sensors

MDPI

Article

Water Quality Indicator Interval Prediction in Wastewater Treatment Process Based on the Improved BES-LSSVM Algorithm

Meng Zhou [1], Yinyue Zhang [1], Jing Wang [1], Yuntao Shi [1,*] and Vicenç Puig [2]

[1] School of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China; zhoumeng@ncut.edu.cn (M.Z.); zyy17718301596@163.com (Y.Z.); jwang@ncut.edu.cn (J.W.)
[2] Advanced Control Systems Research Group at Institutde Robòtical, CSIC-UPC, Universitat Politècnica de Catalunya-BarcelonaTech, 08028 Barcelona, Spain; vicenc.puig@upc.edu
* Correspondence: shiyuntao@ncut.edu.cn

Abstract: This paper proposes a novel interval prediction method for effluent water quality indicators (including biochemical oxygen demand (BOD) and ammonia nitrogen (NH3-N)), which are key performance indices in the water quality monitoring and control of a wastewater treatment plant. Firstly, the effluent data regarding BOD/NH3-N and their necessary auxiliary variables are collected. After some basic data pre-processing techniques, the key indicators with high correlation degrees of BOD and NH3-N are analyzed and selected based on a gray correlation analysis algorithm. Next, an improved IBES-LSSVM algorithm is designed to predict the BOD/NH3-N effluent data of a wastewater treatment plant. This algorithm relies on an improved bald eagle search (IBES) optimization algorithm that is used to find the optimal parameters of least squares support vector machine (LSSVM). Then, an interval estimation method is used to analyze the uncertainty of the optimized LSSVM model. Finally, the experimental results demonstrate that the proposed approach can obtain high prediction accuracy, with reduced computational time and an easy calculation process, in predicting effluent water quality parameters compared with other existing algorithms.

Keywords: water quality monitoring; data pre-processing; improved IBES-LSSVM algorithm; interval prediction method

Citation: Zhou, M.; Zhang, Y.; Wang, J.; Shi, Y.; Puig V. Water Quality Indicator Interval Prediction in Wastewater Treatment Process Based on the Improved BES-LSSVM Algorithm. *Sensors* **2022**, *22*, 422. https://doi.org/10.3390/s22020422

Academic Editor: Assefa M. Melesse

Received: 30 October 2021
Accepted: 23 December 2021
Published: 6 January 2022

1. Introduction

Nowadays, freshwater is considered one of the most critical resources for humans, since it can ensure the availability of an acceptable quantity of water for livelihoods, health, ecosystems and production. Hence, freshwater plays a key role in poverty and disease burden reduction, economic growth and environmental sustainability [1,2]. This fact has long been acknowledged all over the world. However, due to industrial pollution, rapid population growth and farmland sewage caused by the extensive use of chemical fertilizers, pesticides and herbicides, the shortage of freshwater sources is a serious and challenging issue [3,4].

Wastewater treatment is one key technology to potentially provide additional water supplies, and it is very important for the functioning of the economy and society. Wastewater treatment has been attracting a lot of attention, since it can not only remove organic wastes to reduce the environmental burden, but also offer the advantage of producing a renewable source of water [5,6]. Wastewater treatment is a very complex process with a variety of physical and biochemical reactions since it presents nonlinear dynamic behavior, time delay and uncertainty [7]. In wastewater treatment plant processes, effluent water quality monitoring is an important task that involves measuring the evolution of the quality parameters in time.

Note that most traditional methods of measuring these quality indicators for wastewater treatment processes are based on manual lab-based monitoring approaches, with manual sample collection, long-time transportation and biological/microbial testing in a

laboratory, which is cumbersome and time-consuming. Usually, the testing equipment is very expensive and cannot be used online. In addition, since the process of wastewater treatment is complex, some control strategies are necessary and required to be deployed to guarantee that effluent quality indicators behave normally. In recent decades, water quality monitoring has been evolving to the latest wireless sensor networks [8], such that most of the important indicators of effluent water (pressure, pH, level and so on) can be measured by their corresponding sensors online. However, there are still some parameters that cannot be measured quickly due to high costs and the limitations of sensors, such as BOD and NH3-N. Usually, the concentration of the BOD/NH3-N effluent associated with a wastewater treatment process is an important factor to measure the water quality since the discharge of a large amount of NH3-N and BOD wastewater will lead to water eutrophication, which can affect human health. In China's "Pollutant Discharge Standard for Urban Wastewater Treatment Plants (*GB*18918-2002)", the Class A standard stipulates that the maximum discharge for NH3-N is 5 mg/L, while for BOD, it is 10 mg/L. Thus, measuring these effluent quality indicators with high accuracy is an important issue.

