**EVALUATION**

**Q 1**

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

**Reviewer 2** | 25 Sep 2020 | 15:22

**#1**

Four machine learning algorithms were applied, calibrated and then their skill in predicting Karenia brevis blooms was compared to each other, based on 20 years of monitoring data.  
It was concluded that RVM and NB have better skills in bloom prediction than the other two approaches (SVM and ANN) according to k-fold and block cross-validation analyses, however all models produced a similar amount of false negative results. The trained models were applied to investigate the importance of winds, upwelling and river nutrient supply. Necessary quantitative reductions of N and P input from the large rivers were estimated to reduce blooms.

**Q 2**

**Please highlight the limitations and strengths.**

**Reviewer 2** | 25 Sep 2020 | 15:22

**#1**

The principal idea and approach to address bloom prediction using machine learning algorithms is for sure valid and should be use to complement resource intense biogeochemical modeling approaches. One problem is the relative short period of available data (20 years) that hampers the development of models with predictive skill (especially with regard to neural networks, that relying on many data).

*Reply: We agree with the reviewer that a longer data set would be better for the training of machine learning algorithms, particularly for artificial neural networks. However, this 20-year data set has the most complete monitoring of K. brevis blooms over the entire West Florida shelf. In total we have 318 HAB events and 755 non-HAB events/weeks, which should still provide a decent data set for training the machine learning algorithms.*

**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 2** | 25 Sep 2020 | 15:22

**#1**

I have however one major concern with the (description of the) methodology applied here for developing/training the model.  
  
The authors deal with time series data but decide to apply a k-fold cross validation approach for the training of the different algorithms.  
  
However, time series data are ordered data, and care must be taken in splitting the data in order to prevent data leakage. In order to accurately simulate the forecasting environment, in which we are in the present and forecast the future, the forecaster must withhold all data about events that occur chronologically after the events used for fitting the model. This is because it is not the same to forecast 2 years in the midst of a time series or the 2 coming years. So, rather than use k-fold cross-validation, for time series data we must use hold-out cross-validation where a subset of the data (split temporally) is reserved for validating the model performance.

Using the “block cross-validation” method as very briefly described in the paper might have done this. However from that description it is not clear if it was ensured that always observations from the training set occur before the test set.

*Reply:*

*The block cross-validation method was used to address the auto-correlation issue in time series dataset. To ensure independence between cross-validation folds in the time series data, blocks are drawn in chronological order to produce time-slice blocks of similar duration. We have added a more-detailed description of the block cross validation method and provided references. Please see lines 326-329.*

An appropriate method for time series is “nested cross validation” which provides a nearly unbiased estimate of the true error. Especially if you intend to develop a model for forecasting a type of “Time Forward-Chaining” methodology should be used.  
  
The inappropriate use of K-fold cross validation should be removed from the paper. The used method must be explained clearly with respect how training and test data were selected.

*Reply:*

*We thank the reviewer for the suggestion. In fact, we did use the nested k-fold cross-validation and should have emphasized it in the original text. The R-library we used to find the hyperparameters already implemented the nested k-fold cross-validation approach. We have added text to clarify this and revised Table 1. Please see lines 317-323.*

*Since both the nest k-fold cross-validation and block cross-validation are reasonable methods for the time series data, we would like to keep the statistics from both methods in Table 1. We have revised the name from k-fold to nested k-fold in the table.*

**Q 4**

**Check List**

**Reviewer 2** | 25 Sep 2020 | 15:22

**#1**

Is the English language of sufficient quality?  
- Yes  
  
Is the quality of the figures and tables satisfactory?  
- Yes  
  
Does the reference list cover the relevant literature adequately and in an unbiased manner?  
- Yes  
  
Are the statistical methods valid and correctly applied? (e.g. sample size, choice of test)  
- No  
  
Are the methods sufficiently documented to allow replication studies?  
- No  
  
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)  
- Yes  
  
Does the study adhere to ethical standards including ethics committee approval and consent procedure?  
- Not Applicable  
  
Have standard biosecurity and institutional safety procedures been adhered to?  
- Not Applicable

Add comment

**Q 5**

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

**Reviewer 2** | 25 Sep 2020 | 15:22

**#1**

Review of: Machine Learning Classification Algorithms for Predicting Karenia brevis Blooms on the West Florida Shelf  by Li and Glibert  
  
Main achievements  
  
Four machine learning algorithms were applied, calibrated and then their skill in predicting Karenia brevis blooms was compared to each other, based on 20 years of monitoring data.

It was concluded that RVM and NB have better skills in bloom prediction than the other two approaches (SVM and ANN) according to k-fold and block cross-validation analyses, however all models produced a similar amount of false negative results. The trained models were applied to investigate the importance of winds, upwelling and river nutrient supply. Necessary quantitative reductions of N and P input from the large rivers were estimated to reduce blooms.  
  
The principal idea and approach to address bloom prediction using machine learning algorithms is for sure valid and should be use to complement resource intense biogeochemical modeling approaches. One problem is the relative short period of available data (20 years) that hampers the development of models with predictive skill (especially with regard to neural networks, that relying on many data).

*Reply:*

*We agree with the reviewer that a longer data set would be better for the training of machine learning algorithms, particularly for artificial neural networks. However, this 20-year data set has the most complete monitoring of K. brevis blooms over the entire West Florida shelf. In total we have 318 HAB events and 755 non-HAB events/weeks, which should still provide a decent data set for training the machine learning algorithms.*

I have however one major concerns with the (description of the) methodology applied here for developing/training the model.  
  
The authors deal with time series data but decide to apply a k-fold cross validation approach for the training of the different algorithms.  
  
However, time series data are ordered data, and care must be taken in splitting the data in order to prevent data leakage. In order to accurately simulate the forecasting environment, in which we are in the present and forecast the future, the forecaster must withhold all data about events that occur chronologically after the events used for fitting the model. This is because it is not the same to forecast 2 years in the midst of a time series or the 2 coming years. So, rather than use k-fold cross-validation, for time series data we must use hold-out cross-validation where a subset of the data (split temporally) is reserved for validating the model performance.  
  
Using the “block cross-validation” method as very briefly described in the paper might have done this. However from that description it is not clear if it was ensured that always observations from the training set occur before the test set.

*Reply:*

*The block cross-validation method was developed to address the auto-correlation issue in time series dataset. To ensure independence between cross-validation folds in the time series data, blocks are drawn in chronological order to produce time-slice blocks of similar duration. The method is fairly straightforward. We have added a more-detailed description of the block cross validation method and provided references. Please see lines 326-329.*

An appropriate method for time series is “nested cross validation” which provides a nearly unbiased estimate of the true error. Especially if you intend to develop a model for forecasting a type of “Time Forward-Chaining” methodology should be used.  
  
The inappropriate use of K-fold cross validation should be removed from the paper. The used method must be explained clearly with respect how training and test data were selected.

*Reply:*

*We thank the reviewer for the suggestion. In fact, we did use the nested k-fold cross-validation and should have emphasized it in the original text. The R-library we used to find the hyperparameters already implemented the nested k-fold cross-validation approach. We have added text to make it clear and revised Table 1. Please see lines 317-323.*

*Since both the nest k-fold cross-validation and block cross-validation are reasonable methods for the time series data, we would like to keep the statistics from both methods in Table 1. We have revised the name from k-fold to nested k-fold in the table.*