Machine Learning Application in Water Quality Using Satellite Data

N Hassan¹,²,* and C S Woo¹

¹Faculty of Computer Science and Information Technology, Universiti Malaya, 50603 Lembah Pantai, Kuala Lumpur, Malaysia.
²Institute of Oceanography and Environment, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu.
*Corresponding author: zuhrah82@gmail.com

Abstract. Monitoring water quality is a critical aspect of environmental sustainability. Poor water quality has an impact not just on aquatic life but also on the ecosystem. The purpose of this systematic review is to identify peer-reviewed literature on the effectiveness of applying machine learning (ML) methodologies to estimate water quality parameters with satellite data. The data was gathered using the Scopus, Web of Science, and IEEE citation databases. Related articles were extracted, selected, and evaluated using advanced keyword search and the PRISMA approach. The bibliographic information from publications written in journals during the previous two decades were collected. Publications that applied ML to water quality parameter retrieval with a focus on the application of satellite data were identified for further systematic review. A search query of 1796 papers identified 113 eligible studies. Popular ML models application were artificial neural network (ANN), random forest (RF), support vector machines (SVM), regression, cubist, genetic programming (GP) and decision tree (DT). Common water quality parameters extracted were chlorophyll-a (Chl-a), temperature, salinity, colored dissolved organic matter (CDOM), suspended solids and turbidity. According to the systematic analysis, ML can be successfully extended to water quality monitoring, allowing researchers to forecast and learn from natural processes in the environment, as well as assess human impacts on an ecosystem. These efforts will also help with restoration programs to ensure that environmental policy guidelines are followed.

1. Introduction

1.1. Water quality

Water quality describes a state of a water body, as well as its chemical, physical, and biological aspects, including its usefulness for a particular activity (i.e., fishing, swimming or drinking). Substances that can damage aquatic species if found in high enough quantities can also impair water quality. Monitoring water quality is a critical aspect of environmental sustainability. Poor water quality has an impact not just on aquatic life but also on the ecosystem. The following variables are also be used to provide an indicator of water quality: the content of dissolved oxygen (DO); amounts of fecal coliform bacteria from people and animal wastes; levels or ratio of plant nutrients nitrogen and phosphorus; volume of particulate suspended matter (turbidity) and the amount of salt (salinity) in the water. To assess water quality, quantities of substances such as pesticides, herbicides, heavy metals, and other pollutants can be calculated. The abundance of chlorophyll-a (Chl-a), a green pigment...
present in microscopic algae, is often filtered from water samples in many water bodies to provide an indicator of the microalgae living in the water column [1].

1.2. Satellite and remote sensing
Remote sensing is the method of surveying the surface of the earth without making any physical connection. It is used primarily to collect data from the earth's properties and analyze changes in the earth's environment. Along with improvements in satellite technologies and device processing capability, remote sensing has become more widely used in this era. Remote sensing generates spectral, infrared, and radar images that can be interpreted and analyzed to extract useful knowledge about earth elements like water, soil, plants, and the atmosphere, among others. These data are often used to forecast weather and environment, as well as for tracking animal populations, crop health, shoreline changes, and land-use change detection. The resolution of remote sensing data varies depending on the satellite capability. Remote sensing data has recently been produced and effectively utilized to collect water quality information as a solution to the limitations of traditional methods [2]. Remotely sensed data sets are usually more extensive than those collected directly on site by providing better resolution and typically higher temporal frequency and resolution for spatial coverage [3]. Remote satellite sensing examples include Landsat, Sentinel, MODIS, MERIS and VIIRS.

1.3. Machine learning
Machine Learning (ML) is a type of statistical approach that can automatically learn from data and construct a detection, estimation, or classification model that minimizes the variance between the training and prediction datasets without being actively programmed. ML, also known as statistical learning, is providing data to a computer that can be "trained" using known or predetermined attributes or objects to allow semi-automatic or automatic detection, classification, or pattern recognition. ML enabling remotely sensed water quality estimate has grown in popularity in recent years as a result of improvements in algorithm development, computer power, sensor systems, and availability of data [4].

1.4. Systematic review objectives
In this systematic review, the effectiveness of applying ML methodologies were investigated to retrieve water quality parameters from satellite data. Specifically, the objective of studies, the types of satellite data, the ML methodologies, the significance or outcome of the ML application were summarized.

1.5. Nomenclature
Figure 1 provided the list of the abbreviations, acronyms and symbols used in this manuscript.

2. Materials and Methods
The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology was used to prepare and report the results of this study [5]. PRISMA is a standard method to give a systematic review of existing research.

