Modelling Phytoplankton Behaviour in the North and Irish Sea with Transformer Networks

Onatkut Dagtekin *[1] and Nina Dethlefs[1]

[1]School of Computer Science and Technology, University of Hull, United Kingdom

Abstract

Climate change will affect how water sources are managed and monitored. Continuous monitoring of water quality is crucial to detect pollution, to ensure that various natural cycles are not disrupted by anthropogenic activities and to assess the effectiveness of beneficial management measures taken under defined protocols. One such disruption is algal blooms in which population of phytoplankton increase rapidly affecting biodiversity in marine environments. The frequency of algal blooms will increase with climate change as it presents favourable conditions for reproduction of phytoplankton. Machine learning has been used for early detection of algal blooms previously, with the focus mostly on single closed bodies of water in Far East Asia with short time ranges. In this work, we study four locations around the North Sea and the Irish Sea with different characteristics predicting activity with longer time-spans and explaining the importance of the input with regard to the output of the prediction model. This work aids domain experts to monitor potential changes to the ecosystem over longer time ranges and to take action when necessary.

1 Introduction

Harmful algal blooms (HABs) occur when the population of phytoplankton increases rapidly, causing environmental changes such as sunlight blocking and oxygen depletion [15]. These changes affect the ecosystem as well as public health since consumption of aquatic life affected by these blooms pose a health risk [11]. In some cases HABs occur due to eutrophication caused by nutrient overload which may result in further ecosystem disruption. Occurrence of eutrophication involves the creation of oxygen deprived zones due to the extreme number of deceased plants and animals, resulting in dead zones with no ability to support life and require external action to restore the habitat [6].

With the increasing temperatures due to climate change, it is expected that the frequency of algal blooms will increase and will be seen in new regions [28]. In addition to the ecological impacts, occurrence of algal blooms has negative economical impacts. These include drinking water treatment costs and increase to the cost of preservation of biodiversity [10]. Regions where these blooms are frequent see lower sales in sectors related to tourism and lower income from fisheries [2, 16].

To prevent this phenomena from occurring, preventive measures could be taken which includes early detection models that benefit from in-situ data and harness the power of machine learning.

Modelling algal blooms has several challenges. Algal blooms are extreme events, therefore positive labelled samples are extremely low (3-5%) in the dataset which needs to be addressed during training with methods such as SMOTE or label weighting and model evaluation with weighted F1 score. Deep learning models require vast amounts of data for training which is solved with continuous and frequent monitoring. The occurrence of algal blooms is inherently complex as the underlying mechanism is influenced by many factors such as nutrient intake of nitrates and phosphates through industrial pollutants or fertilizers, the water temperature and available light.

In this work, we propose a new model that improves the detection of abnormal activities in certain locations of the North Sea and the Irish Sea using in-situ data and a flexible labelling method with

*Corresponding Author: O.Dagtekin-2019@hull.ac.uk

varying ranges of detection and a longer range of time which was not taken into account in the majority of the approaches, with transformer networks and convolution operations. Our approach generates a possible sequence at day $x + i$, i ranging from 1 to 7, using observations at day x with a representation learning approach and filtering the necessary parts of the generated sequence to predict a label. In addition, we explain the reasoning behind the predictions using SHAP to aid experts in understanding the effects of observations. The scope of this work aims to detect the beginning of these blooms due to mechanics of the phenomenon. We have observed that using a representation learning approach results in a better model, performing 2.7 times better than current methods for this study area.

2 Related Work

The majority of approaches apply thresholding to categorize labels and forecast future behaviour or apply regression to the problem of HAB detection using dissolved oxygen or chlorophyll-a (chl-a) as the target variable, both of which increase with higher photosynthetic activity from aquatic plants or algae, as chlorophyll-a is used to capture sunlight and carry out photosynthesis to produce oxygen and glucose. The chl-a concentration will increase during an algal bloom due to increased photosynthetic activity whereas the oxygen concentration will increase initially with high photosynthetic activity and drop afterwards due to increasing decomposer population. It should be noted that the behaviour of inland waters and seawater differ from one another as seawater bodies can act like large reservoirs so they are less susceptible to change.