Researchers have focused on soft-sensing methods to predict these effluent quality indicators and the prediction task is addressed combining data analytics and water quality control. Soft-sensing methods aim to find some certain relationships between easy-to-measure variables and difficult-to-measure variables in the sewage treatment process. Then, a suitable model is established based on these relationships, and difficult-to-measure variables can be predicted based on the soft-sensing models.

Machine learning approaches are usually considered a subset of artificial intelligence. They focus on some statistical models and algorithms to extract patterns from data so that useful inferences can be used to predict new data. Recently, with the development of machine learning, artificial neural network (ANN), support vector machine (SVM), decision tree, random forest, ensemble learning and many other methods have been researched in depth and have a wide range of applications, including text processing, computer vision, healthcare, finance and robotics. They can also be used for socio-economic and environmental studies [9–12]. In [12], the impacts of flood protection in Bangladesh were evaluated by machine learning methods. In [13], a gray model and ANN method were investigated to predict suspended matter and chemical oxygen demand in the wastewater treatment process. Cong et al. proposed a mixed soft sensor model based on a wavelet neural network and adaptive weighted fusion for the online prediction of effluent COD [14]. M. Hamada carried out the assessment of a wastewater treatment plant's performance based on ANN and a multiple linear regression method [15]. M. Zeinolabedini et al. proved that applying various parent wavelet functions to the neural network structure can improve the accuracy of predicting the wastewater sludge volume [16]. A. K. Kadam et al. used ANN and multiple linear regression to model and predict water quality parameters in river basins [17]. S. Heddam et al. investigated a generalized regression neural network model to predict the BOD of effluent in wastewater treatment plants [18]. Tan et al. predicted the first weighting from the working face roof in a coal mine based on a GA-BP neural network [19]. V. Nourani et al. proved that the prediction ability of a neural network ensemble is more reliable [20].

Compared with the ANN method, SVM is another important prediction technique, which can effectively solve the problem of high-dimensional data model construction under the condition of limited samples, and has strong generalization ability. Hence, many scholars have carried out a lot of research on SVM-based prediction. Cheng et al. proposed a variety of kernel single-class SVMs to monitor and predict the intake conditions of wastewater treatment plants [21]. Han et al. developed a neural network model for predicting the sludge volume index based on information transfer strength and adaptive second-order algorithms [22]. Wu et al. proposed an adaptive multi-output soft sensor model for monitoring wastewater treatment and made several simulation comparisons to prove the superiority of the algorithm [23]. K. Lotfi et al. used a linear–nonlinear hybrid method to predict the effluent index of a wastewater treatment plant, which improves the

prediction ability of the single method [24]. Han et al. proposed a data-based predictive control strategy and proved its superiority through several simulations [25]. In [26], the total solid content of a wastewater treatment plant was predicted by an SVM model, which can enhance performance and durability.

Although SVM is a small-sample learning method and has been widely used to solve the wastewater prediction problem, the calculation process is multifarious, which is difficult to implement for large-scale training samples [27]. To overcome these disadvantages, the least-squares support vector machine (LSSVM) has been proposed. LSSVM improves the performance of the SVM algorithm by solving linear programming rather than quadratic programming. In this way, the calculation process can be reduced and the computation speed greatly improved [28]. Zhang et al. proposed an improved LSSVM model based on SVM to predict river flow [29]. Fei Luo et al. integrated the Gustafson-Kessel algorithm and least-squares support vector machine for line prediction of [30]. D. S. Manu et al. combined SVM and an adaptive neuro-fuzzy reasoning system model to predict the effluent nitrogen content of wastewater treatment plants [31]. Liu et al. investigated the online prediction of effluent COD in an anaerobic wastewater treatment system based on principal component analysis and the LSSVM algorithm [32].

Note that there are some unknown parameters in the kernel functions of LSSVM that need to be selected in advance. Generally, these parameters are determined according to experience, which may be time-consuming, and it is difficult to find the optimal parameters. Nowadays, swarm intelligence optimization algorithms are researched extensively, since the optimal solution can be found by swarm intelligence to perform a collaborative search mechanism. The results of the combination of swarm intelligence optimization algorithms and machine learning methods can be found in a large number of references. In [33], a hybrid model of particle swarm optimization (PSO) and support vector machine is proposed to predict the turbidity and pH value of sand filtered water in irrigation systems. Han et al. use an adaptive PSO algorithm to design self-organizing radial basis function neural networks to improve the accuracy and save time [34]. Chen et al. study the artificial bee colony optimization back-propagation network to predict the water quality of a water diversion project [35]. Fan et al. use the LSSVM model to improve the performance of predicting the safety factor of a circular slope [36]. Mahdi Shariati et al. use the gray wolf algorithm to optimize ELM model parameters to predict the compressive strength of partially replaced cement concrete [37]. However, to the best of the authors' knowledge, these swarm intelligence methods may fall into local optima and do not find the global optimal solutions.