2.1. Eligibility criteria
This study focused on peer-reviewed publications that applied ML to estimate water quality parameters with satellite data. The searches for and screen publications focused on three criteria: (1) water quality parameters, (2) ML techniques, and (3) type of satellite.
| Abbreviation | Description |
|--------------|-------------|
| ADtrees      | Alternating Decision Trees |
| ANFIS        | Adaptive Network-Based Fuzzy Inference System |
| ANN          | Artificial Neural Network |
| AVHRR        | Advanced Very High-Resolution Radiometer |
| BART         | Bayesian Additive Regression Tree |
| BDtrees      | Binary Decision Trees |
| BGA          | Blue-Green Algae |
| BIOC         | Bioclim |
| BR           | Band Ratio |
| BRT          | Boosted Regression Tree |
| C2RCC        | Case-2 Regional/Coast Colour |
| CART         | Classification And Regression Trees |
| CFNN         | Cascade Forward Neural Network |
| Chl-a        | Chlorophyll-a |
| Chl-b        | Chlorophyll-b |
| Chl-c        | Chlorophyll-c |
| CI-OC3M      | Global Ocean Algorithm |
| CNN          | Convolutional Neural Network |
| COD          | Chemical Oxygen Demand |
| COMS         | Communication, Ocean and Meteorological Satellite |
| CZCS         | Coastal Zone Color Scanner Experiment |
| DIN          | Dissolved Inorganic Nitrogen |
| DIP          | Dissolved Inorganic Phosphorus |
| DNN          | Deep Neural Network |
| DO           | Dissolved Oxygen |
| DOC          | Dissolved Organic Carbon |
| OM           | D Domain |
| DSCOVIR      | Deep Space Climate Observatory |
| DT           | Decision Tree |
| E. coli      | Escherichia Coli |
| ELM          | Extreme Learning Machine |
| EnMAP        | Environmental Mapping and Analysis Program |
| EPIC         | Earth Polychromatic Imaging Camera |
| ESA          | European Space Agency |
| ET           | Extremely Randomized Tree |
| ETM+         | Enhanced Thematic Mapper Plus |
| FDA          | Flexible Discriminant Analysis |
| fDOM         | Fluorescent Dissolved Organic Matter |
| FFNN         | Feed-Forward Neural Network |
| GA           | Genetic Algorithm |
| GAM          | Generalized Additive Models |
| GAR          | Generalized Additive Regression |
| GBDT         | Gradient Boosting Decision Tree |
| GBM          | Gradient Boosting Machine |
| GCOM-C       | Global Change Observation Mission – Climate |
| GEE          | Google Earth Engine |
| GLM          | Generalized Linear Model |
| GMM          | Gaussian Mixture Model |
| GNWNR        | Geographically Neural Network Weighted Regression |
| GOCI         | Geostationary Ocean Color Imager |
| GP           | Genetic Programming |
| GPR          | Gaussian Process Regression |
| GWR          | Geographically Weighted Regression |
| HAC          | Hierarchical Ascendant Classification |
| HICO         | Hyperspectral Imager for the Coastal Ocean |
| HMM          | Hidden Markov Model |
| Hyperion     | Hyperspectral Imager |
| IDW          | Inverse Distance Weighting |
| ISS          | International Space Station |
| kNN          | k-nearest neighbour |
| KRR          | Kernel Ridge Regression |
| LightGBM     | Light Gradient Boosting Machine |
| LR           | Linear Regression |
| LSE          | Least Square |
| LSTM         | Long Short-Term Memory |
| MAHA         | Mahalanobis Distance |
| MARS         | Multivariate Adaptive Regression Splines |
| MAXENT       | Maximum Entropy |
| MAXL         | Maxlike |
| MDA          | Mixture Discriminant Analysis |
| MERIS        | Medium Resolution Imaging Spectrometer |
| MERSI        | Medium Resolution Spectral Imager |
| ML           | Machine Learning |
| MLNN         | Multilayer Neural Network |
| MLP          | Multilayer Perceptron Neural Network |
| OC3M         | Ocean Color 3m |
| O4Me         | Ocean Colour for Meris |
| OCI          | Ocean Color Instrument |
| OCN          | Ocean Color Net |
| OLCI         | Ocean And Land Colour Instrument |
| OLI          | Operational Land Imager |
| OLR          | Ordinary Linear Regression |
| PCA          | Principal Component Analysis |
| PCR          | Principal Component Regression |
| PLSR         | Partial Least Squares Regression |
| POC          | Particulate Organic Carbon |
| PPC          | Photoprotective Carotenoids |
| PSC          | Photosynthetic Carotenoids |
| RF           | Random Forest |
| RLR          | Ridge Linear Regressor |
| RPART        | Recursive Partitioning and Regression Trees |
| SC           | Specific Conductance |
| SDD          | Secchi Disk Depth |
| SeaWiFS      | Sea-Viewing Wide Field-Of-View Sensor |
| SEIK         | Singular Evolutive Interpolated Kalman |
| SGLI         | Second-Generation Global Imager |
| SLSTR        | Sea And Land Surface Temperature Radiometer |
| SMOS         | Soil Moisture and Ocean Salinity |
| SNPP         | Suomi National Polar-Orbiting Partnership |
| SOM          | Self-Organizing Map |
| SpecWa       | Spectral Remote Sensing Data and Chlorophyll-a |
| SS            | Value of Inland Waters. |
| SST           | Sea Surface Salinity |
| STA         | Subsurface Temperature Anomaly |
| SVM          | Support Vector Machine |
| TM           | Thematic Mapper |
| TN            | Total Nitrogen |
| TP            | Total Phosphorus |
| TSM          | Total Suspended Matters |
| VIIRS        | Visible Infrared Imaging Radiometer Suite |
| XGBoost      | Extreme Gradient Boosting |

**Figure 1.** List of the abbreviations, acronyms and symbols.
2.2. **Information sources and search**

The peer-reviewed publications were searched in three resources: Scopus, Web of Science and IEEE citation databases. The search was restricted to research articles published in English and in peer-reviewed journals or conference proceedings. The following query is constructed with the Boolean operator `AND` and `OR`. The list of queries is shown in Table 1. The searches were run against the title, keywords and abstract of the publications in different databases separately.

**Table 1.** Keywords used for article search

| Keyword set no. | Search Strategy                                                                 |
|----------------|---------------------------------------------------------------------------------|
| 1              | “machine learning” `AND` “satellite” `OR` “ocean colour” `OR` “organic” `OR`     |
|                | “phytoplankton” `OR` “salinity” `OR` “temperature” `OR` “time series” `OR`     |
|                | “water quality” `OR` “suspended” `OR” “CDOM”                                    |
| 2              | “ocean colour” `AND” “ocean color” `OR” “forecast” `OR” “forecasting” `OR”   |
|                | “predict” `OR” “prediction”                                                     |

2.3. **Study selection**

The eligibility of publications was evaluated and the publications were screened by examining the titles, abstracts and methods, and then obtained eligible publications through reading the full text.

2.4. **Data collection and analysis**

The data were documented with the objectives, methodologies, environments, problems investigated, language and datasets for each eligible publication. A narrative synthesis of all relevant papers was carried out and arranged based on (1) research goal, (2) ML methodologies, and (3) scientific findings. While the first perspective demonstrated satellite data applications for water quality monitoring, the second view gave an insight into current techniques of study and challenges when applying ML to process and analyze water quality parameters. The third viewpoint showed the lessons that may be drawn from water quality concerns.

2.5. **Risk of bias**

This systematic review is biased in certain aspects. To begin with, there is a risk of bias in the review process because there is only one reviewer who screens the literature, and the subjectivity of the inclusion and exclusion criteria may influence the selection of relevant articles. Furthermore, throughout the search procedure, the year range was not specified. This implies that the search results are from all years accessible, starting with the earliest publication discovered in the individual databases and ending with the most current (May 2021). Moreover, though the search was limited to three databases, there are many more databases (e.g., Google Scholar, ACM Digital Library) that may contain more material addressing the ML application in water quality utilizing satellite data discussed in this paper.

3. **Results**

The process of identifying eligible articles is depicted in Figure 2. Initially, the queries returned 1796 publications. After that, the publications were screened to eliminate duplicates. There are 473 duplicates that were removed. The abstracts and titles were read in order to examine the techniques and account for the aforementioned inclusion and exclusion criteria, resulting in the removal of 1196 articles and the retention of 127 for a more in-depth examination. Following the full publication review, 14 studies were excluded due to non-English language publications and studies that were unable to get access to the manuscripts. Finally, 113 publications between the year 2001 until 2021 were included in the systematic review. Table 2 summarizes the publications in terms of their type of satellite used, ML techniques involved, water quality parameters extracted and significance or outcomes of studies.
4. Discussion
The majority of the reviewed studies demonstrated that ML can be effectively applied to learn about water quality monitoring via satellite or remote sensing. This section discusses the insight that can be learned from the reviewed studies.

4.1. Importance of water quality monitoring
A variety of indicators are often used to assess water quality, i.e., turbidity, suspended solids, concentrations of Chl-a, pollution-sediment, DO, CDOM, nutrients (TP, TN, ammonia-nitrogen, nitrate, orthophosphate, silicate), and harmful algae, etc. while water temperature, salinity and many other pollutants are also used as water quality indicators. Nutrient and sediment loads have an impact on water quality. Excess nitrogen and/or phosphorus can lead to eutrophication and fish deaths by increasing algal blooms and aquatic plant growth. The terms suspended-sediment concentration (SSC) and total suspended solids (TSS) are frequently used interchangeably to denote pollution-sediment which is a crucial parameter to consider because of its environmental, economic, and human health implications. [4, 21, 30, 65]. E. coli and cyanobacteria are hazardous organisms that can limit public usage of lakes and coastal waters by lowering dissolved oxygen levels and producing taste and odor problems. Significantly, microcystins, which have been related to liver cancer and tumors in people and animals, have been identified [16, 46, 102, 104]. Monitoring water quality parameters such as Chl-a concentration is crucial in fisheries studies, management, and harvesting since environmental factors impact the number and distribution of fish species for example skipjack tuna [26, 33].
Table 2. Summary of 113 eligible studies.