The detection time-spans of the current approaches are usually short ranging from 12 hours to 4 days. [27] use temporal attention combined with Long Short-Term Memory (LSTM) to predict the chl-a value at most 12 hours ahead in Fujian, China. [24] predict the chl-a value 1 to 3 days ahead, using a combination of an ensemble of Artificial Neural Networks (ANNs) with Discrete Wavelet Transform. [7] use sensory data to predict the chl-a in certain locations in South Korea with LSTMs. They aimed to predict the chl-a concentration a day ahead and 4 days ahead using this approach. [20] compares ANN, generalized regression network and Support Vector Machines (SVM) in the context of predicting chl-a values 7 or 14 days ahead for Tolo Harbour, Hong Kong. [31] uses Extreme Learning Machine to predict chl-a values 7 days ahead along several weirs on the Nakdong River, South Korea.

The most common approaches lean towards using Random Forests (RFs), SVMs and ANNs to predict algal blooms. [29] use RF to predict the ch-a concentration in Urayama Reservoir and Lake Shinji, Japan. [30] use sensory data to predict HABs using AdaBoost with SVM and RF in Yuyuantan Lake, China. [9] use ANNs combined with correlation and feature selection to predict the dissolved oxygen value in Lake Juam, South Korea. [32] predicts chl-a concentration in Dianchi Lake, China using Wavelet Analysis and LSTMs. [22] uses ANNs and SVMs to predict chl-a concentration in Juam and Yeongsan Reservoir, South Korea 7 days ahead.

The majority of the study sites relate to Far East Asia, Lake Erie or the Coast of Florida in the U.S [29, 8, 5]. The increased frequency of blooms results in more focus on these areas [13, 1]. The study of the locations of this work differ from the majority as well since most of the focus is divided between Southeast Asia and United States whereas our study area is the North and Irish Sea [23]. Most of the approaches use models like SVM or RF or using LSTMs to analyse the long/short term temporal patterns in the data. The approaches that classify the blooms use static values or expert information to classify the responses as in the cases of [21] and [30]. Our approach takes the context of the measurements into account as factors such as temperature since those factors affect cellular activity and oxygen solubility in water [19].

The proposed model predicts abnormal activity in monitored locations ranging from 1 day ahead to 7 days ahead, using only data from a single day, with a flexible labelling approach. Explanation models are used to provide insight into how the input influences the output of the model.

3 Dataset & Preprocessing

The data for this work was collected by ESM2 and ESMx data loggers at four different moorings depicted in Figure 1. The data was collected as a part of The National Marine Monitoring Programme

(NMMP) to monitor eutrophication regarding The Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR) and Marine Strategy Framework Directive (MSFD) assessments. The whole dataset was partitioned into four fractions based on location. Each of the datasets has different characteristics due to their locations such that the Liverpool buoy being near a maritime route, WestGab being near wind farms, TH1 being near the delta of the River Thames and Dowsing being in the open sea. It is known that the chl-a concentration has been decreasing in certain hotspots in the Southern North Sea [26].

The periodicity and the relationship between the variables were analysed by [4, 3, 14] with varying date ranges and locations by performing wavelet analysis. The periodicities of variables depend on the season and range between 6 hours to 24 hours. The data consists of eight features; chl fluorescence (*fluors*), turbidity(*ftu*), dissolved oxygen concentration(*o2conc*), salinity(*sal*), temperature(*temp*) and photosynthetically active radiation (PAR) at depths 0, 1 and 2 meters (*depth_0*, *depth_1*, *depth_2*). The majority of the data was collected at 20-30 minute intervals at each station. The data used spans the range between Jan 2009- Dec 2019. Before given as input, the data was normalized with z-score normalization.

