Algal bloom spectrum- Canada

<https://www.canada.ca/en/environment-climate-change/services/water-overview/satellite-earth-observations-lake-monitoring/remote-sensing-algal-blooms.html>

IDEAS: Check correlation between parameters and see if dimensions can be reduced for highly correlated data (How does this affect accuracy?)

Check clusters with K-means

Check principal components with PCA

Use a simple model first to see bias-variance and then improve model based on the results

Formatting:

**TITLE (CITATION)**

**REFERENCE (APA)**

**GIST**

**ALGORITHM**

**PARAMETERS**

1. **A Method for Chlorophyll-a and Suspended Solids Prediction through Remote Sensing and Machine Learning  2020 (3) (Moderately Relevant)**

Silveira Kupssinskü, L., Thomassim  Guimarães, T., Menezes de Souza, E., C Zanotta, D., Roberto Veronez, M.,  Gonzaga, L., & Mauad, F. F. (2020). A method for chlorophyll-a and suspended solids prediction through remote sensing and machine learning.  *Sensors*, *20*(7), 2125.

TSS and Chlorophyll-a mesured using Sentinel-2 Spectral images and UAVs and concentration predicted using supervised ML

Monitoring TSS and Chl-a is essential for sustainability and better management of water resources.

RF and ANN had the highest R^2 values

1. **A New Predictive Model for Evaluating Chlorophyll-a Concentration in Tanes Reservoir by Using a Gaussian Process Regression 2020 (1) (highly relevant)**

García-Nieto, P. J., García-Gonzalo, E., Fernández, J. R. A., & Muñiz, C. D. (2020). A New Predictive Model for Evaluating Chlorophyll-a Concentration in Tanes Reservoir by Using a Gaussian Process Regression. *Water Resources Management*, *34*(15), 4921-4941.

Chl-a concentration predicted using Gaussian process regression (GPR) with LBFGSB optimizer for reservoir using 268 samples over 10 years. R^2 values 0.8597

1. **Application of feature selection and regression models for chlorophyll-a prediction in a shallow lake 2018 (15) (HIGHLY RELEVANT)**

Li, X., Sha, J., & Wang, Z. L. (2018). Application of feature selection and regression models for chlorophyll-a prediction in a shallow lake. *Environmental Science and Pollution Research*, *25*(20), 19488-19498.

Chl-a concentration predicted using lakewater using RF and SVR. The paper also identified relevant inputs: feature selection lead to improved model accuracy. RF had better prediction accuracy than SVR. This approach is useful for lakes with small datasets.

1. **Application of the Random Forest model for chlorophyll-a forecasts in fresh and brackish water bodies in Japan, using multivariate long-term databases 2018 (28) (HIGHLY RELEVANT)**

Yajima, H., & Derot, J. (2018). 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.

Used RF to predict the chl-a concentration in fresh water and saltwater lakes. Limited dataset had significant impact on the prediction performance. Most important prdictors didn't necessarily have a strong statistical correlation with the target parameter (BOD, COD, pH, TN/TP were identified as most influential)

1. **Evaluating complex relationships between ecological indicators and environmental factors in the Baltic Sea: A machine learning approach 2019 (13) (Slightly relevant)**

Lehikoinen, A., Olsson, J., Bergström, L., Bergström, U., Bryhn, A., Fredriksson, R., & Uusitalo, L. (2019). Evaluating complex relationships between ecological indicators and environmental factors in the Baltic Sea: A machine learning approach. *Ecological Indicators*, *101*, 117-125.

Tree augmented naive bayes classifiers were used with Entropy Minimization Discretization algorithm to assess multiple environmental factors and ecological indicators

1. **LSTM Networks to Improve the Prediction of Harmful Algal Blooms in the West Coast of Sabah 2021 (0) (MODERATELY RELEVANT)**

Yussof, F. N., Maan, N., & Md Reba, M. N. (2021). LSTM Networks to Improve the Prediction of Harmful Algal Blooms in the West Coast of Sabah. *International Journal of Environmental Research and Public Health*, *18*(14), 7650.

LSTM and CNN were used to predict harmful algal bloom in Sabah, West Malaysia using Satellite data and chl-a concentration data (15 years from bathymetry data, spectral . Mostly satellite). LSTM outperformed CNN because it can learn long term dependencies but correlation coefficient is still low

1. **Machine Learning Automatic Model Selection Algorithm for Oceanic Chlorophyll-a Content Retrieval 2018 (22) (MODERTELY RELEVANT)**

Blix, K., & Eltoft, T. (2018). Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval. *Remote Sensing*, *10*(5), 775.