| No. | Satellite/remote sensing data | Water quality parameters involved | Significance/outcomes |
|-----|-------------------------------|-----------------------------------|-----------------------|
| 1   | Sentinel-2                    | TP, TN, COD                       | ANN exhibited the best performance followed by RF and SVM regression [6]. |
| 2   | Sentinel-3/OLCI               | Chl-a                             | SVM regression outperformed ANN, KRR and GPR [7]. |
| 3   | GOCI                          | COD, DO, DIP, DIN, oil, pH. salinity | GNNWR outperformed MLP, OLR, GWR and IDW [8]. |
| 4   | MODIS/Aqua                    | temperature                       | The proposed algorithm which has CNN produced promising outcomes [9]. |
| 5   | SpecWa                        | Chl-a                             | One-dimensional CNN (1D CNN) outperforms RF and BR [11]. |
| 6   | Sentinel-3/OLCI               | POC                               | LightGBM outperforms the traditional methods [12]. |
| 7   | SeaWiFS, MERIS, MODIS/Aqua, VIIRS | Salinity                  | XGBoost was the most robust among ANN, SVM and BR [13]. |
| 8   | Sentinel-2, Landsat-8         | Chl-a, CDOM, and suspended solids | MLP was reported significant [15]. |
| 9   | COMS/GOCI, DISCOVR/EPIC, FengYun-3D/ MERSI-II, GCOM-C/SLSTR, ISS/HICO, Landsat-8/OLI, MODIS/Aqua, SeaWiFS, Sentinel-2/MSI, Sentinel-3/OLCI, SNPP/VIIRS | E. coli, intestinal enterococci, total coliforms, faecal coliforms | Prediction of bathing water quality using DT and kNN on selected bands is more accurate than based on all bands [16]. |
| 10  | Sentinel-3/OLCI, Sentinel-3/SLSTR | Chl-a                             | In the winter and summer days, MPR and SVM regression-RBF show the greatest results. Furthermore, LSTM has a lower sensitivity [2]. |
| 11  | Sentinel-2A/MSI, MODIS         | Microphytobenthos                 | RF was reported significant [17]. |
| 12  | Sentinel-2                      | Chl-a                            | XGBoost provides practical approaches [18]. |
| 13  | Landsat-8/OLI                  | Chl-a                             | SVM regression model produced promising outcomes [19]. |
| 14  | SeaWiFS, MODIS/Aqua, VIIRS, OCI | temperature, salinity            | Combining non-linear ANN + HAC produced promising outcomes [20]. |
| 15  | Sentinel-2A                      | Chl-a, suspended solids           | Cubist bestowed significant accuracy [21]. |
| 16  | SeaWiFS, MERIS, MODIS/Aqua, VIIRS | salinity                         | RF outperformed DT and MLP [22]. |
| 17  | VIIRS                          | Chl-a                             | The BRT model was used to predict fish species [23]. |
| 18  | Landsat-8, Sentinel-2          | BGA, Chl-a, fDOM, DO, SC, and turbidity | In comparison to in-situ data, RF has a greater accuracy on satellite observations [24]. |
| 19  | AVHRR, AMSR-E, MODIS/Aqua     | temperature, Chl-a                | The proposed method, SVM regression, MLR and ELM regression performed significantly on particular parameters [25]. |
| 20  | MODIS/Aqua                     | temperature, salinity, Chl-a      | SVM, BRT, RF, MARS, GAM, CART, MLP, RPART, and MAXENT algorithms all performed better than the others, while the FDA, MDA, BIOLC, DOM, MAXL, MAHA, and RBF algorithms performed poorly [26]. |
| 21  | MODIS/Aqua, VIIRS              | POC + temperature, salinity, density, spiciness | Both RF regressor and RLR show promising results [28]. |
| 22  | MERIS, Sentinel-3/OLCI         | Chl-a                             | SVM outperformed the other ML models in other studies [29]. |
| 23  | MODIS                          | salinity, Chl-a, suspended solids, temperature | RF and SVM outperformed GLM and GAM [30]. |
| 24  | MODIS/2A/B                     | Chl-a                             | CNN performs better than SVM regression [31]. |
| 25  | MODIS                          | temperature, Chl-a                | GPR yields better results than SVM, MLP, RFR and MLR [32]. |
| 26  | MODIS/Aqua                     | Chl-a, suspended solids           | DT acquires a better performance than GLM [33]. |
| 27  | Landsat-8/OLI                  | salinity                          | ANN gives accurate results for a particular level of atmosphere [35]. |
| 28  | GEE, Landsat-8, Sentinel-2     | BGa phycocyanin, Chl-a, DO, SC, fDOM, turbidity, and pollution-sediments | DNN and LSTM outperformed PLSR and SVM regression for certain water quality parameters [4]. |
| 29  | MERIS                          | Turbidity, SDD                    | The models performed well with slight outperformance for GPR followed by LR, SVM regression and RF regression [37]. |
### Table 3. Summary of 113 eligible studies (continued).