Figure 1: Locations of moorings

Depending on environmental conditions the maximum amount of dissolved oxygen in a water body can differ. The labelling process used the following equation to calculate the maximum amount of dissolved oxygen concentration in the water given the temperature and salinity [12]:

$$\begin{aligned} D_O = In(A_0 + A_1T_+A_2T^2 + A_3T^2 + \\ A_3T^3 + A_4T^4 + A_5T^5 + \\ S(B_0 + B_1T + B_2T^2 + B_3T^3) + CS^2) \end{aligned} \quad (1)$$

where $A_0, ..., A_5$, $B_0, ..., B_3$ and C are coefficients of the equation given in Table 1, S is the salinity and T is $In[(298.15 - T_O)(273.15 + T_O)^{-1}]$ where T_O is the observed temperature value at time t. Algal bloom starts with the increased algal activity in a body of water which results in increased dissolved oxygen therefore thresholding was used, comparing the current dissolved oxygen to the maximum percentage of dissolved oxygen the water can hold at time t. If the percentage is above 105% the maximum threshold the label will be 1, else 0. The labelling process is done per day based on mean dissolved oxygen. The positive label percentages for each location is as follows: 1.44% for TH1, 3.89% for Dowsing, 3.98% for WestGab and 11.44% for LivBay.

Coefficient	Value
A_0	2.00907
A_1	3.22014
A_2	4.05010
A_3	4.944457
A_4	$-2.56847 * 10^{-1}$
A_5	3.887674
B_0	$-6.24523 * 10^{-3}$
B_1	$-7.37614 * 10^{-3}$
B_2	$-1.03410 * 10^{-2}$
B_3	$-8.17083 * 10^{-3}$
C	$-4.88682 * 10^{-7}$

Table 1: Coefficients for Equation 1

4 Methodology

The baseline models for this work were chosen as the SVM and RF as they were the most popular machine learning models for this task. We also include an isolation forest (IF) method to observe if the abnormalities could be identified in an unsupervised fashion by identifying the differences between normal occurrences and abnormalities. A convolutional variational autoencoder (VAE) is also included to see if relevant information could be extracted from a latent space regarding these abnormalities with varying filter sizes. Luong attention model is also included to observe if any improvements could be made over LSTM models.

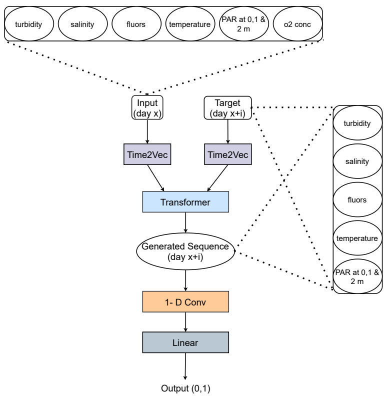

Figure 2: Proposed model for predicting oxygen thresholds. The input consists of all of the observed variables at day x, whereas the target consists of all variables except dissolved oxygen at day $x + i$. The transformer generates the target sequence for day $x + i$ except the dissolved oxygen. The output is a binary variable denoting if the average dissolved oxygen at day $x + i$ is below or above a threshold or not.