Most of the above-mentioned methods only focus on point prediction, without providing information regarding accuracy. The prediction results have strong uncertainty that affects the decision-making process, increasing the risk of not making good decisions. Prediction interval (PI) is a standard tool for quantifying prediction uncertainty. PI not only provides the range where the target value is most likely to exist, but also indicates its accuracy. Yao et al. combined the mean variance estimation (MVE) method with a recurrent neural network to measure the uncertainty in prediction [38]. Yuan et al. combined beta distribution with the PSO-LSTM model to obtain the wind power prediction interval with high reliability and a narrow interval width, so as to provide decision support for the safe and stable operation of power systems [39]. Liao et al. combined the bootstrap method with the long and short memory network to realize the uncertain prediction of the remaining service life of the machine [40]. Marin et al. obtained the prediction interval of power consumption by combining the delta method with a fuzzy prediction model [41]. Sun et al. constructed a high-quality prediction interval based on the two-step method of dual ELM and applied it to the scheduling of a gas system [42]. In recent years, a direct interval prediction method called upper and lower bound estimation (LUBE) has been proposed. The main idea of this method is to directly construct the upper and lower bounds of PI by optimizing the coefficients of the neural network according to the interval quality evaluation index. This approach can provide good performance and does not consider

strict data distribution assumptions, such that it can provide more information about the prediction results, which motivates the work of this paper.

The main objective of this paper is to obtain a soft-sensor-based interval prediction method with high prediction accuracy and less computational time to predict the effluent water quality parameters, which is significant for water quality monitoring and control. Aiming at the online prediction of BOD/NH3-N effluent in a wastewater treatment plant within a smart data-driven framework, the main contributions of this paper are the following:

- Data pre-processing methods, i.e., abnormal data elimination and normalization, are taken into consideration after the data and their related auxiliary variables are collected. Then, some key factors of the wasterwater quality indicators are selected based on the gray correlation analysis algorithm.
- In order to improve the prediction accuracy of BOD/NH3-N effluent, a novel IBES-LSSVM algorithm is proposed, in which an improved bald eagle search (IBES) optimization algorithm is used to find the optimal parameters of the least-squares support vector machine (LSSVM). The superiority of the proposed method is verified by comparing it with the existing soft-sensing models (such as GWO, WOA, PSO, SSA) using some benchmark functions and providing higher prediction accuracy.
- In order to estimate the uncertainty of the model prediction results and make better decisions, after obtaining the point prediction results, the interval prediction bounds of effluent quality are also generated. Compared with some existing soft-sensing models, the proposed interval prediction method can obtain a more accurate prediction range.

The structure of this paper is as follows: In Section 2, the problem description is given, including the real data collection, data pre-processing and gray-correlation-analysis-based data selection. Section 3 describes the model uncertainty analysis by using the proposed IBES-LSSVM algorithm and LUBE algorithm. In Section 4, the simulation examples are depicted, demonstrating the effectiveness of the proposed method based on the BOD and NH3-N data. Section 5 draws the main conclusions of this paper.

2. Problem Description

In this paper, a soft-sensing-based method is investigated to analyze and predict the water quality indicators, including three main aspects: data collection, data pre-processing and data interval prediction. The main steps of the approach presented in this paper are shown in Figure 1.

Under a smart data-driven framework, in order to predict water quality tendencies and analyze the mechanisms behind the considered data sources, enough relevant experimental data in real time must be collected based on the prediction quality indicators. Most collected data may present several issues, such as data sparsity and data synchronization, among others. After the data are collected, they must be pre-processed in advance by applying several procedures, such as data cleaning, abnormal data elimination or normalization. Then, correlation analysis from different dimensions of water quality indicators should be considered to extract the relations between these auxiliary variables and find the key factors.

2.1. Data Collection

Due to the complexity of the wastewater treatment process and the large number of parameters that need to be set, it is necessary to determine the characteristic variables related to the water quality to be determined as auxiliary variables. The data that can evaluate the quality or impact of water quality in wastewater treatment plants are mainly divided into the following four categories [43]:

- **Physical data:** Physical properties are the ones that must be monitored throughout the treatment process, including total suspended solids, temperature, conductivity, transparency, total dissolved solids, etc.