Water quality retrieval using regression algorithms from multispectral data (Ocean colour monitoring). This paper proposes Automatic Model Selection Algorithm (AMSA) that selects best model based on the dataset. The AMSA used in this paper is used to estimate oceanic chl-a. ML algo evaluated are Gaussian Process Regression (GPR), Support Vector Regression (SVR) and Partial Least Square Regression (PLSR) models.

Helpful tool for water quality analysis using remote sensing data. Improved understanding of the underlying physical processes due to feature ranking methods. Paper showed combining ML feature ranking and regression methods in AMSA can reduce computational time and result in improved regression . GPR and SVR were confirmed to show strong regression power.

1. **Machine Learning-Based Ensemble Prediction of Water-Quality Variables Using Feature-Level and Decision-Level Fusion with Proximal Remote Sensing - 2019 (16) MODERATELY RELEVANT**

Peterson, K. T., Sagan, V., Sidike, P., Hasenmueller, E. A., Sloan, J. J., & Knouft, J. H. (2019). Machine learning-based ensemble prediction of water-quality variables using feature-level and decision-level fusion with proximal remote sensing. *Photogrammetric Engineering & Remote Sensing*, *85*(4), 269-280.

Several key water quality parameters including chl-a were predicted using multiple linear regression, partial least-squares regression, Gaussian process regression, support vector machine regression, and extreme learning machine regression with Spectral Reflectance data. Canonical correlation analysis feature-level fusion was developed for spectral analysis of water quality variables. VArious ML models were combined for ensemble forecasting method  using decision level fusion approach and it proved to be the most effective

1. **Merged-LSTM and multistep prediction of daily chlorophyll-a concentration for algal bloom forecast 2019 (4) (HIGHLY RELEVANT)**

Cho, H., & Park, H. (2019, October). Merged-LSTM and multistep prediction of daily chlorophyll-a concentration for algal bloom forecast. In *IOP Conference Series: Earth and Environmental Science* (Vol. 351, No. 1, p. 012020). IOP Publishing.

3 LSTMs merged which uses data from diverse sources to predict (data over 7 days) or  forecast algal blooms. Parameters used to predict are TN, TP, TOC, Chl-a, TN, TP, TOC, Temperature, Solar radiation rainfall, flow rate

1. **Modelling Chlorophyll-a Concentration using Deep Neural Networks considering Extreme Data Imbalance and Skewness 2019 (8) (HIGHLY RELEVANT)**

Choi, J. H., Kim, J., Won, J., &  Min, O. (2019, February). Modelling chlorophyll-a concentration using  deep neural networks considering extreme data imbalance and skewness. In  *2019 21st International Conference on Advanced Communication Technology (ICACT)* (pp. 631-634). IEEE.

Chl-a predicted (concentration of chl-a in 7 days) using water temperature, pH, electrical conductivity, dissolved oxygen, total organic carbons, total nitrogen, total phosphorus and chlorophyll-a while handling data skewness and imbalance using convolutional neural network. Log transformation and oversampling techniques help improve the performance of the predictive model.

1. **Monthly chlorophyll-a prediction using neuro-genetic algorithm for water quality management in Lakes 2016 (9) (HIGHLY RELEVANT)**

Lee, G., Bae, J., Lee, S., Jang, M.,  & Park, H. (2016). Monthly chlorophyll-a prediction using  neuro-genetic algorithm for water quality management in Lakes. *Desalination and Water Treatment*, *57*(55), 26783-26791.

Chl-a concentration in Lakes used for drinking water was predicted (1 month forward [future] prediction) using temperature (temp), dissolved oxygen (DO),

pH, biological oxygen demand (BOD), chemical oxygen demand (COD), suspended solid (SS), ammoniacal nitrogen (NH3-N), nitrate (NO3-N), dissolved total nitrogen (DTN), total nitrogen (TN), phosphate phosphorus (PO3-P), dissolved total phosphorus (DTP), total phosphorus (TP), electrical conductivity(EC), and Chl-a.

Neuro genetic algorithm NGA (GA combined with ANN ) algo with 0.9 R^2 where double hidden layer showed better performance. GA used to determine effective number of nodes and activation functions.

1. **Multistep-ahead forecasting of chlorophyll a using a wavelet nonlinear autoregressive network 2018 (16) (HIGHLY RELEVANT)**

Du, Z., Qin, M., Zhang, F., & Liu,  R. (2018). Multistep-ahead forecasting of chlorophyll a using a wavelet  nonlinear autoregressive network. *Knowledge-Based Systems*, *160*, 61-70.

