MDPI

A Novel Early Warning System (EWS) for Water Quality, Integrating a High-Frequency Monitoring Database with Efficient Data Quality Control Technology at a Large and Deep Lake (Lake Qiandao), China

Liancong Luo [1,*], **Jia Lan** [2], **Yucheng Wang** [2], **Huiyun Li** [3,*], **Zhixu Wu** [2], **Chrisopher McBridge** [4], **Hong Zhou** [5], **Fenglong Liu** [6], **Rufeng Zhang** [1], **Falu Gong** [1], **Jialong Li** [1], **Lan Chen** [1] and **Guizhu Wu** [1]

1   Institute for Ecological Research and Pollution Control of Plateau Lakes,
    School of Ecology and Environmental Science, Yunnan University, Kunming 650504, China;
    hangrufeng1997@126.com (R.Z.); gongfl2022@163.com (F.G.); jlli@mail.yun.edu.cn (J.L.);
    15111718997@163.com (L.C.); wuguizhu0223@163.com (G.W.)
2   Chunan Branch, Hangzhou Municipal Ecology and Environment Bureau, Chunan 311700, China;
    lanika0916@163.com (J.L.); iceman0011@163.com (Y.W.); caepb@126.com (Z.W.)
3   Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China
4   Environmental Research Institute, The University of Waikato, Hamilton 3240, New Zealand;
    christopher.mcbride@waikato.ac.nz
5   Waikato Institute of Technology, Hamilton 3204, New Zealand; hong.zhou@wintec.ac.nz
6   Financial Department, Hunan University of Humanities, Science and Technology, Loudi 417000, China;
    13907384452@139.com
*   Correspondence: billluo@ynu.edu.cn (L.L.); hyli@niglas.ac.cn (H.L.);
    Tel.: +86-(0)-181-1290-7530 (L.L.); +86-(0)-159-0515-7570 (H.L.)

Abstract: To assess water quality (WQ) online for assuring the safety of drinking water, a novel early warning system integrating a high-frequency monitoring system (HFMS) and data quality control (QC) was developed at Lake Qiandao. The HFMS was designed for monitoring water quality, nutrient inputs by main tributaries, water currents and meteorology at different sites at Lake Qiandao. The EWS focused on data availability, a QC method, a statistical analysis method and data applications instead of technological aspects for sondes, wireless data transfer and interface software development. QC was implemented before use to delete the abnormal values of outliers, to detect change points, to analyse the change trend, to interpolate discrete missing measurements, and find continuous missing or wrong observations caused by technical problems with the sonde. For demonstrating advantages and data availability, surface and profiling measurements at two sites were plotted. The plots show obvious seasonal and diel variations, demonstrating the success of integration of the system with advanced automated technology and good QC. This successfully developed system is now not only giving early warning signals, but also providing critical WQ information for the security of drinking water diverted to Hangzhou city through a tunnel of 110 km length. The automatic monitoring data with QC is also being used to produce initial conditions for WQ prediction based on a three dimensional hydrodynamic-ecosystem model.

Keywords: early warning system; high-frequency monitoring; data quality control; water quality; Lake Qiandao

1. Introduction

Lake eutrophication is a long-term global problem caused by excess nutrient inputs [1], and exacerbated by long water residence times that delay WQ responses to management actions. It is a common problem in the lakes located at the middle and lower

catchment of the Yangtze River even in the "good WQ" lakes classified by the Ministry of Ecology and Environment of the People's Republic of China (e.g., Lake Qiandao). Impairments associated with eutrophication include poor water clarity, harmful algal blooms (HABs) [2–4], and the loss of biodiversity, which affect drinking water supplies and the recreational use of lakes. Some impairments are highly dynamic (e.g., HABs, loss of dissolved oxygen, etc.), which has resulted in the rapid proliferation of EWS for monitoring key variables that can cause rapid changes in the water quality of coastal water [5] and freshwater systems [6–10].

An EWS is an integrated system consisting of in situ autonomous sensors for continuous rapid monitoring. The measured data are analysed and interpreted for the purpose of forecasting changes in water quality by the system. It provides a fast and accurate way to distinguish abnormal/abrupt variations in WQ due to biochemical and physical interactions over short time scales. EWS requires the fast detection of abnormalities in WQ parameters, which calls for high-frequency real-time monitoring technologies, wireless communication and appropriate data storage and analysis. A new generation of online monitoring tools based on sensor sonde technology and satellite-based remote sensing (RS) has emerged in recent years [11–16]. However, the effective implementation of these tools has not been fully realised due to their limitations relating to meeting practical utility needs, high costs, unsatisfied reliability, hardware maintenance demands, and cumbersome data management and analysis approaches, with respect to practical operations.

Conventional sample collection and laboratory-based methods are too slow to achieve operational response and temporal–spatial continuity. There is a clear and increasing need to rapidly detect WQ parameters to ensure an appropriate and timely response to instances of accidental or deliberate contamination [13]. For the past two decades, Wireless Sensor Networks (WSNs) technology has been applied increasingly to environmental monitoring for providing high-frequency scientific data. These high-tech smart devices have offered a vital approach to environmental monitoring and have monitored some lesser-studied fundamental processes, due to their inaccessibility [12]. Generally, the sensor nodes acquire data autonomously, process them locally, and transfer the information to a base station with an internet connection [17].

WSN technology integrated with floating buoys has been widely used to acquire high-frequency WQ data for lakes in the world. Due to severe environmental problems, the Chinese great lakes, which were or are currently supplying drinking water, have been the focus of many efforts to build dense buoy monitoring networks. For example, on Lake Taihu (Jiangsu Province, surface area 2338 km², mean water depth 1.9 m), 18 WSN buoy stations are operated by the Chinese Academy of Science, 21 stations by the Jiangsu Environmental Monitoring Station [18], and one each by the Suzhou Meteorological Bureau and Nanjing Normal University. At Lake Dianchi (Yunnan Province, surface area ~300 km²), there are 30 monitoring buoy stations (MBSs) [19]. Lake Taihu and Lake Dianchi are both key lakes which have been invested in tremendously by the Chinese central government for ecological restoration over the past two decades. At Lake Qiandao, which is a drinking water supply reservoir with an area of 580 km² [20], there are four buoy profilers with a meteorological station, 10 MBSs for surface WQ detection with a meteorological station, 13 MBSs for inflow river WQ detection, and four for both surface WQ detection and current measurements by an Acoustic Doppler Current Profiler (ADCP).