The proposed model (TF-Conv) consists of four components: a time embedding component (Time2Vec), a transformer, convolutional layer and linear layer with softmax [25, 17]. The embedding layer maps the input to two domains: time and frequency, the transformer is used to generate the sequence for x day(s) ahead, which is ranged between 1-7. Separate embedding components are used for input and target sequences as they differ in their number of features. The input is the measurements of day x and the target is the measurements of day $x + i$ where i is the number of days into the future ranging between 1 and 7. The input data is used to generate the target observations using the transformer network. The target variable is used during training to compute the loss between the generated sequence and the ground truth. Masking is used at the decoding stage of the transformer. During training teacher forcing is used for the transformer. The ground truth is given as the target value during decoding. During testing, the previous output of the transformer is used as the target tensor, initially a tensor of zeros of shape $(1, seq\_len, num\_features)$ is given as target. The convolutional layer is used for feature selection. The generated sequence does not include the dissolved oxygen so as not to overfit the convolution part of the model to only the dissolved oxygen. The generated sequence is taken through a 1-D convolution layer to serve as a feature selector. Lastly, the filtered observation is passed through a linear layer to classify the sequence. The labels were inversely weighted during training due to label imbalance as the ratio of labels is close to 96:4 in three of the four monitoring sites. The final output of the network is a binary variable which denotes if the daily average dissolved oxygen is above the threshold or not. Figure 2 illustrates the proposed architecture. The training and testing procedures are provided in pseudocode format in Algorithm 1 and 2.

GradientShap[1] was used as the explanation model. For the explanation model's baselines, we have used the training data of the prediction model. The output of the explanation model is per sample and per time-step. To give an overall view of the explanations we have decided to aggregate the explanations per day and compute the averages per feature.

5 Results

The predictions are done i days into the future given the observation at day x. i ranges between 1 to 7. 70% of data of TH1 buoy was used for training, 30% for validation. This location was chosen due to nutrient flow from the River Thames. A single location was used for training to test the generalisability of the model, to assess the model performance with data gathered from various locations with different properties. By modelling different nutrient concentrations, we aimed to make a more generalised model for this task. The other three sites are used for testing. Only one site is used for training to observe if the model is able to generalise different settings.

The F1 scores of each day for each site are presented in Figure 3. The mean F1 scores for all test locations are illustrated in Figure 4. F1 score was used as the performance metric due to the issue of label imbalance in the datasets. The weights

[1] https://captum.ai/api/gradient_shap.html

Algorithm 1 TF-Conv training (single batch)

Ensure: X_{src} = tensor of$(seq_len, batch_size, num_features)$
Ensure: X_{tgt} = tensor of$(seq_len, batch_size, num_features - 1)$
$X_{src} \leftarrow time2vec(X_{src})$
$X_{tgt} \leftarrow time2vec(X_{tgt})$
$X_{src} \leftarrow tf\_encode(X_{src})$
$X_{src} \leftarrow tf\_decode(X_{src}, X_{tgt}, masks)$
$X_{src} \leftarrow avg_pool(GeLU(conv_1d(X_{src})))$
$X_{src} \leftarrow softmax(linear(X_{src}))$

Algorithm 2 TF-Conv testing(single batch)

Ensure: X_{src} = tensor of$(seq_len, batch_size, num_features)$
Ensure: X_{tgt} = tensor of zeros$(seq_len, 1, num_features - 1)$
$X_{src}, X_{tgt} \leftarrow time2vec(X_{src}), time2vec(X_{tgt})$
$X_{src} \leftarrow tf\_encode(X_{src})$
$outputs = [\,]$
while $cur\_seq \neq seq\_len$ **do**
$output \leftarrow tf\_decode(X_{src}[cur\_seq], X_{tgt}, masks)$
$X_{tgt} \leftarrow output$
$outputs.append(output)$
end while
$outputs \leftarrow avg\_pool(GeLU(conv\_1d(outputs)))$
$outputs \leftarrow softmax(linear(outputs))$

of recall and precision were equal for the F1-score. An Adam optimizer was used for this task with 200 epochs and earlystopping with a patience of 15 epochs [18]. The embedding size of time2vec was set to 10 and the convolution window was set to 2 for all experiments. The rest of the hyperparameters are given in Table 2 based on prediction day. The hyperparameter optimization was done using grid search.