- **Chemical data:** Chemical water quality indices of the national comprehensive discharge standard for water pollutants, including: pH, biochemical oxygen demand, biochemical oxygen consumption, heavy metals, nitrates, etc.
- **Biological data:** Biomarkers include a variety of microorganisms in the water, such as mayflies, E. coli, etc.
- **Environmental data:** Environmental data cover the whole process of water supply, including indexes of weather, hydrology, soil or ecology.

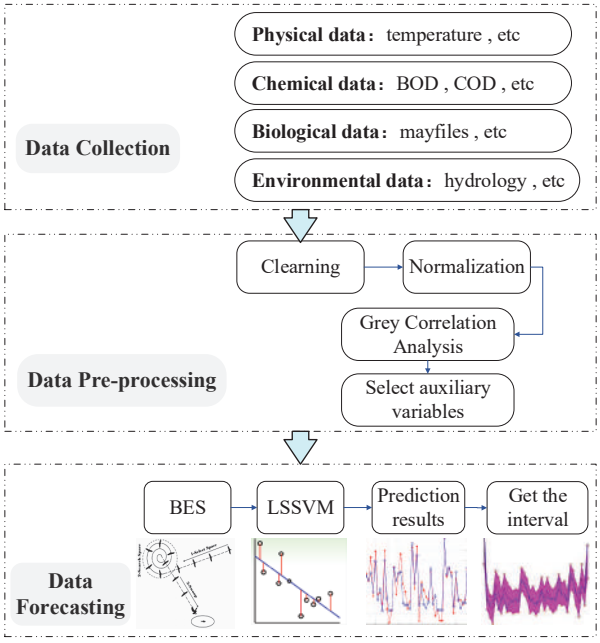


Figure 1. Main steps of the proposed approach.

This paper focuses on a real wastewater treatment plant in Beijing, China, from August 2014 to September 2014 [7,44]. Two data sets are collected first, which are used to predict the BOD/NH3-N effluent, separately. (1) BOD data set: containing 360 batches of data with 23 variables (including the BOD effluent parameters)—the detailed information is shown in Table 1; (2) NH3-N data set: including 10 characteristic variables related to NH3-N effluent parameters, as shown in Table 2.

2.2. Elimination of Abnormal Data

Data collected from wastewater treatment plants can contain erroneous values because of improper instrument operation, human or environmental interference and other factors. As a result, we need to analyze the collected data first, and eliminate some abnormal or meaningless data.

In this paper, we use the 3σ criterion to handle the abnormal data of the two collected data sets. The sample data are denoted as $x_1, x_2, \cdots, x_n$. η_i is used to represent the data residual error. Then, the standard deviation is calculated as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} {\eta_i}^2}{n-1}} \tag{1}$$

$$\eta_i = x_i - \bar{x} \tag{2}$$

where n represents the number of elements in the data set, and $\bar{x}$ is the data average.

If the residual error of particular data sample x_i satisfies

$$|\eta_i| > 3\sigma \tag{3}$$

this means that it corresponds to an abnormal sample and needs to be eliminated. Otherwise, x_i is accepted.

Table 1. Effluent BOD data set.

Number	Auxiliary Variable	
01	Influent pH (IPH)	
02	Effluent pH (EPH)	
03	Influent SS	(mg/L)
04	Effluent SS (ESS)	(mg/L)
05	Influent BOD (IBOD)	(mg/L)
06	Influent COD (ICOD)	(mg/L)
07	Effluent COD (ECOD)	(mg/L)
08	Sludge settling ratio of biochemical tank	(mg/L)
09	MLSS in biochemical tank (MLSS)	(mg/L)
10	Biochemical pool Do	(mg/L)
11	Influent oil (IOil)	(mg/L)
12	Effluent oil (EOil)	(mg/L)
13	Influent NH3-N (INH3-N)	(mg/L)
14	Effluent NH3-N	(mg/L)
15	Influent Chroma (IC)	(d)
16	Effluent Chroma (EC)	(d)
17	Influent TN (IT)	(mg/L)
18	Effluent TN	(mg/L)
19	Influent phosphate concentration (IPC)	(mg/L)
20	Effluent phosphate concentration	(mg/L)
21	Inlet water temperature	(°C)
22	Outlet water temperature	(°C)
23	Effluent BOD (EBOD)	(mg/L)

Table 2. Effluent NH3-N data set.