Key issues that need to be addressed for a EWS in order to assure the accuracy and precision of measurements, are data quality control (QC) and quality assurance (QA). QC and QA are fundamental for decision making based on reliable data analysis. For a specific water quality parameter, QC generally involves a number of internal consistency tests, a threshold test, a step change point and trend detections for finding potential outliers at a particular station [21]. Measured data at a given site may also be compared with measurements from surrounding sites for an accuracy assessment. An effective QC and

QA system is critical to the success of any environmental project, which has been successfully applied to the fields of climatology, oceanography and other geosciences. However, there has been limited application to developing an EWS with real-time high-frequency monitoring. Therefore, the aims of this study were to introduce a comprehensive EWS. developed for Lake Qiandao, and the corresponding QC method for real-time high-frequency monitoring data.

2. Methodology

2.1. Study Area

Built in 1959, Lake Qiandao is located at Chun'an County, which is at the west of Zhejiang Province, China (29°22'–9°50'N, 118°34'–119°15'E, Figure 1) [22]. It is one of the largest reservoirs in China with a surface water area of 573 km² and a water capacity of 178.4×10^8 m³, when the water level is 108 m [23]. The mean water depth is 34 m and the maximum depth is 100 m. It is used to supply drinking water for the 450,000 people in Chun'an County. Now, it is also providing drinking water for five million people in Hangzhou City, through a tunnel with a length of ~110 km. There are 34 inflow tributaries around the lake. The largest one (Xinan River, Figure 1) is from the northwest, carrying 51.4% of the total inflow to the lake from all sources, not including rainfall and ground water. It carries 34.3% of the total phosphorus (TP) loading and 63.7% of the total nitrogen (TN) loading [19]. The multi-year average inflow and outflow are 103.44×10^8 m³ and 97.45×10^8 m³, respectively [23], with a residence time of ~668 days.
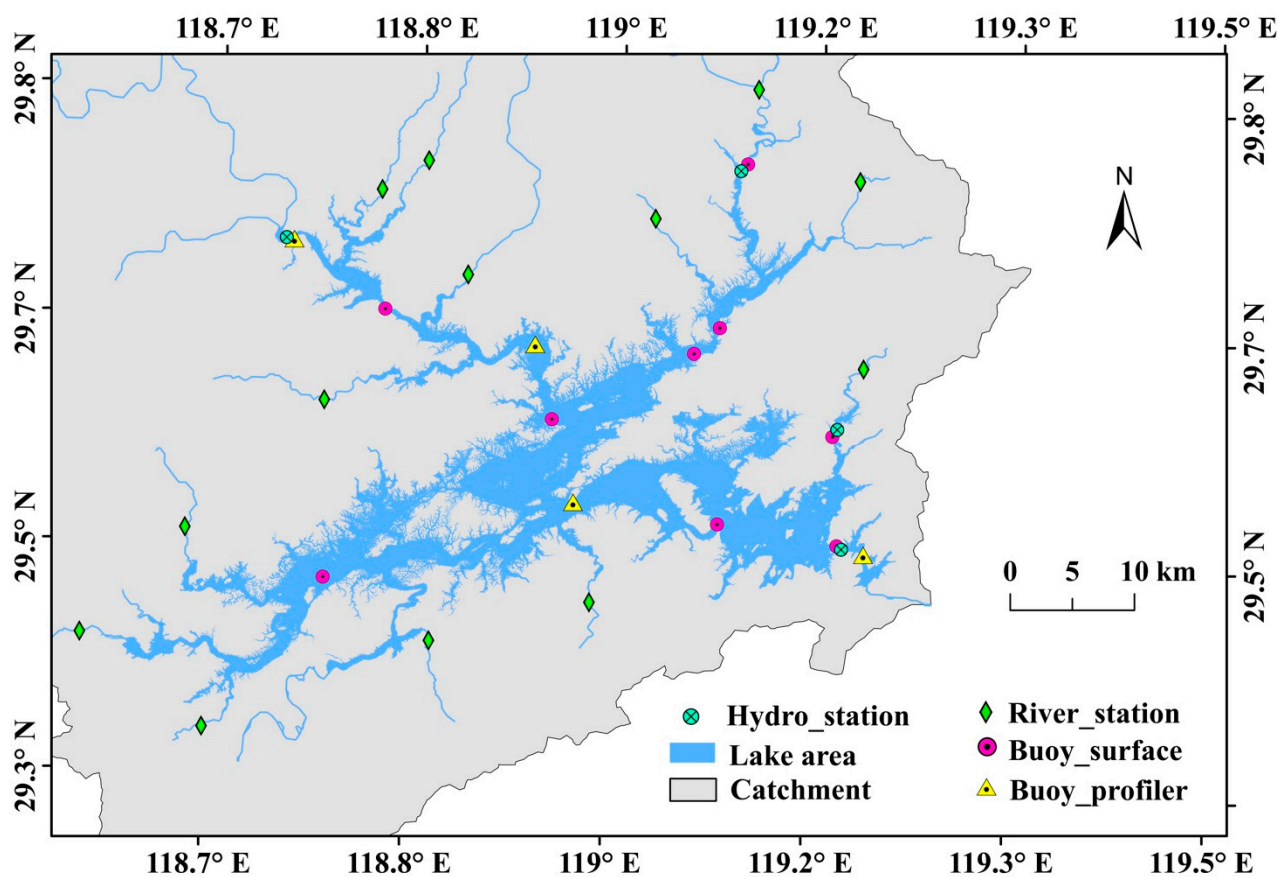

Figure 1. Monitoring sites at Lake Qiandao and its location within China.

2.2. Monitoring Stations

The HFMS at Lake Qiandao includes thirteen river stations; fourteen buoy stations—including ten buoys measuring the surface WQ (Buoy_surface, model EMM700, YSI Incorporated, Yellow Springs, USA); four 'profiler' buoys (Buoy_profiler, model EMM2500, Yellow Springs, USA); four hydrological stations (Hydro_station)—these were deployed at the three main tributaries and the only river outflow and measure water current speed; and four flux stations (Flux_station, model Tenghai HZF3, Tenghai Science & Technology Ltd., Hangzhou, China), located alongside the hydrological stations, measuring WQ parameters. There are also thirteen river stations (River_station, model EMM700, YSI Incorporated, Yellow Springs, USA) measuring WQ, deployed at the main inflow tributaries. Meteorological sensors (Met_station) are deployed at the top of each buoy station, except for one site (Figure 1). The Buoy_profilers were deployed at sites 1–4. The profiling information at site 3 is representative of the lake because it is at the centroid, while site 4 is the deepest monitoring point, which is located at the biggest outflow channel. Sites 5 and 6 with Buoy_surface are near the middle, capturing surface WQ variation.