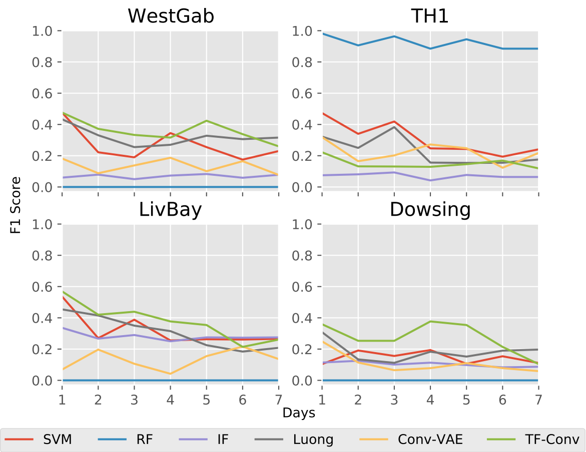

Figure 3: F1 scores for abnormality prediction for all 4 buoys

6 Discussion

In terms of mean f-score, the proposed model TF-Conv is the most suitable model for the majority of the cases. RF had problems such as overfitting as it performs nearly perfectly in the training site, TH1, whereas it performs poorly in other locations, SVM suffers from the same phenomenon for the Dowsing buoy. To obtain satisfactory results for RF, it could be trained on all four locations which might cause memory issues and maintenance costs. IF assumes that there are outliers in the data which can be predicted due to their different properties

Day	Batch Size	# of Encoder/Decoder Layers	# of Attention Heads	Transformer Network Dimensions	Learning Rate	Dropout Rate
1	16	5	2	32	0.003163	0.389
2	6	4	5	256	0.003837	0.314
3	6	3	5	32	0.001766	0.177
4	4	1	5	128	0.004316	0.284
5	32	2	2	32	0.000756	0.380
6	16	4	5	32	0.004591	0.213
7	8	4	5	32	0.003786	0.226

Table 2: Hyperparameters used for each model where the value of day is i days into the future.

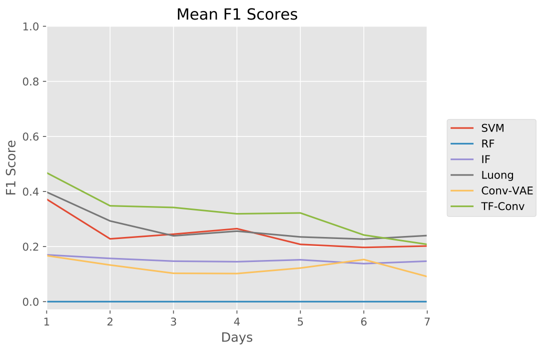

Figure 4: Mean F1 scores for abnormality prediction for testing buoys: WestGab, LivBay and Dowsing

and low occurrence rates. The results show that the increased activity in all of the sites were not outliers due to their properties and the assumptions made by IF does not hold.

The decreasing performance of the attention model between day 2 and 6 indicates that Luong attention is not suitable for predicting the near future blooms but it may be suitable for prediction for days further into the future. The inputs for the deep learning models are aggregated based on observation day whereas the machine learning models use averages of features based on observation day due to the model's limitation of not being able to model tensors more than 2 dimensions. The use of aggregation aids the deep learning models' generalisability since these models are exposed to raw data rather than a summarized version. Even with a summarized version of data, RF performs better in singular site comparison but the trade-off is made in generalisability. It can be noticed that as the percentage of positive samples increases, the performance of TF-Conv also increase as seen by the LivBay monitoring site.

The explanation model we used was *GradientShap* which works by adding random noise to data samples that were sampled between the baseline and the input and computing the gradients. The explanations differ from site to site as seen on Figure 5. It also shows that the order and the magnitude of the importances change from day to day. The model used assumes feature independence and the explanation model is linear. The explanation models show that each site has their own properties and the site with most positive labels (LivBay) and the best performance out of all sites has *o2conc* as the most important future which indicates that tracking the *o2conc* in the water might be useful where abnormalities frequently occur while using the TF-Conv model. The explanations also give insight into how input features differ from one another depending on prediction day, empirically showing the requirement of training a model for each prediction day.