| No. | Satellite/remote sensing data | Water quality parameters involved | Significance/outcomes |
|-----|--------------------------------|----------------------------------|-----------------------|
| 35  | MERIS                           | Chl-a                            | The MLP outperforms the SVM regression to capture satellite Chl-a [38]. |
| 36  | MODIS/Aqua                      | salinity, Chl-a, temperature, CDOM | Polynomial regression was reported significant [39]. |
| 37  | Sentinel-2-MSI                  | Chl-a                            | OCN outperformed SVM, GPR, C2RCC, BRT, BR and OC4M [40]. |
| 38  | GEE                            | DO, salinity, Chl-a, and pH       | RF bestowed significant accuracy [41]. |
| 39  | CZCS, SeaWiFS                   | Chl-a                            | SVM has high efficiency for the study [42]. |
| 40  | Landsat-8                       | Chl-a, TP, TN                     | The ensemble of ANN, SVM, RF and KNN increased the inversion water quality parameter results [43]. |
| 41  | Aquarius                        | salinity                          | The DNN algorithm outperforms SVM, GPR and KRR in terms of performance [44]. |
| 42  | MODIS                           | Chl-a                            | SVM regression showed the best performance among ANN, RF, CI-OC3M, MLR, GAR, PCR and GBDM [45]. |
| 43  | Sentinel-3/OLCI                 | cyanobacterial                    | Hierarchical Bayesian Spatio-temporal modeling shows high performance with low Deviance Information Criterion (DIC) value [46]. |
| 44  | VIIRS                           | Chl-a                            | SVM method produced promising outcomes [47]. |
| 45  | Sentinel-3/OLCI                 | Chl-a                            | GPR demonstrated to have strong generalization capabilities [48]. |
| 46  | MODIS/Aqua                      | temperature, salinity, Chl-a     | The RF regression ensemble-based model showed satisfactory performance among SVM, DT, MLR, PCR and MNR [49]. |
| 47  | MODIS, VIIRS                    | temperature                       | The ADtrees classifier outperformed the BDtrees [50]. |
| 48  | Landsat-5, Landsat-8, Sentinel-2, Sentinel-3, EnMAP and Hyperion | Chl-a                            | The performance of ANN, SVM, RF and MARS in terms of regression is comparable [51]. |
| 49  | MODIS/Aqua                      | Chl-a                            | XGBoost produced promising outcomes [52]. |
| 50  | Sentinel-1, Sentinel-2           | Suspended solids, dissolved solids | SVM outperforms RF, DT and kNN [53]. |
| 51  | Landsat-5, Landsat-7, Landsat-8 | Suspended solids, Chl-a, turbidity | ANN outperformed SVM, RF and Cubist [54]. |
| 52  | AMSR-2, SMOS                    | salinity                          | RF model outperforms GBDT model [55]. |
| 53  | MODIS                           | Chl-a                            | The RF regression ensemble produced promising outcomes [56]. |
| 54  | Landsat-8, GEE, Sentinel-2      | turbidity, suspending solids, TP, TN | SVM provided higher accuracy than ANN [57]. |
| 55  | MODIS/Terra                     | Turbidity, temperature            | ANN relatively good accuracy [58]. |
| 56  | SeaWiFS, MERIS, MODIS/Aqua      | Chl-a, temperature                | ET model shows better performance than the RF model [59]. |
| 57  | MODIS/Terra, Landsat-7/ETM+     | Chl-a, cyanobacterial             | RF model produced promising outcomes [60]. |
| 58  | MODIS/Aqua                      | Chl-a                            | LR was used to compare retrieval algorithm, OC3M [61]. |
| 59  | VIIRS, MODIS, SeaWiFS           | temperature, salinity, Chl-a      | ANN demonstrates a very good ability to generalize in terms of both space and time [62]. |
| 60  | AVHRR                           | temperature, salinity             | ANN technique successfully predicts early drift paths of green tides [63]. |
| 61  | Landsat-8/OLI                   | Chl-a, suspending solids, TP, TN  | Multiple Regression was reported significant [64]. |
| 62  | MODIS                           | Suspended solids                  | The best result was achieved by RF, which was followed by SVM, GAM, GLM, BART, MARS and CART [65]. |
| 63  | SeaWiFS, AVHRR                  | Chl-a, temperature                | The best algorithm was RF, which was followed by SVM, ANN, and PLSR [66]. |
| 64  | Landsat-5, Landsat-7            | Chl-a, CDOM, suspended solids     | The BRT model predictive performance was significantly better than MLR [67]. |
| 65  | Sentinel-2                      | algal                            | LSE, RBF and ANFIS have good performance proving the reliability and accuracy [68]. |
| 66  | MODIS/Aqua, MERIS               | Chl-a, CDOM                       | GPR and SVM regression models are confirmed to show stronger regression power than PLSR [69]. |
| 67  | MODIS/Aqua                      | Chl-a, temperature                | Multivariate Regression model was reported significant [70]. |
| 68  | Landsat-8/OLI                   | Chl-a                            | SVM regression showed slightly superior performance than ANN [71]. |
| 69  | SeaWiFS, MODIS/Aqua, MODIS/Terra | CDOM, POC, DOC                  | MLR outperform RF tree-bagger [72]. |
| 70  | Sentinel-3/OLCI                 | Chl-a, CDOM, suspended solids     | GPR was reported significant [73]. |
| 71  | Sentinel-3/OLCI                 | Chl-a, CDOM, suspended solids     | SVM, RF, KRR and GPR are very efficient except for RLR [74]. |
| 72  | SMOS                            | Temperature, salinity             | RF outperformed SVM regression [75]. |
| 73  | Landsat-4, Landsat-5, Landsat-7/ETM+, Landsat-8/OLI | Suspended solids | By a significant margin, ELM outperformed FFNN and CFNN, as well as RF and SVM [76]. |
| 74  | MODIS/Aqua                      | temperature                       | ANN produced promising outcomes [77]. |
| 75  | SeaWiFS, MODIS/Aqua, MERIS, VIIRS | suspended solids             | The validation of the generated product is quite important when using HMM [78]. |
| 76  | MERIS, MODIS/Aqua, SeaWiFS      | Chl-a                            | GMM uses for clustering and assimilated into the 1D models using the SEIK filter [79]. |
| 77  | MERIS, MODIS/Aqua, SeaWiFS      | salinity                         | MLP showed the best performance among SVM, RF and DT [80]. |
Table 4. Summary of 113 eligible studies (continued).

| No. | Satellite/remote sensing data | Water quality parameters involved | Significance/outcomes |
|-----|-------------------------------|----------------------------------|------------------------|
| 78  | GOCI                          | temperature, salinity, Chl-a, CDOM | RF generally performed better than SVM regression [81]. |
| 79  | VIIRS/NPP, MODIS/Terra, MODIS/Aqua | Chl-a, TP, TN, SDD | ANN strongly demonstrates effectiveness and reliability [82]. |
| 80  | AVHRR, SeaWiFS                | Chl-a, temperature | BRT showed slightly higher prediction performance than GAM and GLM [83]. |
| 81  | VIIRS, MERIS, MODIS/Aqua, SeaWiFS | Chl-a | BRT performs the best in terms of prediction [84]. |
| 82  | MODIS/Aqua, VIIRS             | Chl-a | MLP produced promising outcomes [85]. |
| 83  | MODIS                         | Salinity | SVM outperforms the classical approach [86]. |
| 84  | GOCI                          | Phytoplankton, suspended solids, CDOM | SVM produced promising outcomes [87]. |
| 85  | MERIS, MODIS                  | Chl-a | SVM combine with Linear, polynomial, RBF, sigmoid regression analysis improves the precision of the algorithm [88]. |
| 86  | MODIS/Aqua                    | Salinity | GAM outperformed an ANN [89]. |
| 87  | VIIRS                         | temperature, salinity, Chl-a | MLP method produced promising outcomes [90]. |
| 88  | Landsat-5/TM                  | suspended solids | LR and quadratic models perform better than Logarithmic, Power and Exponential [91]. |
| 89  | MERIS, MODIS                  | Chl-a | LR, polynomial, exponential functions and PCA were used for the partitioning mechanism. SVM method is used for the iterative classification process [92]. |
| 90  | SMOS, Aquarius                | salinity, temperature | SVM produced promising outcomes [93]. |
| 91  | MODIS                         | suspended solids | MLP method produced a superior performance [94]. |
| 92  | MODIS/Terra                   | total organic carbon (TOC) | The ANN model was chosen among GP and ELM for the forecasting method [95]. |
| 93  | Landsat-5/TM, Landsat-7/ETM+, MODIS/Terra | Chl-a, SDD | GP shows satisfactory results [96]. |
| 94  | AVHRR, SeaWiFS, MODIS/Aqua   | Chl-a, CDOM, temperature, | ANN was used for classification data and MLR was used to reconstructed pCO₂ [97]. |
| 95  | MERIS                         | Chl-a, Chl-b, Chl-c, PSC, PPC | MLR significant on ship-based reflectance measurements [98]. |
| 96  | MODIS                         | Chl-a | Algorithms using SVM are able to give better results than DT and Log-linear [99]. |
| 97  | MODIS/Terra, MERIS, Landsat/TM | microcystin | GP use for data mining purpose [100]. |
| 98  | GOCI                          | Chl-a, suspended solids | SVM regression outperformed RF, Cubist [101]. |
| 99  | Landsat, MODIS               | microcystin | GP model performed slightly better than ANN [102]. |
| 100 | MODIS                        | TP | GP use for spatiotemporal mapping [103]. |
| 101 | MODIS/Terra, Landsat-5, Landsat-7 | microcystin | GP was able to produce a positive correlation with the observed results [104]. |
| 102 | MODIS/Aqua                   | Salinity, Chl-a, temperature | AN was reported significant [105]. |
| 103 | MODIS/Terra                  | SDD, suspended solids, Chl-a | GP advantage was identified and been selected among ANN and MLR for estimation [106]. |
| 104 | MODIS                        | Chl-a, temperature | Bayesian network produced promising outcomes [107]. |
| 105 | SPOT-5                       | COD, ammonia-nitrogen, permanganate index (COD₇₅) | Combine the GA–SVM regression model is significantly better than MLR [108]. |
| 106 | MERIS                        | Chl-a | MLP was able to show prediction performance [109]. |
| 107 | MODIS, SeaWiFS, AVHRR, GEE   | Chl-a, CDOM, temperature, salinity, suspended solids | Gaussian use for creating the 3D surface optical forecast [110]. |
| 108 | MERIS                        | Suspended solids, Chl-a | SVM use for estimation [111]. |
| 109 | SeaWiFS                      | Suspended solids, Chl-a, nitrate, orthophosphate, silicate, salinity, temperature | In terms of determining temporal and spatial variability, MLR performed well [112]. |
| 110 | SeaWiFS, AVHRR               | Chl-a, temperature | SOM has produced robust estimation [113]. |
| 111 | SeaWiFS                      | CDOM, suspended solids, temperature, salinity | MLR produced promising outcomes [114]. |
| 112 | SeaWiFS                      | Chl-a, temperature | “Broken-stick” regression able to show prediction performance [115]. |
| 113 | AVHRR                        | suspended solids, SDD | LR was reported significant [116]. |
4.2. Remote sensing for water quality