2.3. Sensor Information and Alert Range

Table 1 shows all the sensor metadata. The alert range for a specific sensor, in the seventh column of Table 1, was decided by analysing historical data manually, or by using the sonde measurements. The alert thresholds are equal to the corresponding reasonable "minimum-maximum-range" (MMR) of each sonde. When the measured value is out of the alert range, an alert report will be recorded and the EWS can find the report by searching alert report once an hour. Once alert information for a specific sonde is found, a message will be sent to the EWS manager's cellphone. The integrated sonders with a metal protective cage move up and down through the water column at a constant speed The average return times for sonders moving at site 1, site 2, site 3 and site 4, are 55 min through a water column of 65 m, 45 min (water column 40 m), 50 min (water column 46 m), and 30 min (water column 16 m), respectively. The measurement values are recorded every minute at all the four sites. All the buoy systems are solar-powered and the data are transferred to a computer server at the Chunan Branch of Hangzhou Ecology and Environment Bureau by 4G wireless telemetry. The whole monitoring system is being maintained by Hangzhou Tenghai Science and Technology Limited, with the sondes cleaned to wipe bio-fouling once a month and calibrated once every three months for data assurance. The power supply system with a solar panel and wireless data transfer are also regularly checked and maintained by this company.

Table 1. Sensor metadata for all monitoring buoys at Lake Qiandao.

Buoy Type	Number of Buoys	Measured Parameter	Unit	Sensor Model	Measurement Range	Alert Range	Monitoring Frequency
Buoy_surface	10	WT	°C		−5–+50	9–35	
		PH			0–14	6–9	
		ORP	mV		−999–+999	−30–+500	
		COND	mS cm⁻¹		0–200	80–170	
		DO_con	mg L⁻¹	YSI	0–50	4–13	
		DO_sat	%	EXO2	0–500	39–170	30 min
		TURB	NTU		0–4000	0–43	
Buoy_profiler	4	CHLA	µg L⁻¹		0–400	0–25	
		PC	µg L⁻¹		0–100	0–7	
		FDOM	QSE		0–300	0–7	
River_station (EMM700)	13	WT	°C	YSI	−5–+50		
		COND	µS cm⁻¹	EXO2	0–200		2 h
		TURB	NTU		0–4000		
		FDOM	QSE		0–300		
Met_station	13	RH	%		0–100	35–90	
		BP	hpa		600–1100	970–1030	
		Wind_spd	m s⁻¹	VAISALA	0–60	0–13	30 min
		Wind_dir	°C	WXT520	0–360	0–360	
		TEMP	°C		−52–+60	1–33	
		RAIN	mm		0–200 mm h⁻¹	0–162 mm day⁻¹	
Flux_station	4	TN	mg L⁻¹			0.4–1.5	
		TP	mg L⁻¹			0–0.03	
		COD	mg L⁻¹	TriOS OPUS	0–500	4–11	1 h
		NO₃-N	mg L⁻¹		0–100	0–0.08	
		TOC	mg L⁻¹		0–500		
Hydro_Station	4	Current Speed	m s⁻¹	ADCP TRDI WHR600k	0–5	0–1	30 min

Abbreviations: water temperature (WT), oxidation reduction potential (ORP), electrical conductivity (COND), dissolved oxygen concentration (DO_con), dissolved oxygen saturation (DO_sat), turbidity (TURB), chlorophyll *a* (CHLA), phycocyanin (a pigment specific to cyanobacteria, PC), fluorescent dissolved organic matter (FDOM), total nitrogen (TN), total phosphorus (TP), chemical oxygen demand (COD), nitrate (NO_3-N), total organic carbon (TOC), relative humidity (RH), air pressure (BP), wind speed (Wind_spd), wind direction (Wind_dir) and air temperature (TEMP).

2.4. Data Quality Control

The monitoring stations produce large volumes of data, requiring specialised tools to facilitate quality control and to ensure that data are fit-for-purpose. We developed bespoke software in Fortran, employing two principle methods of quality control. Firstly, an MMR was adopted whereby the minimum and maximum values of the raw data measured by each sensor were specified, by assessing the range of previous observations and defining a 'reasonable range' (larger than or equal to the alert range at Table 1) for each variable based on a large volume of historical measurements from the lake area and inflows/outflows. The lowest value from both historical observations was adopted as the minimum value for MMR, with the maximum value defined with a similar method. Table 2 shows all the maximum (Max)/minimum (Min)/average (Avg) values for WQ measurements including WT, pH, DO, permanganate index (PI), chemical oxygen demand (COD), five-day biochemical oxygen demand (BOD₅), ammonia (NH₄-N), TP,

TN, CHLA and Secchi depth (SD). Unfortunately, only WT, pH and DO were observed by the monitoring buoys. Subsequently, data outside of the specified range for each variable were quarantined with a unique flag number (e.g., '8888') and will be further investigated.

Table 2. Statistical value of measured water quality parameters at the four sites of Lake Qiandao from April 2001 to May 2021.

Site	Number of Samples	Statistical Value	WT (°C)	pH	DO (mg L^{-1})	PI (mg L^{-1})	COD (mg L^{-1})	BOD$_5$ (mg L^{-1})	NH$_4$-N (mg L^{-1})	TP (mg L^{-1})	TN (mg L^{-1})	CHLA (μg L^{-1})	SD (m)
Site 1	195	Max	32.8	9.1	14.3	3.53	17.0	3.50	0.51	0.173	2.47	72	6.0
		Min	9.0	6.4	4.9	1.23	5.0	0.34	0.01	0.002	0.59	0.6	0.1
		Avg	20.5	7.7	8.7	1.96	7.0	1.05	0.08	0.029	1.24	9.5	2.39
Site 2	184	Max	34.3	8.8	14.9	2.67	14.0	2.70	0.18	0.050	1.63	47.0	7.8
		Min	9.3	6.7	6.7	0.78	1.1	0.28	0.01	0.002	0.33	0	0.8
		Avg	20.9	7.8	9.3	1.54	5.1	1.04	0.03	0.013	1.04	7.0	3.81
Site 3	230	Max	33.5	8.8	11.8	2.00	12.0	1.80	0.09	0.027	1.45	20.2	11.0
		Min	9.6	6.7	6.4	0.71	0.7	0.22	0.01	0.002	0.42	0.3	1.7
		Avg	20.7	7.8	8.9	1.37	4.1	0.85	0.02	0.009	0.84	4.0	5.48
Site 4	230	Max	32.9	8.5	11.6	2.13	12.0	1.60	0.03	0.025	1.48	15.6	11.0
		Min	6.7	6.6	6.0	0.61	0.0	0.19	0.01	0.002	0.40	0.0	2.4
		Avg	20.7	7.7	8.6	1.27	4.0	0.80	0.01	0.007	0.82	3.3	6.02