7 Conclusion and Future Work

In this paper, we proposed a novel model for detecting algal blooms by predicting dissolved oxygen concentration 1 to 7 days ahead using time embeddings, transformer network and a convolutional layer. The proposed model increases the prediction performance by 5% in terms of F-score on average ranging from 1 to 7 days ahead of occurrence. The importance of each feature is provided with SHAP values per day increasing interpretability of the model. We have observed that the most important feature changes based on monitoring site and prediction day.

Data with different frequencies such as ship-based data or data with different modalities could be used to improve the detection process. This work could be extended to closed bodies of water. The current results indicate that models could be tested for different day ranges than they were trained on

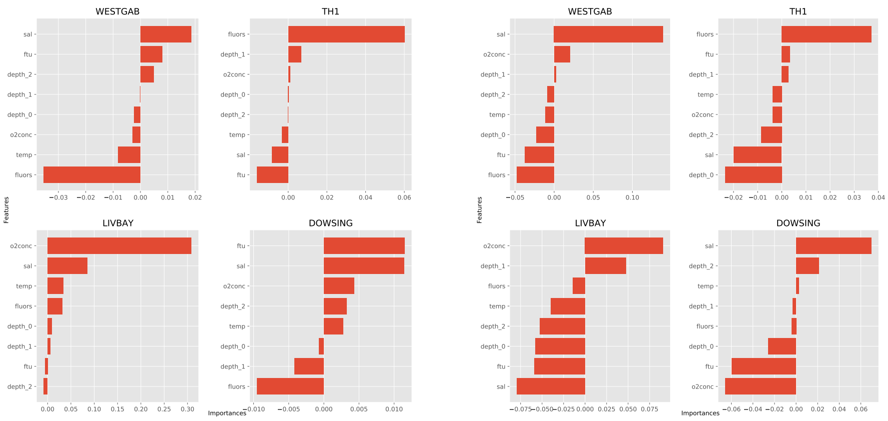

Figure 5: Left: Feature importances of SHAP for predictions 1-day ahead. Right: Feature importances of SHAP for predictions 7-days ahead.

to test the model's generalisability. The stability of the model could be checked by predicting bloom events further than seven days. The model's performance could be assessed by training it per location. Generalisability among different locations was not included in the scope of this work, transfer learning methods could be used in the future to test the efficiency of this architecture.

Acknowledgements

The authors would like to thank CEFAS for the expertise in platform and sensor development and supplying the data for this work and the University of Hull VIPER team for supplying the infrastructure for experimentation. The data was collected under the National Programme of Eutrophication Monitoring (Sla25), supported by Department for Environment, Food and Rural Affairs.

References

[1] D. M. Anderson, E. Fensin, C. J. Gobler, A. E. Hoeglund, K. A. Hubbard, D. M. Kulis, J. H. Landsberg, K. A. Lefebvre, P. Provoost, M. L. Richlen, et al. Marine harmful algal blooms (habs) in the united states: history, current status and future trends. *Harmful Algae*, 102:101975, 2021. doi:10.1016/j.hal.2021.101975.

[2] A. Bechard. The economic impacts of harmful algal blooms on tourism: an examination of southwest florida using a spline regression approach. *Natural Hazards*, 104(1):593–609, 2020. doi:10.1007/s11069-020-04182-7.

[3] A. N. Blauw, E. Beninca, R. W. Laane, N. Greenwood, and J. Huisman. Dancing with the tides: fluctuations of coastal phytoplankton orchestrated by different oscillatory modes of the tidal cycle. *PLoS One*, 7(11), 2012. doi:10.1371/journal.pone.0049319.

[4] A. N. Blauw, E. Benincà, R. W. Laane, N. Greenwood, and J. Huisman. Predictability and environmental drivers of chlorophyll fluctuations vary across different time scales and regions of the north sea. *Progress in Oceanography*, 161:1–18, 2018. doi:10.1016/j.pocean.2018.01.005.