Optical and thermal sensors collect water quality information with a great spectrum and spatial resolution. Watershed scale models based on ocean color satellite data have been constructed for determining optical active components (OAC) such as Chl-a, suspended solids and CDOM. However, existing satellites cannot directly monitor all water quality parameters, including nutrient concentrations, DO and COD levels, and microorganisms/pathogens, because some of these variables are not optically active, or because there is an absence of hyperspectral data at precise spatial resolutions. Therefore, some studies used OAC as a proxy to estimate non-OAC parameters by determining their relationship [57] and also use possible band compositions from satellite imagery bands [6].

4.3. Machine learning application

Numerous studies have been conducted to determine water quality using satellite data. The majority of the research relied on empirical relationships between satellite-derived reflectance and target water quality parameters to apply relatively simple linear or nonlinear regressions on satellite data. The empirical models produced have a limitation in that they may not operate effectively in diverse environments (such as the open ocean, coastal, river or inland waters). As a result, additional in-situ data is required, as well as parameter values that have been optimized [101]. Moreover, numerous ML models, which are sophisticated nonlinear data-driven approaches that have been tested and widely utilized. Some studies applied ML model comparison and select the best performance ML method to implement for their research. Other studies use the ML method to make improved measuring techniques such as fluorescence line height measurement [7], SSS measurement [9], estimation of POC [13], reduce spectral noise [19], reconstruct missing value [24] and atmospheric correction [35].

4.4. Limitation

This systematic review has numerous limitations that should be acknowledged. Firstly, ML-related keywords included in the search queries are not enough to cover as many related publications as possible. Therefore, this process might miss some studies that failed to be retrieved. Secondly, there is no review of performance used in method evaluation for ML. Thirdly, this review does not include bibliometric analysis to show the research trends.

5. Conclusion

This systematic review summarized how ML has been applied on satellite data to study water quality issues. The initial search process resulted in 1796 publications, and by refining the search by removing 473 duplicates publication, excluded 1196 non-related topics publications. Through the screening of 127 publications, 113 papers have been selected for data extraction and synthesis. Results also showed that there is a huge variety of ML methods suggested especially on the retrieval of water quality parameters. The most common ML approaches were ANN, SVM, RF, DT, MLP, cubist and GP for monitoring water quality at regional and global scales. According to the systematic analysis, ML can be successfully extended to water quality monitoring, allowing researchers to forecast and learn from natural processes in the environment, as well as assess human impacts on an ecosystem. These initiatives will also aid policymakers and water resource managers in taking proactive actions to prevent the negative consequences of water pollution through restoration projects, as well as ensure that environmental regulatory rules are followed.

Disclosure statement

The authors declared that there is no conflict of interest that could have influenced any part or the entirety of the work reported in this manuscript.
References
[1] Diersing N, Keys F and Marine N 2009 Water quality: frequently asked questions Florida Keys Natl. Mar. Sanctuary 8 5–6.
[2] Mohebzadeh H and Lee T 2021 Spatial downscaling of MODIS chlorophyll-a with machine learning techniques over the west coast of the Yellow Sea in South Korea J. Oceanogr. 77 103–22.
[3] Kim H C, Son S, Kim Y H, Khim J S, Nam J, Chang W K, Lee J H, Lee C H and Ryu J 2017 Remote sensing and water quality indicators in the Korean West coast: spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids Mar. Pollut. Bull. 121 425–34.
[4] Sagan V, Peterson K T, Maimaitijiang M, Sidike P, Sloan J, Greeling B A, Maalouf S and Adams C 2020 Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing Earth-Science Rev. 205 103187.
[5] Page M J, McKenzie J E, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D, Shamseer L, Tetzlaff J M, Aké E A, Brennan S E, Chou R, Glanville J, Grimshaw J M, Hróbjartsson A, Lalu M M, Li T, Loder E W, Mayo-Wilson E, McDonald S, McGuinness L A, Stewart L A, Thomas J, Tricco A C, Welch V A, Whiting P and Moher D 2021 The PRISMA 2020 statement: an updated guideline for reporting systematic reviews BMJ 372.
[6] Guo H, Huang J J, Chen B, Guo X and Singh V P 2021 A machine learning-based strategy for estimating non-optically active water quality parameters using Sentinel-2 imagery Int. J. Remote Sens. 42 1841–66.
[7] Tenjo C, Ruiz-verdú A, Van Wittenberghe S, Delegido J and Moreno J 2021 A new algorithm for the retrieval of sun induced chlorophyll fluorescence of water bodies exploiting the detailed spectral shape of water-leaving radiance Remote Sens. 13 1–19.
[8] Du Z, Qi J, Wu S, Zhang F and Liu R 2021 A spatially weighted neural network based water quality assessment method for large-scale coastal areas Environ. Sci. Technol. 55 2553–63.
[9] Bao S, Zhang R, Wang H, Yan H, Chen J and Wang Y 2021 Correction of satellite sea surface salinity products using ensemble learning method IEEE Access 201–1.
[10] Xavier Prochaska J, Cornillon P C and Reiman D M 2021 Deep learning of sea surface temperature patterns to identify ocean extremes Remote Sens. 13 1–18.
[11] Maier P M, Keller S and Hinz S 2021 Deep learning with WASI simulation data for estimating chlorophyll a concentration of inland water bodies Remote Sens. 13 1–27.
[12] Su H, Lu X, Chen Z, Zhang H and Lu W 2021 Estimating coastal chlorophyll-a concentration from time-series OLCI data based on machine learning Remote Sens. 13 576.
[13] Liu H, Li Q, Bai Y, Yang C, Wang J, Zhou Q, Hu S, Shi T, Liao X and Wu G 2021 Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods Remote Sens. Environ. 256 112316.
[14] Bayati M and Danesh-Yazdi M 2021 Mapping the spatiotemporal variability of salinity in the hypersaline Lake Urmia using Sentinel-2 and Landsat-8 imagery J. Hydrol. 595 126032.
[15] Fan Y, Li W, Chen N, Ahn J H, Park Y J, Kratzer S, Schroeder T, Ishizaka J, Chang R and Stamnes K 2021 OC-SMART: A machine learning based data analysis platform for satellite ocean color sensors Remote Sens. Environ. 253 112236.
[16] Senta A and Šerić L 2021 Remote sensing data driven bathing water quality assessment using Sentinel-3 Indones. J. Electr. Eng. Comput. Sci. 21 1634–47.
[17] Oiry S and Barillé L 2021 Using sentinel-2 satellite imagery to develop microphytobenthos-based water quality indices in estuaries Ecol. Indic. 121 107184.
[18] Cao Z, Ma R, Duan H, Pahlevan N, Melack J, Shen M and Xue K 2020 A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes Remote Sens. Environ. 248 111974.
[19] Hu C, Feng L and Guan Q 2020 A machine learning approach to estimate surface chlorophyll a
concentrations in global oceans from satellite measurements. *IEEE Trans. Geosci. Remote Sens.* **59** 4590–607.