The second approach is an "abnormal" value detection method, as follows:

(1) Suspected abnormal value judgement. For a target value, not including the first and last ones (e.g., 'x_i' in Equations (1)–(3)), if it is either larger or smaller than its adjacent values, then the target value will be regarded as a suspected abnormal value and flagged.

$$f_1 = x_i - x_{i-1} \tag{1}$$

$$f_2 = x_{i+1} - x_i \tag{2}$$

$$ff = f_1 \times f_2 \tag{3}$$

where x_i (I = 2, 3, 4, … n − 1) represents the time series of buoy measurements, excluding the first and last values. So if ff < 0, the measurement at the ith time will be regarded as a suspected abnormal value.

(2) Abnormal value confirmation. We calculate the average value $\bar{x}$ of raw data after MMR control, the anomaly $|x_{max} - \bar{x}|$ between the maximum value x_{max} and $\bar{x}$, and the anomaly $|x_{min} - \bar{x}|$ between the minimum value x_{min} and $\bar{x}$. The larger value $|x - \bar{x}|$ between $|x_{max} - \bar{x}|$ and $|x_{min} - \bar{x}|$ will be chosen to compare with the absolute value $|ff|$ of ff. If $|ff|$ is larger than, or equal to, $|x - \bar{x}|^2$, then the measurement at the ith time will be confirmed as an abnormal value.

2.5. Change Point and Trend Detections

The Pettitt test was used to automatically detect change points in the data series once a week. Pettitt's test is a nonparametric test to detect a single change point in a time series with continuous data. Its calculation procedures can be found in detail in [24]. The identified change points were then compared to the minimum and maximum values for each sensor. If their values are all in MMR, the validity of change point values will be

confirmed. Otherwise the values will be removed from the time series or marked for further check. An exploratory analysis was also carried out to detect the trend of hourly and daily data for all the parameters using the Mann-Kendall method once a week. If the serial data kept increasing or decreasing for more than one week, its validity would be manually and carefully investigated.

2.6. Data Availability, Daily and Hourly Data Calculation

Most of the buoy monitoring datasets at the lake area (Buoy_surface, Buoy_profiler, Met_station) commenced in September 2015. The River_station data collection began in August 2016 and the Flux_station and Hydro_station data collection began in April 2017. A software developed by the authors is used to analyse and summarise the high-frequency data, including the calculation of daily and hourly values, based on quality-controlled raw data. Small data gaps without measurements (≤days) are interpolated by the software and the large data gaps are arbitrarily set up with a unique flag number (e.g., '8888'), which will be not included for calculating daily and hourly values.

3. Results

3.1. Buoy Photographs

Figure 2 shows photographs of the Buoy_surface system at site 5 (Figure 2A), located at the mouth of largest tributary (Xinan River, Figure 1), and the Buoy_profiler system at site 4 (Figure 2B), located at the deepest area in front of the dam for the power station (Figure 1), which is the only outflow. The web interface, which dynamically updates all station data from the database, allows the user to make requests for time periods of interest, review data from specific sites, visualise data as a function of time, and perform simple statistical analyses of the real-time data. All historical data from the monitoring system can be downloaded through the web interface by authorised users.

Figure 2. Pictures of (**A**) the Buoy_surface system at site 5, (**B**) the Buoy_profiler at site 4.

3.2. Surface Measurements

The measurements from site 5 are presented here as an example, and show the daily and hourly variations in surface WQ measured by the buoy probes. Figure 3A shows the time series of daily surface WT from 30 September 2015 to 1 August 2020, and daily surface DO, CHLA and PC from 27 January 2016 to 1 August 2020. The values of maximum (Max), minimum (min), average and standard deviation (Stdev) for WT, DO, CHLA and PC at site 5 are given in Table 3. All the maximum values for WT, CHLA and PC occurred in summer, but their lowest observations occurred in winter (PC) or spring (WT, DO and CHLA). CHLA and PC showed higher variability over time than those of WT and

DO during the study period, based on their statistical Stdevs compared to their average values.

Table 3. Statistical summaries of surface WT, DO, CHLA and PC measured by monitoring buoys at site 5 and site 6.