[5] J. P. Cannizzaro, C. Hu, D. C. English, K. L. Carder, C. A. Heil, and F. E. Müller-Karger. Detection of karenia brevis blooms on the west florida shelf using in situ backscattering and fluorescence data. *Harmful Algae*, 8(6):898–909, 2009. doi:10.1016/j.hal.2009.05.001.

[6] M. F. Chislock, E. Doster, R. A. Zitomer, and A. E. Wilson. Eutrophication: causes, consequences, and controls in aquatic ecosystems. *Nature Education Knowledge*, 4(4):10, 2013. `https://www.nature.com/scitable/knowledge/library/eutrophication-causes-consequences-and-controls-in-aquatic-102364466/`.

[7] H. Cho, U. Choi, and H. Park. Deep learning application to time-series prediction of daily chlorophyll-a concentration. *WIT Trans. Ecol. Environ.*, 215:157–63, 2018. doi:10.2495/EID180141.

[8] H. Cho and H. Park. Merged-lstm and multistep prediction of daily chlorophyll-a concentration for algal bloom forecast. In *IOP Conference Series: Earth and Environmental Science*, volume 351, page 012020. IOP Publishing, 2019. doi: 10.1088/1755-1315/351/1/012020.

[9] S. Cho, B. Lim, J. Jung, S. Kim, H. Chae, J. Park, S. Park, and J. K. Park. Factors affecting algal blooms in a man-made lake and prediction using an artificial neural network. *Measurement*, 53:224–233, 2014. doi:10.1016/j.measurement.2014.03.044.

[10] W. K. Dodds, W. W. Bouska, J. L. Eitzmann, T. J. Pilger, K. L. Pitts, A. J. Riley, J. T. Schloesser, and D. J. Thornbrugh. Eutrophication of us freshwaters: analysis of potential economic damages, 2009. doi:10.1021/es801217q.

[11] I. R. Falconer, M. D. Burch, D. A. Steffensen, M. Choice, and O. R. Coverdale. Toxicity of the blue-green alga (cyanobacterium) microcystis aeruginosa in drinking water to growing pigs, as an animal model for human injury and risk assessment. *Environmental toxicology and Water quality*, 9(2):131–139, 1994. doi:10.1002/tox.2530090209.

[12] H. E. Garcia and L. I. Gordon. Oxygen solubility in seawater: Better fitting equations. *Limnology and oceanography*, 37(6):1307–1312, 1992. doi:10.4319/lo.1992.37.6.1307.

[13] H. Gu, Y. Wu, S. Lü, D. Lu, Y. Z. Tang, and Y. Qi. Emerging harmful algal bloom species over the last four decades in china. *Harmful Algae*, page 102059, 2021. doi:10.1016/j.hal.2021.102059.

[14] J. Heffernan, J. Barry, M. Devlin, and R. Fryer. A simulation tool for designing nutrient monitoring programmes for eutrophication assessments. *Environmetrics: The official journal of the International Environmetrics Society*, 21(1):3–20, 2010. doi:10.1002/env.980.

[15] M. Kahru and B. G. Mitchell. Ocean color reveals increased blooms in various parts of the world. *Eos, Transactions American Geophysical Union*, 89(18):170–170, 2008. doi:10.1029/2008EO180002.

[16] B. Karlson, P. Andersen, L. Arneborg, A. Cembella, W. Eikrem, U. John, J. J. West, K. Klemm, J. Kobos, S. Lehtinen, et al. Harmful algal blooms and their effects in coastal seas of northern europe. *Harmful Algae*, page 101989, 2021. doi:10.1016/j.hal.2021.101989.

[17] S. M. Kazemi, R. Goel, S. Eghbali, J. Ramanan, J. Sahota, S. Thakur, S. Wu, C. Smyth, P. Poupart, and M. Brubaker. Time2vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*, 2019. `https://arxiv.org/abs/1907.05321`.