[20] Arnault S, Thiria S, Crépon M and Kaly F 2020 A tropical Atlantic dynamics analysis by combining machine learning and satellite data *Adv. Sp. Res.* **68** 467–486.

[21] Saberioon M, Brom J, Nedbal V, Souček P and Císař P 2020 Chlorophyll-a and total suspended solids retrieval and mapping using Sentinel-2A and machine learning for inland waters *Ecol. Indic.* **113** 106236.

[22] Campbell A M, Racault M F, Goul t S and Laur enson A 2020 Cholera risk: A machine learning approach applied to essential climate variables *Int. J. Environ. Res. Public Health* **17** 1–24.

[23] Welch H, Brodie S, Jacox M G, Robinson D, Wilson C, Bograd S J, Oliver M J and Hazen E L 2020 Considerations for transferring an operational dynamic ocean management tool between ocean color products *Remote Sens. Environ.* **242** 111753.

[24] Park J, Kim H C, Bae D and Jo Y H 2020 Data reconstruction for remotely sensed chlorophyll-a concentration in the Ross Sea using ensemble-based machine learning *Remote Sens.* **12**.

[25] Peterson K T, Sagan V and Sloan J J 2020 Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing *GScience Remote Sens. Environ.* **57** 510–25.

[26] Mugo R and Saitoh S I 2020 Ensemble modelling of skipjack tuna (Katsuwonus pelamis) habitats in the western north pacific using satellite remotely sensed data; a comparative analysis using machine-learning models *Remote Sens. Environ.* **12** 2591.

[27] Fu Z, Hu L, Chen Z, Zhang F, Shi Z, Hu B, Du Z and Liu R 2020 Estimating spatial and temporal variation in ocean surface pCO$_2$ in the Gulf of Mexico using remote sensing and machine learning techniques *Sci. Total Environ.* **745** 140965.

[28] Sauzéde R, Johnson J E, Claustre H, Camps-Valls G and Ruescas A B 2020 Estimation of Oceanic Particulate Organic Carbon with Machine Learning *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **5** 949–56.

[29] Guan Q, Feng L, Hou X, Schurgers G, Zheng Y and Tang J 2020 Eutrophication changes in fifty large lakes on the Yangtze Plain of China derived from MERIS and OLCI observations *Remote Sens. Environ.* **246** 111890.

[30] DeLuca N M, Zaichik B F, Guikema S D, Jacobs J M, Davis B J K and Curriero F C 2020 Evaluation of remotely sensed prediction and forecast models for Vibrio parahaemolyticus in the Chesapeake Bay *Remote Sens. Environ.* **250** 112016.

[31] Yu B, Xu L, Peng J, Hu Z and Wong A 2020 Global chlorophyll-a concentration estimation from moderate resolution imaging spectroradiometer using convoluted neural networks *J. Appl. Remote Sens.* **14** 034520.

[32] Nguyen H Q, Ha N T and Pham T L 2020 Inland harmful cyanobacterial bloom prediction in the eutrophic Tri An Reservoir using satellite band ratio and machine learning approaches *Environ. Sci. Pollut. Res.* **27** 9135–51.

[33] Alfatinah A and Chu H J 2020 Integration of satellite-based environmental data for Skipjack tuna fishing ground determination *40th Asian Conf. Remote Sensing, ACRS 2019: Progress Remote Sens. Technol. Smart Futur.* 1–9.

[34] Fan Y, Li S, Han X and Stammes K 2020 Machine learning algorithms for retrievals of aerosol and ocean color products from FY-3D MERSI-II instrument *J. Quant. Spectrosc. Radiat. Transf.* **250** 107042.

[35] Medina-Lopez E 2020 Machine learning and the end of atmospheric corrections: A comparison between high-resolution sea surface salinity in coastal areas from top and bottom of atmosphere Sentinel-2 imagery *Remote Sens.* **12** 1–21.

[36] Ansari M and Akhoondzadeh M 2020 Mapping water salinity using Landsat-8 OLI satellite images (Case study: Karun basin located in Iran) *Adv. Sp. Res.* **65** 1490–502.

[37] Arias-Rodriguez L F, Duan Z, Sepúlveda R, Martinez-Martinez S I and Disse M 2020 Monitoring water quality of Valle de Bravo reservoir, Mexico, using entire lifespan of
MERIS data and machine learning approaches Remote Sens. 12 1586.

[38] Martinez E, Brini A, Gorgues T, Drumetz L, Roussillon J, Tandeo P, Maze G and Fablet R 2020 Neural network approaches to reconstruct phytoplankton time-series in the global ocean Remote Sens. 12 1–13.

[39] Wouthuysen S, Kusmanto E, Fadli M, Harsono G, Salamena G, Lekalette J and Syahailatua A 2020 Ocean color as a proxy to predict sea surface salinity in the Banda Sea IOP Conf. Ser. Earth Environ. Sci. 618 012037.

[40] Asim M, Brekke C, Mahmood A, Eltoft T and Reigstad M 2020 Ocean Color Net (OCN) for the Barents Sea Int. Geo sci. Remote Sens. Symp. 5881–4.

[41] Yñiguez A T and Ottong Z J 2020 Predicting fish kills and toxic blooms in an intensive mariculture site in the Philippines using a machine learning model Sci. Total Environ. 707 136173.

[42] Martinez E, Gorgues T, Lengaigne M, Fontana C, Sauzède R, Menkes C, Uitz J, Di Lorenzo E and Fablet R 2020 Reconstructing global chlorophyll-a variations using a non-linear statistical approach Front. Mar. Sci. 7 1–20.

[43] Xiaojuan L, Mutao H and Jianbao L 2020 Remote sensing inversion of lake water quality parameters based on ensemble modelling E3S Web Conf. 143 02007.

[44] Zhang L, Zhang R and He Q 2020 Sea surface salinity retrieval from aquarius in the south china sea using machine learning algorithm Int. Geosci. Remote Sens. Symp. 5643–6.

[45] He J, Chen Y, Wu J, Stow D A and Christakos G 2020 Space-time chlorophyll-a retrieval in optically complex waters that accounts for remote sensing and modeling uncertainties and improves remote estimation accuracy Water Res. 171 115403.