Site	Parameter	Max		Min		Avg.	Stdev	*n*
		Value	Date	Value	Date			
Site 5	WT (°C)	32.4	30 July 2017	10.3	1 March 2019	20.2	6.3	1617
	DO (mg L^{-1})	13.9	18 April 2020	1.7	3 April 2016	9.3	1.6	1546
	CHLA (μg L^{-1})	24.3	17 June 2019	0.1	9 April 2016	3.6	3.0	1530
	PC (μg L^{-1})	2.15	23 July 2017	0.1	10 February 2016	0.5	0.4	1318
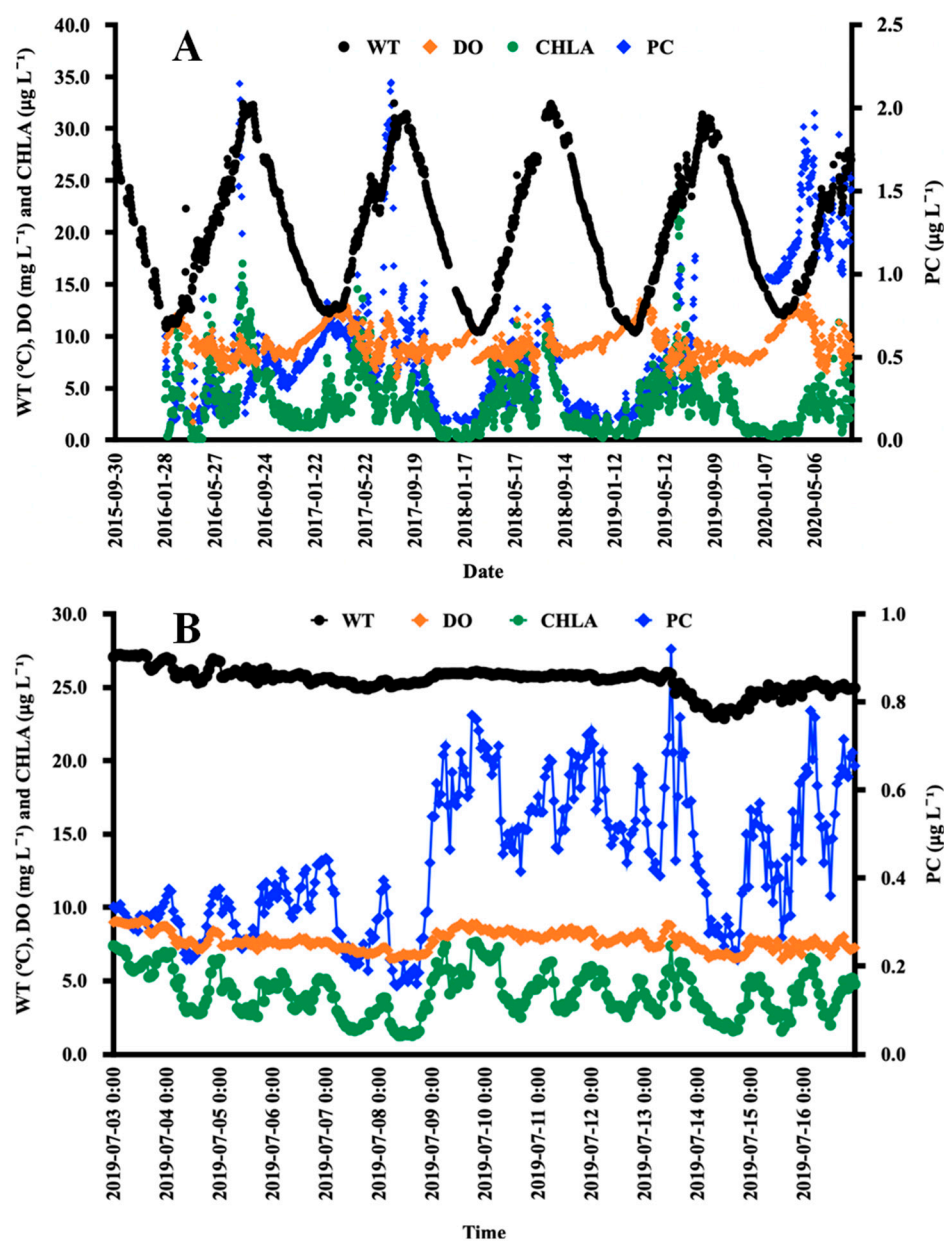
Figure 3. Time serials of (**A**) daily and (**B**) hourly WT (black dots, °C), DO (orange diamonds, mg L^{-1}), CHLA (green dots, μg L^{-1}) and PC (blue diamonds, μg L^{-1}) at site 5. Left Y-axes is for WT, DO and CHLA, and right Y-axes is for PC. X-axes is for date in format of year-month-day.

To show diel variation, hourly data of WT, DO, CHLA and PC at site 5 for the period of 00:00 a.m. 3 July 2019–11:00 p.m. 16 July 2019, without data gaps, are presented as an example in Figure 3B. CHLA and PC show obvious diel variation with higher values in daytime relative to night time, while WT and DO keep more constant than CHLA and PC, showing no diel variation. The Pearson correlation coefficient between WT and DO is 0.8 (n = 336), suggesting that surface DO was mainly controlled by WT for the lake.