Multistep-ahead forecasting model wavelet nonlinear autoregressive network (WNARNet) that integrates the wavelet transform and a nonlinear autoregressive neural network (NAR) is proposed to forecast Chl-a concentration. Model performs well in predicting the dynamics of chl-a and can forecast 20 steps ahead. Wavelet transform decreases accumulative errors and NAR decreases dependencies between time series in COASTAL WATERS. It predicts using time series Chl-a data but remote buoy continuously measure  chl-a, temperature, DO, salinity, etc.

Chl-a is predicted multiple steps ahead with high accuracywhere r is 0.08 higher and the RMSE is  0.04 lower than the values of the benchmark models

1. **Prediction of Chlorophyll-a Concentrations in the Nakdong River Using Machine Learning Methods 2020 (15) (HIGHLY RELEVANT)**

Shin, Y., Kim, T., Hong, S., Lee, S.,  Lee, E., Hong, S., ... & Heo, T. Y. (2020). Prediction of  chlorophyll-a concentrations in the Nakdong River using machine learning  methods. *Water*, *12*(6), 1822.

weather variables (AvgTemp, Sunshine, Rainfall, Inflow, and Outflow) and water quality variables (WaterTemp, pH,EC, DO, and TOC) and chl-awere used as the explanatory variables to predict chl-a concentration nakdong River

Out of several ML models tested (as Support Vector Regression, Bagging, Random Forest, Extreme Gradient Boosting (XGBoost), Recurrent Neural Network (RNN), and Long–Short-Term Memory(LSTM)), RNN model combined with rolling window learning method outperformed Variable selection using the forward selection method and 1-step ahead recursive learning can increase the model prediction accuracy.

1. **Reconstructing Global Chlorophyll-a Variations Using a Non-linear Statistical Approach 2020 (5) (MODERATELY RELEVANT)**

Martinez, E., Gorgues, T., Lengaigne,  M., Fontana, C., Sauzède, R., Menkes, C., ... & Fablet, R. (2020).  Reconstructing global chlorophyll-a variations using a non-linear  statistical approach. *Frontiers in Marine Science*, *7*, 464.

Oceanic and atmospheric variables used to reconstruct surface chl-a (or just Chl) spatio-temporal variations were reconstructed using SVR with 13 year training period to simulate chl variability 32 years global physical-biogeochemical simulation. SVR reconstruct satellite chl observation accurately reproduce some aspectsof chl variability.

Interannual variations are reproduced for tropical pacific and indian oceans an it also accurately captures chl trends estimated by satellite data in exratropical and subtropical gyres

1. **Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach 2020 (59) (MODERATELY RELEVANT)**

Pahlevan, N., Smith, B., Schalles, J.,  Binding, C., Cao, Z., Ma, R., ... & Stumpf, R. (2020). Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI)  in inland and coastal waters: A machine-learning approach. *Remote Sensing of Environment*, *240*, 111604.

Inland and coastal water chl-a concentration retrieved using hyperspectral in-situ radiometric data and ocean and satellite imaging using Mixture Density Network (MDN)

Retrieved chla from reflectance data from sentinel-2&3 and trained using coincident chl-a and reflectance dataset

1. **Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model 2020 (32) (HIGHLY RELEVANT)**

Barzegar, R., Aalami, M. T., &  Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, 1-19.

chla and DO was predicted using EC, pH, ORP & water temperature using DL models (CNN & LSTM) both standalone and

combined. Hybrid CNN-LSTM captured both low and high levels of water quality and performed well for both DO and Chla, especially DO

1. **Transfer learning for neural network model in chlorophyll-a dynamics prediction 2019 (11) (HIGHLY RELEVANT)**

Tian, W., Liao, Z., & Wang, X. (2019). Transfer learning for neural network model in chlorophyll-a dynamics prediction. *Environmental Science and Pollution Research*, *26*(29), 29857-29871.

Water temperature, pH, electronic conductivity, ORP, turbidity, DO, and current chlorophyll-a were the inputs, and the chlorophyll-a of the next point in time was the output.

feedforward neural networks (FNN) model with transfer learning is the most suitable method for chl-a prediction. TL was tested on FNN, RNN & LSTM

1. **WaterNet: A Convolutional Neural Network for Chlorophyll-a Concentration Retrieval 2020 (9) (MODERATELY RELEVANT)**

Syariz, M. A., Lin, C. H., Nguyen, M.  V., Jaelani, L. M., & Blanco, A. C. (2020). WaterNet: A  convolutional neural network for chlorophyll-a concentration retrieval. *Remote Sensing*, *12*(12), 1966.

**Chl-a concentration estimation retrieval from sentinel images coupled with insitu measurements using CNN based model . Two step training was implemented nd WaterNet model could accurately capture nonlinearities**