[46] Myer M H, Urquhart E, Schaeffer B A and Johnston J M 2020 Spatio-temporal modeling for forecasting high-risk freshwater cyanobacterial harmful algal blooms in Florida Front. Environ. Sci. 8 1–13.

[47] Ye H, Yang C, Tang S and Chen C 2020 The phytoplankton variability in the Pearl River estuary based on VIIRS imagery Cont. Shelf Res. 207 104228.

[48] Blix K and Eltoft T 2019 A generalized chlorophyll-a estimation model for complexity-diverse Arctic waters Int. Geosci. Remote Sens. Symp. 7861–4.

[49] Chen S, Hu C, Barnes B B, Wanninkhof R, Cai W J, Barbero L and Pierrot D 2019 A machine learning approach to estimate surface ocean pCO2 from satellite measurements Remote Sens. Environ. 228 203–26.

[50] Kilpatrick K A, Podestá G, Williams E, Walsh S and Minnett P J 2019 Alternating decision trees for cloud masking in MODIS and VIIRS NASA sea surface temperature products J. Atmos. Ocean. Technol. 36 387–407.

[51] Maier P M and Keller S 2019 Application of different simulated spectral data and machine learning to estimate the chlorophyll a concentration of several inland waters Work. Hyperspectral Image Signal Process. Evol. Remote Sens. 1–5.

[52] Ghatkar J G, Singh R K and Shanmugam P 2019 Classification of algal bloom species from remote sensing data using an extreme gradient boosted decision tree model Int. J. Remote Sens. 40 9412–38.

[53] Bangira T, Alifieri S M, Menenti M and van Niekerk A 2019 Comparing thresholding with machine learning classifiers for mapping complex water Remote Sens. 11 1351.

[54] Hafeez S, Kong H, Wong M S and Nazeer M 2019 Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-II Waters: A Case Study of Hong Kong. Remote Sens. 11 617.

[55] Su H, Yang X and Yan X H 2019 Estimating Ocean Subsurface Salinity from Remote Sensing Data by Machine Learning Int. Geosci. Remote Sens. Symp. 8139–42.

[56] Chen S, Hu C, Barnes B B, Xie Y, Lin G and Qiu Z 2019 Improving ocean color data coverage through machine learning Remote Sens. Environ. 222 286–302.

[57] Govedarica M and Jakovljevic G 2019 Monitoring spatial and temporal variation of water
quality parameters using time series of open multispectral data SPIE Proceedings 11174 55.
[58] Laureano-Rosario A E, Duncan A P, Symonds E M, Savie D A and Muller-Karger F E 2019 Predicting culturable enterococci exceedances at Escambron Beach, San Juan, Puerto Rico using satellite remote sensing and artificial neural networks J. Water Health 17 137–48.
[59] Park J, Kim J H, Kim H C, Kim B K, Bae D, Jo Y H, Jo N and Lee S H 2019 Reconstruction of ocean color data using machine learning techniques in polar regions: Focusing on off Cape Hallett, Ross Sea Remote Sens. 11 1366.
[60] Hu Y and Li L 2019 Remote sensing mapping of cyanobacteria blooms in chaohu based on spatio-temporal-spectrum fusion: Improvement on spatial scales J. Eng. Sci. Technol. Rev. 12 182–94.
[61] Abbas M M, Melesse A M, Scinto L J and Rehage J S 2019 Satellite estimation of chlorophyll-a using moderate resolution imaging spectroradiometer (MODIS) sensor in shallow coastal water bodies: Validation and improvement Water (Switzerland) 11 1621.
[62] Krasnopolsky V, Nadiga S, Mehra A and Bayler E 2018 Adjusting neural network to a particular problem: Neural network-based empirical biological model for chlorophyll concentration in the upper ocean Appl. Comput. Intell. Soft Comput.
[63] Hu P, Liu Y, Hou Y and Yin Y 2018 An early forecasting method for the drift path of green tides: A case study in the Yellow Sea, China Int. J. Appl. Earth Obs. Geoinf. 71 121–31.
[64] Lim J and Choi M 2015 Assessment of water quality based on Landsat 8 operational land imager associated with human activities in Korea Environ. Monit. Assess. 187 1–17.
[65] DeLuca N M, Zaitchik B F and Curriero F C 2018 Can multispectral information improve remotely sensed estimates of total suspended solids? A statistical study in Chesapeake Bay Remote Sens. 10 1–16.
[66] Hu S, Liu H, Zhao W, Shi T, Hu Z, Li Q and Wu G 2018 Comparison of machine learning techniques in inferring phytoplankton size classes Remote Sens. 10 1–18.
[67] Lin S, Qi J, Jones J R and Stevenson R J 2018 Effects of sediments and coloured dissolved organic matter on remote sensing of chlorophyll-a using Landsat TM/ETM+ over turbid waters Int. J. Remote Sens. 39 1421–40.
[68] Nascimento Silva H A and Panella M 2018 Eutrophication Analysis of Water Reservoirs by Remote Sensing and Artificial Networks Prog. Electromagn. Res. Symp. 458–63.
[69] Blix K and Eltoft T 2018 Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval Remote Sens. 10 775.
[70] Karki S, Sultan M, Elkadiri R and Elbayoumi T 2018 Mapping and forecasting onsets of harmful algal blooms using MODIS data over coastalwaters surrounding charlotte county, Florida Remote Sens. 10 1656.
[71] Kown Y S, Baek S H, Lim Y K, Pyo J C, Ligaray M, Park Y and Cho K H 2018 Monitoring coastal chlorophyll-a concentrations in coastal areas using machine learning models Water (Switzerland) 10 1–17.
[72] Aurin D, Mannino A and Lary D J 2018 Remote sensing of CDOM, CDOM spectral slope, and dissolved organic carbon in the global ocean Appl. Sci. 8 2687.
[73] Blix K, Pálffy K, Tóth V R and Eltoft T 2018 Remote sensing of water quality parameters over Lake Balaton by using Sentinel-3 OLCI Water (Switzerland) 10 1–20.
[74] Ruescas A B, Mateo-Garcia G, Camps-Valls G and Hieronymi M 2018 Retrieval of case 2 water quality parameters with machine learning Int. Geosci. Remote Sens. Symp. 124–7.
[75] Su H, Li W and Yan X H 2018 Retrieving Temperature Anomaly in the Global Subsurface and Deeper Ocean From Satellite Observations J. Geophys. Res. Ocean. 123 399–410.
[76] Peterson K T, Sagan V, Sidike P, Cox A L and Martinez M 2018 Suspended sediment concentration estimation from landsat imagery along the lower missouri and middle Mississippi Rivers using an extreme learning machine Remote Sens. 10 1503.
[77] Sunder S, Raaj R and Ramakrishnan B 2017 ANN based estimation of daily sea surface temperature over Arabian sea using MODIS data 38th Asian Conf. Remote Sens. - Sp. Appl.
Touching Hum. Lives, ACRS 2017.

[78] Renosh P R, Jourdin F, Charantonis A A, Yala K, Rivier A, Badran F, Thiria S, Guillou N, Leckler F, Gohin F and Garlan T 2017 Construction of multi-year time-series profiles of suspended particulate inorganic matter concentrations using machine learning approach Remote Sens. 9 1320.

[79] Dreano D, Tsiaras K, Triantafyllou G and Hoteit I 2017 Efficient ensemble forecasting of marine ecology with clustered 1D models and statistical lateral exchange: application to the Red Sea Ocean Dyn. 67 935–47.