3.3. Profiling Measurements

Profiles of WT, DO, CHLA and PC at site 4 (deepest area) for 1 January 2016–10 July 2020 are shown in Figure 4. The maximum measurement depth was 65 m at this site. WT (Figure 4A) profiles show a monomictic pattern of mixing with thermal stratification in summer and mixing in winter, although some periods and layers lacked measurements. The TDs were 9.2 m, 9.4 m, 12.1 m and 7.8 m in the summer (July–September) of 2016, 2017 and 2018 and July of 2019 (data not available in August and September). The TD in 2018 were greater than other years, which suggests that the stratification in the summer of 2018 was more intensive than in 2016, 2017 and 2019. Correspondingly, bottom hypoxia events were observed during the stratification of all years. Average bottom DO values (Figure 4B) during stratification were 7.0 mg L^{-1}, 7.7 mg L^{-1}, 8.2 mg L^{-1} and 7.7 mg L^{-1} in 2016, 2017, 2018 and 2019, respectively. The lowest DO values were 1.0 mg L^{-1}, 2.0 mg L^{-1}, 1.6 mg L^{-1} and 2.9 mg L^{-1}, observed on 4 April 2016, 9 December 2017, 1 January 2018 and 13 January 2019, respectively.
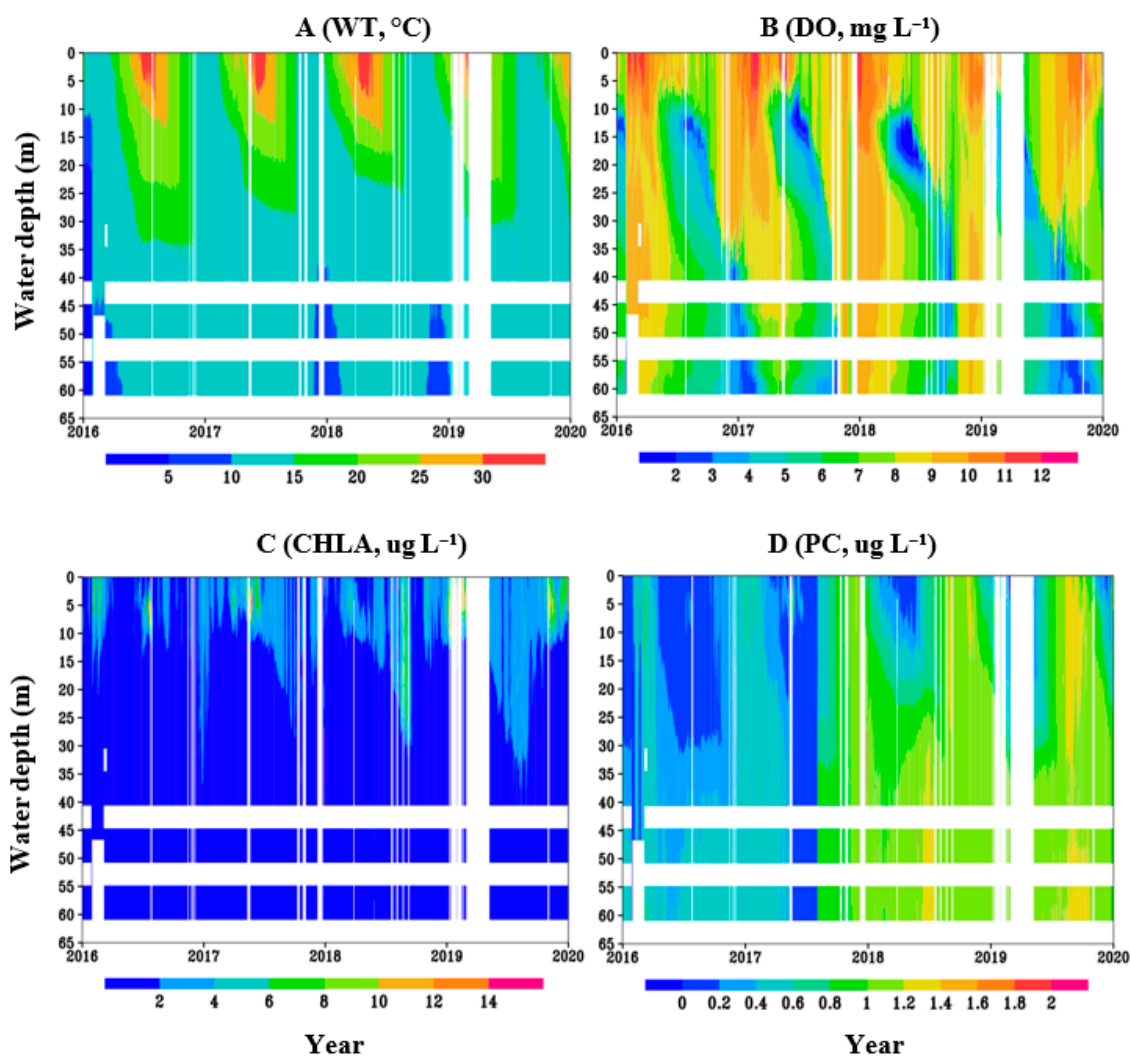
Figure 4. Daily (**A**)WT, (**B**) DO, (**C**) CHLA and (**D**) PC profiles at site 4 for 1 January 2016–10 July 2020. Y-axes for (**A–D**) represent water depth (m) and X-axes represent years from 2016 to 2020. White areas denote lack of measurements.

CHLA (Figure 4C) followed a similar pattern to WT, with higher values in summer than in winter, suggesting that the biomass of phytoplankton is mainly regulated by water temperature instead of nutrients. The phytoplankton was mostly distributed in the upper 15 m except for late 2018 and early 2019, when phytoplankton could still be found at a depth of 30 m. PC values (Figure 4D) were much smaller than CHLA at the same depth. It didn't have distinct seasonal variation, but was obviously stratified in the summer of 2018 with a higher concentration in the lower layer than the upper layer.

3.4. Real-Time Early Warning Information

The Ministry of Ecology and Environment of China (MEEC) issued state standards for surface water quality in 2002 [25] in order to better manage surface water in China. Lake Qiandao was required to meet Grade I (Table 4, the requirements for heavy metal were not shown in the table) since it provides drinking water for approximately half a million people in Chun'an County (located to the northeast of lake), and a total of 10 million people in both Hangzhou City and Jiaxing City, with water diverted through a tunnel of more than 110 km in length. DO and pH are the only two parameters which were measured by wireless sonde, deployed with monitoring buoys at the lake. The statistical analysis results for DO and pH are shown in Table 5. There were totally 275 and 130 samples of pH and DO out of their MMRs at site 5, accounting for 11.0% and 8.0% of all the valid samples, respectively. The percentages of pH (21.5%) and DO (11.3%) at site 6 were more than those at site 5. The observed maximum/minimum/average values of pH at site 5 and site 6 were 9.9/9.002/9.22 and 12.2/9.0002/9.62, respectively. Thus, the maximum pH at site 6 is much greater than that at site 5, showing that pH was more variable at site 6 than at site 5. However, DO followed a reverse pattern, with more varied values at site 5 compared to site 6. Its lowest value was as low as 1.74 mg L^{-1}, observed on 3 April 2016.

Table 4. Threshold of WQ parameters required for Lake Qiandao.

Parameter	pH	DO (mg L^{-1})	PI (mg L^{-1})	COD (mg L^{-1})	BOD5 (mg L^{-1})	NH₄-N (mg L^{-1})	TP (mg L^{-1})	TN (mg L^{-1})
Threshold for Grade I	6–9	≥7.5	≤2	≤15	≤3	≤0.15	≤0.01	≤0.2

Table 5. Statistical analysis results of DO and pH for early warning at site 5 and site 6.

Statistical Results	Site 5		Site 6	
	pH	DO	pH	DO
Number of all available data	1573	1618	1128	1128
Number of EWS data	275	130	242	128
Max	9.9	7.5	12.2	7.5
Min	9.002	1.74	9.0002	6.2
Avg.	9.22	6.92	9.62	6.99

4. Discussion

The EWS is now being used for giving real-time early warning signals by judging if the measurements of each sensor are out of the corresponding MMR and meet the required WQ grade. Unfortunately, DO and pH are the only two parameters which were directly measured by the buoy sensor and which can be used for the early warning of WQ at this lake. Therefore, it should be considered seriously whether sensors measuring

nutrients (e.g., TN, TP, NH_4-N, COD and PI) should be added to the buoys, or whether a model (e.g., AEM3D) simulating these nutrients based on high-frequency monitoring data, needs to be developed for providing more satisfactory early warning signals.