Number	Auxiliary Variable	
01	Effluent TP	(mg/L)
02	Influent TP (ITP)	(mg/L)
03	Temperature (T)	(°C)
04	Anaerobic terminal ORP (ATORP)	(mv)
05	Aerobic front end DO	(mg/L)
06	Aerobic terminal DO	(mg/L)
07	Total suspended solids TTS (TTS)	(mg/L)
08	Effluent PH (EPH)	
09	Effluent ORP (EORP)	(mL)
10	Effluent nitrate (EN)	(mg/L)
11	Effluent NH3-N (ENH3-N)	(mg/L)

2.3. Data Normalization

Different variables often have different dimensions and dimensional units. In order to eliminate the dimensional influence between indicators, it is necessary to normalize the data to achieve uniformity among the different data indicators. There are four classes of normalization methods, i.e., rescaling, mean normalization, standardization and scaling to

unit length. In this paper, the rescaling method is selected. The normalization formula is as follows:

$$\tilde{x_i} = \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}} \tag{4}$$

where x_i is any value of a variable; $x_{i\min}$ and $x_{i\max}$ are, respectively, the minimum and maximum value of the variable.

After this kind of normalization, all the values of the data are set in the range of $[0, 1]$.

2.4. Correlation Degree Analysis

Since different characteristic variables will have different influences on the predicted variables, to obtain a soft-sensing model with a simpler structure, it is necessary to choose the quality indicators with high correlations. Selecting $\acute{m}$ auxiliary variables from m variable, it has $\acute{m} < m$. In practice, the larger m is, the smaller $\acute{m}$ is compared to m.

In this paper, the gray relational degree analysis method is investigated to select the characteristic variables of BOD and NH3-N effluents. Gray relational degree analysis is a multi-factor statistical method, which describes the strength of the relationship between various factors according to the gray relational degree. This method looks for the inconsistency between quantitative results and quantitative analysis in the traditional mathematical statistics method and reduces the amount of calculation.

The gray correlation coefficient is formulated as follows:

$$\beta = \left| x_0(k) - x_j(k) \right| \tag{5}$$

$$\mu_{\mathrm{j}}(k) = \frac{\min\limits_{j}\min\limits_{k}\beta + \rho \cdot \max\limits_{j}\max\limits_{k}\beta}{\beta + \rho \cdot \max\limits_{j}\max\limits_{k}\beta} \tag{6}$$

where j means the j-th variable, k is the k-th iteration, $x_0(k)$ is the output variable, $x_j(k)$ is the input variable, μ_{j} is the gray correlation coefficient and ρ is the resolution coefficient. If ρ is smaller, the difference between correlation coefficients is larger, and the distinguishing ability is stronger.

Then, the gray correlation degree can be calculated as follows:

$$\gamma_j = \frac{1}{n}\sum_{k=1}^{n}\mu_j(k) \tag{7}$$

where n is the number of variables.

If the gray correlation degree is larger, this means that the corresponding variable has a higher correlation with the effluent quality indicators. Then, according to the gray correlation degree, the characteristic variables are sorted from front to back. Usually, a threshold is determined in advance as $\hbar$, and then the key indicators can be selected as the input of the soft-sensing model if

$$\gamma_j > \hbar \tag{8}$$

is satisfied.

3. Methodology

In this section, a novel IBES-LSSVM method is proposed to find the optimal kernel function parameters of the LSSVM in Figure 2.

