Fig.3&4 appear to be very consistent between measurement and the model prediction. If that is true, it seems unlikely that the F1-score could be that low. Later on, I realized that the plotted green dots in Fig 3&4 are only the correct predictions. The authors should emphasize that in the Figure caption. Right now it is not mentioned in the Figure caption of Fig 3&4, and this could be very misleading to the readers. The authors should also put in the unsuccessful predictions in this time series analyses. It is not meaningful to only look at the correct predictions.

*Reply: Thank you for your suggestion! We have changed the color scheme and added orange circles to mark incorrect predictions in the revised Figures 3 and 4.*

**Q 4**

**Check List**

**Reviewer 1** | 09 Aug 2020 | 03:33

**#1**

Is the English language of sufficient quality?  
- Yes  
  
Is the quality of the figures and tables satisfactory?  
No answer given.  
  
Does the reference list cover the relevant literature adequately and in an unbiased manner?  
No answer given.  
  
Are the statistical methods valid and correctly applied? (e.g. sample size, choice of test)  
No answer given.  
  
Are the methods sufficiently documented to allow replication studies?  
No answer given.  
  
Are the data underlying the study available in either the article, supplement, or deposited in a repository? (Sequence/expression data, protein/molecule characterizations, annotations, and taxonomy data are required to be deposited in public repositories prior to publication)  
No answer given.  
  
Does the study adhere to ethical standards including ethics committee approval and consent procedure?  
No answer given.  
  
Have standard biosecurity and institutional safety procedures been adhered to?  
No answer given.

Add comment

**Q 5**

**Please provide your detailed review report to the editor and authors (including any comments on the Q4 Check List):**

No answer given.

|  |  |
| --- | --- |
|  |  |