The system is also providing data for horizontal interpolation, to produce an initial condition for AEM3D predicting HABs and WQ at a time scale of seven days. For a WQ and HAB prediction system, initial and boundary conditions are very important to improve prediction accuracy. The initial conditions include horizontal and vertical WQ distribution within the waterbody (e.g., TN, TP and CHLA concentrations at each grid location). A huge challenge with initial conditions derives from the limited number of monitoring buoys collecting high-frequency WQ data, due to economic considerations, which leads to an inaccurate spatial distribution of WQ (e.g., high spatial patchiness of cyanobacterial bloom). However, many advanced interpolation methods are now available to address this issue. Inaccurate spatial WQ distribution in model initialisation can lead to inaccurate WQ prediction at each grid location, and thus unconvincing algal aggregation caused by winds. However, temporal and spatial difficulties prevent conventional methods for water sampling and laboratory analysis to meet the requirements for producing model initial conditions. A comprehensive approach is required, integrating high-frequency buoy monitoring data, laboratory data, satellite images and other available resources to provide a satisfactory spatial WQ concentration for the initialization of predictive simulation systems.

WQ sonde (multi-sensor probes) measurements can efficiently provide data wirelessly at a high temporal resolution, but potential problems could include data distortion due to sensor faults, or data gaps because of failed data storage and/or transfer. Therefore, data QC procedures are necessary before using the sensor measurements. The first step for QC is typically to detect missing series and estimate missing values by relying on neighbouring observations, then to detect unreasonable values out of range between the upper and the lower limits for each parameter, ideally guided by experience for a specific water body and measurement type. Unreasonable values can be removed and substituted with interpolated values. If many successive measurements from the same sensor are of the same exact value, they should usually also be removed and interpolated with neighbouring values, or excluded from data analysis. The final step is to detect outliers (incorrect or out-of-range) measurements, which can be removed or assumed to be missing [26]. Outliers are typically those observations which represent abrupt increases or decreases compared to the neighbouring values. There are many methods [26–28] and pieces of software [29] available for data QC. In this paper, we adopted the outlier detection method for finding anomalous values, which were removed and generally replaced with interpolated values. The detections of change points and trends will further help to find abnormal values or sonde problems. In our system, interpolation was not implemented to reproduce actual missing values due to the high potential for erroneous measurement generation when interpolating over longer time periods. Therefore, interpolation methods or software need to be integrated into this system in order to produce data without measurement.

The whole buoy monitoring system was originally designed to provide essential WQ information, in order to meet Grade I at the required sites. Therefore, Buoy_profilers were deployed at site 1, site 3 and site 4, and Buoy_surface was deployed at site 5. However, the buoys can only monitor WT, DO, pH, CHLA, PC and TURB. It is very difficult and expensive to directly and accurately measure TN, TP, NH_4-N, PI, COD and BOD_5 at a high frequency and in near-real-time [30]. An optional solution in this early warning system, is to calculate these parameters based on their regressed relations with sonde-measured values. The calculated values from regressive equations can then provide vital WQ information at different zones of the lake, feeding the AEM3D (http://www.hydronumerics.com.au, accessed on 15 January 2022) model for WQ prediction.

The collected data may assist water environmental managers in identifying and predicting the impacts of climatic extreme events [31]. For example, at Lake Qiandao, rainstorms with high rainfall typically result in a large inflow of water and nutrient loading, including N, P, and organic/inorganic matters, leading to an abrupt increase in water level and a significant increase of regional N and P concentrations [22]. This remarkably alters the spectral absorption properties of chromophoric dissolved organic matter (CDOM) and particles at the northwestern, southwestern and northeastern areas [32]. HFMS can also provide a useful basis for theory and model developments, improving our understanding of lake (reservoir) responses to perturbations caused by human activities and climate change at different time scales (e.g., sub-hourly, hourly, daily, monthly, seasonally, annually and every decade).

Although HFMS has a wide range of applications, it is now still hampered by several factors. For example, we have a limited choice of water quality sensors that are robust, economic and low-maintenance. The accuracy of the sondes measuring chemical parameters (i.e., phosphorus, ammonium, ammonia and nitrite) and biological parameters (i.e., bacterial enumeration, cyanobacteria, biota and cyanotoxins) still needs to improve, although the fast spread of HFMS is encouraging sensor developers to improve technology as quickly as they can. For giving better early warning signals and real-time WQ assessment, in the future, it will be necessary to add sondes measuring chemical and biological parameters to the current HFMS.

Author Contributions: Conceptualization, L.L. and H.L.; methodology, L.L.; software, L.L.; validation, J.L. (Jia Lan), Y.W., Z.W. and C.M.; formal analysis, (J.L.) Jialong Li; investigation, Y.W.; resources, Y.W.; data curation, F.L., L.C. and G.W.; writing—original draft preparation, L.L.; writing—review and editing, C.M. and H.Z.; visualization, L.L., R.Z. and F.G.; supervision, L.L.; project administration, L.L. and H.L.; funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Greeson, P.E. Lake eutrophication—A natural process. *J. Am. Water Resour. Assoc.* **1969**, *5*, 16–30.
2. Anderson, D.M.; Glibert, P.M.; Burkholder, J.M. Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries* **2002**, *25*, 704–726.
3. Smith, V.H.; Schindler, D.W. Eutrophication science: Where do we go from here? *Trends Ecol. Evol.* **2009**, *24*, 201–207.
4. Conley, D.J.; Paerl, H.W.; Howarth, R.W.; Boesch, D.F.; Seitzinger, S.P.; Havens, K.E.; Lancelot, C.; Likens, G.E. Controlling eutrophication: Nitrogen and phosphorus. *Sciences* **2009**, *323*, 1014–1015.
5. Lee, J.H.W.; Hodgkiss, I.J.; Wong, K.T.M.; Lam, I.H.Y. Real time observations of coastal algal blooms by an early warning system. *Estuar. Coast. Shelf Sci.* **2005**, *65*, 172–190.
6. Wang, X.; Su, J.; Cai, Y.; Tan, Y.; Yang, Z.; Wang, F. An integrated approach for early warning of water stress in shallow lakes: A case study in Lake Baiyangdian, North China. *Lake Reserv. Manag.* **2013**, *29*, 285–302.
7. Burchard-Levine, A.; Liu, S.; Vince, F.; Li, M.; Ostfeld, A. A hybrid evolutionary data driven model for river water quality early warning. *J. Environ. Manag.* **2014**, *143*, 8–16.
8. Park, Y.; Cho, K.H.; Park, J.; Cha, S.M.; Kim, J.H. Development of early-warning protocol for predicting chlorophyll-a concentration using machine warning models in freshwater and estuarine reservoirs, Korea. *Sci. Total Environ.* **2015**, *502*, 31–41.