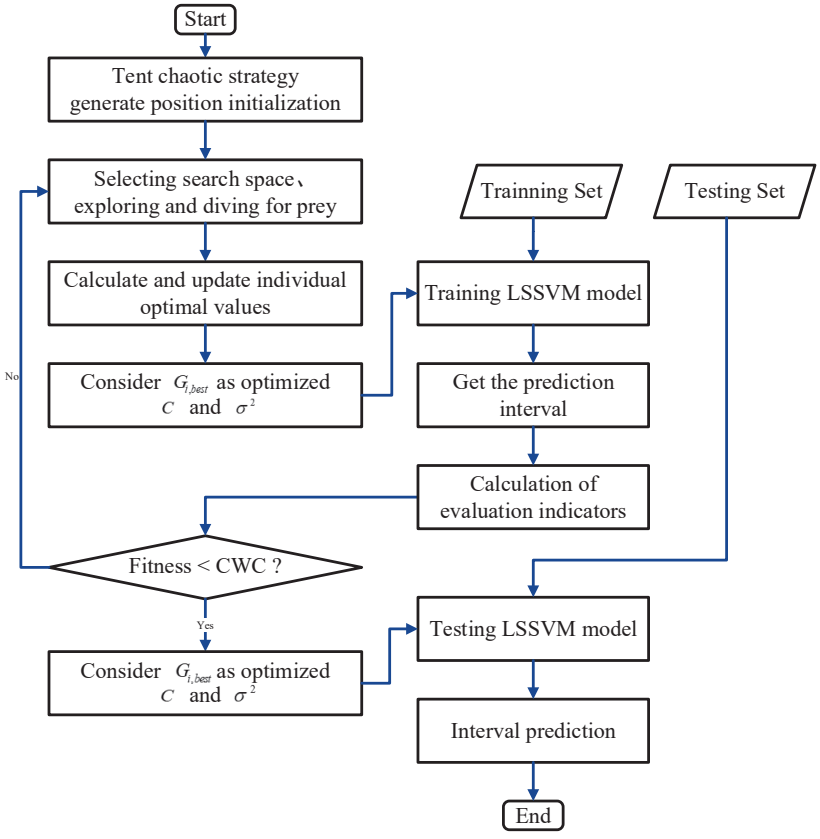


Figure 2. Flow chart of IBES-LSSVM model.

3.1. LSSVM Algorithm

The theory of LSSVM was first proposed by Suykens in 1994. LSSVM is a kernel learning machine following the principle of structural risk minimization and is suitable for analyzing the issue of sample classification and regression estimation [45].

In LSSVM theory, firstly, the sample data are mapped to higher dimensions through nonlinear changes, and linear functions are used for fitting in this high-dimensional feature space:

$$y(x) = w \cdot \phi(x) + b \tag{9}$$

where $y(x)$ is the output variable, x is the input variables, and w and b are weight and bias terms, respectively.

The optimization objectives of the LSSVM regression algorithm can be formulated as

$$\begin{aligned} &\min J(w, \xi_i) = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{n} \xi_i^2 \\ &s.t. \\ &\quad y_i = w \cdot \phi(x) + b + \xi_i, \quad i = 1, 2, \cdots, n \end{aligned} \tag{10}$$

where C is the regularization coefficient, ξ_i is the relaxation variable, and $\sum_{i=1}^{n} \xi_i^2$ is the experience risk.

By means of Lagrange multipliers α_i, (10) can be expressed as:

$$L(w, b, \xi_i, \alpha_i) = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{N} \xi_i^2 - \sum_{i=1}^{n} \alpha_i [w \cdot \phi(x) + b + \xi_i - y_i] \tag{11}$$

According to Karush–Kuhn–Tucker (KKT) optimization conditions:

$$\begin{cases} \dfrac{\partial L}{\partial \mathrm{b}} = 0 \Rightarrow \sum\limits_{i=1}^{n} \alpha_i = 0 \\ \dfrac{\partial L}{\partial w} = 0 \Rightarrow w = \sum\limits_{i=1}^{n} \alpha_i \phi(x_i) \\ \dfrac{\partial L}{\partial \xi_i} = 0 \Rightarrow \alpha_i = C\xi_i \\ \dfrac{\partial L}{\partial a} = 0 \Rightarrow w \cdot \phi(x_i) + b + \xi_i - y_i \end{cases} \tag{12}$$

By defining kernel functions, the optimization problem (11) can be transformed into a linear solution issue:

$$\begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & \frac{K(x_1,x_1)+1}{C} & \cdots & K(x_1, x_n) \\ \vdots & \vdots & & \vdots \\ 1 & K(x_n, x_1) & \cdots & \frac{K(x_n,x_n)+1}{C} \end{pmatrix} \begin{pmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} 0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{13}$$

where $K(x, x_i)$ is the kernel function.

The Lagrange multiplier and its parameters can be obtained from (13). Therefore, the output of LSSVM can be obtained:

$$\hat{y}(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i) + b \tag{14}$$

For LSSVM, there are many different types of kernel functions, such as linear function, polynomial kernel function, radial basis function (RBF), sigmoid kernel function, etc. Different kernel functions will produce difference types of LSSVM. In this paper, we select RBF as the kernel function of the model:

$$K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2}) \tag{15}$$

where σ is the variance of RBF.

Through the aforementioned analysis, LSSVM has two tunable parameters (regularization coefficient C and variance of radial basis kernel function σ with RBF), which are important and need to be determined. To obtain the optimal two parameters, the next step is to use an improved PSO algorithm to optimize them.

3.2. IBES-LSSVM Algorithm

The BES algorithm is an optimization algorithm that simulates the hunting strategy of vultures when looking for fish. It can obtain a single optimal solution through multiple iterations and finally obtain the overall optimal solution, such that the position of the optimal solution corresponds to the optimal parameter value.

BES hunting is divided into three stages. In the first stage (selection space), the eagle selects the space with the largest prey number. In the second stage (spatial search), the eagle moves in the selected space to find the prey. In the third stage (dive), the eagle swings from the best position determined in the second stage and determines the best hunting.