[80] Chen S and Hu C 2017 Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean color measurements Remote Sens. Environ. 201 115–32.

[81] Jang E, Im J, Park G H and Park Y G 2017 Estimation of fugacity of carbon dioxide in the east sea using in situ measurements and geostationary ocean color imager satellite data Remote Sens. 9 821.

[82] Chang N Bin, Bai K and Chen C F 2017 Integrating multisensor satellite data merging and image reconstruction in support of machine learning for better water quality management J. Environ. Manage. 201 227–40.

[83] Zhang X, Saitoh S I and Hirawake T 2017 Predicting potential fishing zones of Japanese common squid (Todarodes pacificus) using remotely sensed images in coastal waters of south-western Hokkaido, Japan Int. J. Remote Sens. 38 6129–46.

[84] Nieto K and Mélín F 2017 Variability of chlorophyll-a concentration in the Gulf of Guinea and its relation to physical oceanographic variables Prog. Oceanogr. 151 97–115.

[85] Sauzède R, Clausbre H, Uitz J, Jamet C, Dall’Olmo G, D’Ortenzio F, Gentili B, Poteau A and Schmechtig C 2016 A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: Retrieval of the particulate backscattering coefficient J. Geophys. Res. Ocean. 121 2552–71.

[86] Wattelez G, Dupouy C, Mangeas M, Lefèvre J, Touraivane and Frouin R 2016 A statistical algorithm for estimating chlorophyll concentration in the New Caledonian lagoon Remote Sens. 8 1–23.

[87] Zhang Y N and Zheng X S 2016 An improved algorithm for retrieval of aerosol optical properties over the Yellow Sea from Geostationary Ocean Color Imager Int. Geosci. Remote Sens. Symp. 4077–9.

[88] Davila J C and Zaremba M B 2016 An iterative learning framework for multimodal chlorophyll-a estimation IEEE Trans. Geosci. Remote Sens. 54 7299–308.

[89] Vogel R L and Brown C W 2016 Assessing satellite sea surface salinity from ocean color radiometric measurements for coastal hydrodynamic model data assimilation J. Appl. Remote Sens. 10 036003.

[90] Krasnopolsky V, Nadiga S, Mehra A, Bayer E and Behringer D 2016 Neural networks technique for filling gaps in satellite measurements: Application to ocean color observations Comput. Intell. Neurosci.
sensing-based water quality forecasting model *ICNSC 2015* - *2015 IEEE 12th Int. Conf. Networking, Sens. Control* 51–7.

[96] Doña C, Chang N Bin, Caselles V, Sánchez J M, Camacho A, Delegido J and Vannah B W 2015 Integrated satellite data fusion and mining for monitoring lake water quality status of the Albufera de Valencia in Spain *J. Environ. Manage.* 151 416–26.

[97] Parard G, Charantonis A A and Rutgerson A 2015 Remote sensing the sea surface CO2 of the Baltic Sea using the SOMLO methodology *Biogeosciences* 12 3369–84.

[98] Bracher A, Taylor M H, Taylor B, Dinter T, Röttgers R and Steinmetz F 2015 Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations *Ocean Sci.* 11 139–58.

[99] Wattelez G, Dupouy C, Mangeas M, Léfevre J, Touraiavane T and Frouin R J 2014 A statistical algorithm for estimating chlorophyll concentration from MODIS data *Ocean Remote Sens. Monit. from Sp.* 9261 92611S.

[100] Chang N Bin, Vannah B and Jeffrey Yang Y 2014 Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in lake erie *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7 2426–42.

[101] Kim Y H, Im J, Ha H K, Choi J K and Ha S 2014 Machine learning approaches to coastal water quality monitoring using GOCI satellite data *GIScience Remote Sens.* 51 158–74.

[102] Chang N Bin and Vannah B 2013 Comparative data fusion between genetic programing and neural network models for remote sensing images of water quality monitoring *Proc. - 2013 IEEE Int. Conf. Syst. Man, Cybern. SMC* 2013 1046–51.

[103] Chang N Bin, Xuan Z and Yang Y J 2013 Exploring spatiotemporal patterns of phosphorus concentrations in a coastal bay with MODIS images and machine learning models *Remote Sens. Environ.* 134 100–10.

[104] Chang N Bin and Vannah B 2013 Intercomparisons between empirical models with data fusion techniques for monitoring water quality in a large lake *2013 10th IEEE Int. Conf. Networking, Sens. Control, ICNSC* 2013 258–63.

[105] Geiger E F, Grossi M D, Trembanis A C, Kohut J T and Oliver M J 2013 Satellite-derived coastal ocean and estuarine salinity in the Mid-Atlantic *Cont. Shelf Res.* 63 S235–42.

[106] Chang N Bin, Yang Y J, Daranpob A, Jin K R and James T 2012 Spatiotemporal pattern validation of chlorophyll-a concentrations in Lake Okeechobee, Florida, using a comparative MODIS image mining approach *Int. J. Remote Sens.* 33 2233–60.

[107] Williamson R, Field J G, Shillington F A, Jarre A and Potgieter A 2011 A Bayesian approach for estimating vertical chlorophyll profiles from satellite remote sensing: Proof-of-concept *ICES J. Mar. Sci.* 68 792–9.

[108] Wang X, Fu L and He C 2011 Applying support vector regression to water quality modelling by remote sensing data *Int. J. Remote Sens.* 32 8615–27.

[109] González Vilas L, Spyrakos E and Torres Palenzuela J M 2011 Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician rias (NW Spain) *Remote Sens. Environ.* 115 524–35.

[110] Ladner S, Arnone R, Sandidge J, Ko D S, Casey B and Hall C 2009 “Ocean weather” in the gulf of Mexico: Exploiting real-time satellite ecological properties and circulation models for coastal ocean monitoring *MTS/IEEE Biloxi - Mar. Technol. Our Futur. Glob. Local Challenges, Ocean.* 2009 0–7.

[111] Tang S, Dong Q, Chen C, Liu F and Jin G 2009 Retrieval of suspended sediment concentration in the Pearl River estuary from MERIS using support vector machines *Int. Geosci. Remote Sens. Symp.* 3 239–42.

[112] Green R E and Gould R W 2008 A predictive model for sateHite-derived phytoplankton absorption over the Louisiana shelf hypoxic zone: Effects of nutrients and physical forcing *J. Geophys. Res. Ocean.* 113 1–17.

[113] Demarcq H, Richardson A J and Field J G 2008 Generalised model of primary production in the
southern Benguela upwelling system Mar. Ecol. Prog. Ser. 354 59–74.

[114] Green R E, Gould R W and Ko D S 2008 Statistical models for sediment/detritus and dissolved absorption coefficients in coastal waters of the northern Gulf of Mexico Cont. Shelf Res. 28 1273–85.

[115] Anderson T R, Spall S A, Yool A, Cipollini P, Challenor P G and Fasham M J R 2001 Global fields of sea surface dimethylsulfide predicted from chlorophyll, nutrients and light J. Mar. Syst. 30 1–20.

[116] Budd J W, Drummer T D, Nalepa T F and Fahnenstiel G L 2001 Remote sensing of biotic effects: Zebra mussels (Dreissena polymorpha) influence on water clarity in Saginaw Bay, Lake Huron Limnol. Oceanogr. 46 213–23.