[18] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. `https://arxiv.org/abs/1412.6980`.

[19] J. R. Lepock. How do cells respond to their thermal environment? *International journal of hyperthermia*, 21(8):681–687, 2005. doi:10.1080/02656730500307298.

[20] X. Li, J. Yu, Z. Jia, and J. Song. Harmful algal blooms prediction with machine learning models in tolo harbour. In *2014 International Conference on Smart Computing*, pages 245–250. IEEE, 2014. doi:10.1109/SMARTCOMP.2014.7043865.

[21] N. Mellios, S. J. Moe, and C. Laspidou. Machine learning approaches for predicting health risk of cyanobacterial blooms in northern european lakes. *Water*, 12(4):1191, 2020. doi:10.3390/w12041191.

[22] Y. Park, K. H. Cho, J. Park, S. M. Cha, and J. H. Kim. Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, korea. *Science of the Total Environment*, 502:31–41, 2015. doi:10.1016/j.scitotenv.2014.09.005.

[23] M.-T. Sebastiá-Frasquet, J.-A. Aguilar-Maldonado, I. Herrero-Durá, E. Santamaría-del Ángel, S. Morell-Monzó, and J. Estornell. Advances in the monitoring of algal blooms by remote sensing: A bibliometric analysis. *Applied Sciences*, 10(21):7877, 2020. doi:10.3390/app10217877.

[24] S. Shamshirband, E. Jafari Nodoushan, J. E. Adolf, A. Abdul Manaf, A. Mosavi, and K.-w. Chau. Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. *Engineering Applications of Computational Fluid Mechanics*, 13(1):91–101, 2019. doi:10.1080/19942060.2018.1553742.

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017. `https://arxiv.org/abs/1706.03762`.

[26] K. Von Schuckmann, P.-Y. Le Traon, N. Smith, A. Pascual, S. Djavidnia, J.-P. Gattuso, M. Grégoire, G. Nolan, S. Aaboe, E. Á. Fanjul, et al. Copernicus marine service ocean state report, issue 4. *Journal of Operational Oceanography*, 13(sup1):S1–S172, 2020. doi:10.1080/1755876X.2020.1785097.

[27] X. Wang and L. Xu. Unsteady multi-element time series analysis and prediction based on spatial-temporal attention and error forecast fusion. *Future Internet*, 12(2):34, 2020. doi:10.3390/fi12020034.

[28] M. L. Wells, V. L. Trainer, T. J. Smayda, B. S. Karlson, C. G. Trick, R. M. Kudela, A. Ishikawa, S. Bernard, A. Wulff, D. M. Anderson, et al. Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful algae*, 49:68–93, 2015. doi:10.1016/j.hal.2015.07.009.

[29] H. Yajima and J. Derot. Application of the random forest model for chlorophyll-a forecasts in fresh and brackish water bodies in japan, using multivariate long-term databases. *Journal of Hydroinformatics*, 20(1):206–220, 2018. doi:10.2166/hydro.2017.010.

[30] Y. Yang, Y. Bai, X. Wang, L. Wang, X. Jin, and Q. Sun. Group decision-making support for sustainable governance of algal bloom in urban lakes. *Sustainability*, 12(4):1494, 2020. doi:10.3390/su12041494.

[31] H.-S. Yi, S. Park, K.-G. An, and K.-C. Kwak. Algal bloom prediction using extreme learning machine models at artificial weirs in the nakdong river, korea. *International journal of environmental research and public health*, 15(10):2078, 2018. doi:10.3390/ijerph15102078.

[32] Z. Yu, K. Yang, Y. Luo, and C. Shang. Spatial-temporal process simulation and prediction of chlorophyll-a concentration in dianchi lake based on wavelet analysis and long-short term memory network. *Journal of Hydrology*, 582:124488, 2020. doi:10.1016/j.jhydrol.2019.124488.