In the selection stage, firstly, this paper optimizes the initial prey position and adopts the tent chaos strategy, which has the advantages of simple structure and strong ergodicity. Then, the linear decreasing method is used to improve the control parameters of the vulture iterative update position. The optimal model parameters of the model can be found that improve the quality of the fitting. The tent chaotic mapping function is described as:

$$P_{i+1} = \begin{cases} P_i/\lambda, & P_i \in [0, \lambda) \\ (1 - P_i)/(1 - \lambda), & P_i \in [\lambda, 1] \end{cases} \tag{16}$$

where λ is $[0, 1]$.

Then, the vultures hunt for food. The formula is:

$$P_{\text{new},i} = P_{\text{best}} + R_1 \cdot C_1 \cdot (P_{mean} - P_i) \tag{17}$$

where R_1 is a parameter controlling the position change, and C_1 is a random number between $(0, 1)$. P_{best} is the current optimal location. P_{mean} is the average distribution location of vultures after the previous search. P_i is the location of the i-th vulture.

In the search phase, vultures search for prey in the selected search space and move in different directions in the spiral space to speed up the search. The best position for subduction is:

$$P_{i,\text{new}} = P_i + b(i) \cdot (P_i - P_{i+1}) + a(i) \cdot (P_i - P_{mean}) \tag{18}$$

where:

$$a(i) = \frac{ar(i)}{\max(|ar|)} \tag{19}$$

$$b(i) = \frac{br(i)}{\max(|br|)} \tag{20}$$

$$ar(i) = r(i) \cdot \sin[(\theta(i))] \tag{21}$$

$$\text{br}(i) = r(i) \cdot \cos[(\theta(i))] \tag{22}$$

$$r(i) = \theta(i) + R_2 \cdot C_3 \tag{23}$$

$$\theta(i) = \pi \cdot \omega \cdot C_2 \tag{24}$$

$$\omega = (1 - \frac{i}{i_{max}})^2 \cdot (\omega_{max} - \omega_{min}) + \omega_{min} \tag{25}$$

where $\theta(i)$ and $r(i)$ are the polar angle and polar diameter of the spiral equation, respectively. ω and R_2 are the parameters controlling the spiral trajectory. C_2 and C_3 are a random number within $(0, 1)$. The $a(i)$ and $b(i)$ represent the position of the vulture in polar coordinates, and the values are $(-1, 1)$.

During the dive phase, vultures swing from the best position in the search space to their target prey. All points also move towards the best point according to

$$\begin{aligned} P_{i,\text{new}} = {} & C_4 \cdot P_{best} + a_1(i) \cdot (P_i - R_3 \cdot P_{man}) \\ & + b_1(i) \cdot (P_i - R_4 \cdot P_{best}) \end{aligned} \tag{26}$$

where:

$$a_1(i) = \frac{ar(i)}{\max(|ar|)} \tag{27}$$

$$b_1(i) = \frac{br(i)}{\max(|br|)} \tag{28}$$

$$ar(i) = r(i) \cdot \sinh[(\theta(i))] \tag{29}$$

$$\text{br}(i) = r(i) \cdot \cosh[(\theta(i))] \tag{30}$$

$$r(i) = \theta(i) \tag{31}$$

$$\theta(i) = \pi \cdot \omega \cdot C_5 \tag{32}$$

where R_3 and R_4 represent the moving speed of the vulture to the optimal point. C_4 and C_5 are random numbers within $(0, 1)$.

3.3. Interval Prediction

The traditional point prediction cannot deal with the uncertainty in the operation of the system. In order to obtain the numerical estimation and its reliability, the practical application requires the calculation of the prediction interval. Interval prediction indicates the estimation interval of the range of predicted values in a certain confidence interval. Therefore, the prediction interval is composed of the upper and lower line of prediction, which provides its accuracy within a certain confidence level. Assuming that the confidence level is $(1-\mu)\%$, l and u are the lower and upper limits, respectively, when $P(l < y < u) = 1 - \mu\%$, and PI can be expressed as $[l, u]$. For a given confidence interval, the smaller the range of prediction interval, the smaller the uncertainty of prediction and the higher the accuracy.

The evaluation indexes of interval prediction are as follows [46].

$PICP$: The ratio of the real value to the upper and lower bounds of the prediction interval

$$PICP = \frac{1}{n}\sum_{i=1}^{n} c_i \tag{33}$$

If the predicted value is within the $[l_i, u_i]$ range, c_i is 1. Otherwise, c_i is 0. If all predicted values are included in the prediction interval, $PICP$ = 100%. n is the number of prediction points. In theory, $PICP \geqslant (1-\mu)\%$; otherwise, PI is invalid or unreliable. When comparing the PIs by the model, the other indexes should be as small as possible under the condition that the $PICP$ is as close to the confidence level as possible.