9. Biddanda, B.A.; Weinke, A.D.; Kendall, S.T. et al., Chronicles of hypoxia: Time-series buoy observations reveal annually recurring seasonal basin-wide hypoxia in Muskegon Lake—A Great Lakes estuary. *J. Great Lakes Res.* **2018**, *44*, 219–229.

10. Delpla, I.; Florea, M.; Rodriguez, M.J. Drinking water source monitoring using early warning systems based on data mining techniques. *Water Resour. Manag.* **2019**, *33*, 129–140.

11. Glasgow, H.B.; Burkholder, J.M.; Reed, R.E.; Lewitus, A.J.; Kleinman, J.E. Real-time remote monitoring of water quality: A review of current applications, and advancements in sensor, telemetry, and computing technologies. *J. Exp. Mar. Biol. Ecol.* **2004**, *300*, 409–448.

12. Ungureanu, F.; Lupu, R.G.; Stan, A.; Craciun. I.; Teodosiu. C. Towards real time monitoring of water quality in river basins. *Environ. Eng. Manag. J.* **2010**, *9*, 1267–1274.

13. Storey, M.V.; van der Gaag, B.; Burns, B.P. Advances in on-ling drinking water quality monitoring and early warning systems. *Water Res.* **2011**, *45*, 741–747.

14. Qi, X.; Wang, S.; Jiang, Y.; Liu, P.; Li, Q.; Hao, W.; Han, J.; Zhou, Y.; Huang, X.; Liang, P. Artificial electrochemically active biofilm for improved sensing performance and quickly devising of water quality early warning biosensors. *Water Res.* **2021**, *198*, 117164. https://doi.org/10.1016/j.watres.2021.117164.

15. Zhang, C.; Pei, H.; Jia, Y.; Bi, Y.; Lei, G. Effects of air quality and vegetation on algal bloom early warning systems in large lakes in the middle–lower yangtze river basin. *Environ. Pollut.* **2021**, *285*, 117455. https://doi.org/10.1016/j.envpol.2021.117455.

16. Kumar, T.; Naik, S.; Jujjavarappu, S.E. A critical review on early-warning electrochemical system on microbial fuel cell-based biosensor for on-site water quality monitoring. *Chmesphere* **2022**, *291*, 133098. https://doi.org/10.1016/j.chemosphere.2021.133098.

17. Ungureanu, F.; Balasa, M.I. Sensor networks in waste water treatment monitoring. *Environ. Manag. J.* **2007**, *6*, 65–70.

18. Tang, M.; Si, W.; Zhong, S. Development of automatic water quality monitoring network at Lake Taihu, Jiangsu Province. *Environ. Dev.* **2017**, *2*, 91–95. (In Chinese)

19. Wu, J.; Kong, L.; Yang, J.; Li, Y. Design and development of the integrated platform for environmental monitoring and information management for Lake Dianchi. *Technol. Innov. Appl.* **2018**, *15*, 79–80. (In Chinese)

20. Zhang, Y.; Wu, Z.; Liu, M.; He, J.; Shi, K.; Wang, M.; Yu, Z. Thermal structure and response to long-term climatic changes in Lake Qiandaohu, a deep subtropical reservoir in China. *Limnol. Oceanogr.* **2014**, *59*, 1193–1202.

21. Durre, I.; Menne, M.J.; Vose, R. Strategies for evaluating quality assurance procedures. *J. Appl. Meteorol. Climatol.* **2008**, *47*, 1785–1791.

22. Zhang, Y.; Lan, J.; Li, H.; Liu, F.; Luo, L.; Wu, Z.; Yu, Z.; Liu, M. Estimation of external nutrient loadings from the main tributary (Xin'anjiang) into Lake Qiandao, 2006–2016. *J. Lake Sci.* **2019**, *31*, 1534–1546. (In Chinese)

23. Xiang, C.; Yan, L.; Han, Y.; Wu, Z.; Yang, W. Valuation of ecosystem services of the Thousand-Island Lake, Zhejiang, China. *Chin. J. Appl. Ecol.* **2019**, *30*, 3875–3884. (In Chinese)

24. Pettitt, A.N. A nonparametric approach to the change-point problem. *J. R. Stat. Society. Ser. C* **1979**, *28*, 126–135.

25. Ministry of Ecology and Environment of China (MEEC). *State Standards for Surface Water Quality (GB 3838–2002)*; China Environmental Press: Beijing, China. **2002**,

26. Gonzalez-Rouco, J.F.; Jimenez, J.L.; Quesada, V.; Valero, F. Quality control and homogeneity of precipitation data in the southwest of Europe. *J. Clim.* **2001**, *14*, 964–978.

27. Li, X.; Li, F. Quality control of online oceanic and meteorological data. *Mar. Forecast.* **1997**, *14*, 71–78. (In Chinese)

28. Gokturk, O.M.; Bozkurt, D.; Sen, O.L.; Karaca, M. Quality control and homogeneity of Turkish precipitation data. *Hydrol. Processes* **2008**, *22*, 3210–3218.

29. Araya-Lopez, J.L.; Kalyuzhnaya, A.V.; Kosukhin, S.S.; Ivanov, S.V. Data quality control for St. Petersburg flood warning system. *Procedia Comput. Sci.* **2016**, *80*, 2128–2140.

30. Banna, M.H.; Imran, S.; Francisque, A.; Najjaran, H.; Sadiq, R.; Rodriguez, M.; Hoorfar, M. Online drinking water quality monitoring: Review on available and emerging technologies. *Crit. Rev. Environ. Sci. Technol.* **2014**, *44*, 1370–1421.

31. Marcé, R.; George, G.; Buscarinu, P.; Deidda, M.; Dunalska, J.; de Eyto, E.; Flaim, G.; Grossart, H.-P.; Istvanovics, V.; Jennings, E.; et al. Automatic High Frequency Monitoring for Improved Lake and Reservoir Management. *Environ. Sci. Technol.* **2016**, *50*, 10780–10794.

32. Shi, L.; Mao, Z.; Liu, M.; Zhang, Y. Effects of rainstorm on the spectral absorption properties of chromophoric dissolved organic matter and particles in Lake Qiandao. *J. Lake Sci.* **2018**, *30*, 358–374. (In Chinese)