$PINAW$: The narrow PI has more information and practical value than the wide PI according to

$$PINAW = \frac{1}{nR}\sum_{i=1}^{n}(u_i - l_i) \tag{34}$$

where R is the range of predicted values, respectively.

$PINRW$: Represents the standard square root width of the predicted interval. The expression is:

$$PINRW = \frac{1}{R}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(u_i - l_i)^2} \tag{35}$$

CWC: In practical application, it is often hoped that a narrow prediction interval width can still be obtained under the condition of high prediction probability, i.e., the prediction interval range probability and interval width will conflict. Therefore, the comprehensive index CWC is proposed:

$$CWC = PINAW\left(1 + \varrho(PICP) \cdot e^{-\tau\cdot(PICP-(1-\mu))}\right) \tag{36}$$

where τ and μ are constants.

When working with training data, the set $\varrho(PICP)$ is 1. In addition, in data verification, $\varrho(PICP)$ is a step function:

$$\varrho = \begin{cases} 0 & PICP \geq 1-\mu \\ 1 & PICP < 1-\mu \end{cases} \tag{37}$$

LUBE is a method based on neural networks to directly calculate the lower and upper bound of the prediction interval. Assuming that the two node values of the output layer of the neural network are the upper and lower limits of the interval, respectively, all the predicted values are included in this range at the confidence level $(1-\mu)\%$. The training purpose of a neural network is to minimize the objective function CWC. In this way, the probability and

width of the prediction interval are considered at the same time, and the advantages and disadvantages of the prediction interval PI can be comprehensively evaluated.

The flow-chart of the proposed IBES-LSSVM algorithm is shown in Figure 2, which mainly includes the procedure presented in Algorithm 1.

Algorithm 1 LUBE interval prediction based on IBES-LSSVM model.

Input: Measured data of wasterwater treatment plant.
Output: Prediction interval of BOD/NH3-N effluent.
Step 1: Abnormal data elimination, normalization of the data according to Equations (1)–(4).
Step 2: Analyzing and selecting the key indicators with high correlation degree by Equations (5)–(8).
Step 3: The bald eagle population is initialized by tent chaos strategy based on Equation (16).
Step 4: Local optimal solution.
1: for all X_i do:
2: for all X_i do:
3: Obtain predicted value by means of Equations (9)–(15), (17).
4: end for
5: Using confidence, mean, standard deviation and other parameters, the prediction interval is obtained according to $norminv()$ formula.
6: Evaluate interval fitness by means of Equations (33)–(37).
7: end for
8: Obtain the local optimal solution.
Step 5 : Global optimal solution.
1: While $t \leq iter$ do:
2: for all X_i do:
3: Update parameter X, C, σ by using Equations (18)–(25).
4: Obtain different predictions by using Equations (9)–(15).
5: end for
6: Using confidence, mean, standard deviation and other parameters, the prediction interval is obtained according to $norminv()$ formula.
7: Judge and update by Equations (33)–(37).
8: for all X_i do:
9: Update parameter X, C, σ by using Equations (26)–(32).
10: Obtain different predictions by using Equations (9)–(15).
11: Using confidence, mean, standard deviation and other parameters, the prediction interval is obtained according to $norminv()$ formula.
12: Judge and update by means of Equations (33)–(37).
13: end for
14: $t = t + 1$
15: end while
16: Obtain the global optimal solution.
Step 6: Return the global optimal prediction interval.
Step 7: Output C, σ, fitness and other index values by using Equations (33)–(37), (38)–(41).

4. Simulation Results

In this section, the data sets of BOD/NH3-N effluents are collected from a wastewater treatment plant in Beijing and are used to verify the effectiveness of the proposed approach.

The following evaluation indices of several certainty point predictions are evaluated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{38}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{39}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{40}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y} - y_i)^2}{\sum_{i=1}^{n}(\hat{y} - \bar{y})^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(\hat{y} - \bar{y})^2} \tag{41}$$

4.1. Experiment of Benchmark Functions

The proposed approach is based on the six functions listed in Table 3 with the corresponding ranges and parameters. The range is the boundary of the function search space.

In order to verify the superiority of the proposed approach, it is compared with the WOA, GWO, PSO and SSA algorithms. Statistical results are presented in Table 4. Moreover, the iteration process is depicted in Figures 3–8. From the results, we can see that the convergence rate of IBES is better than that of the other algorithms and the proposed IBES method is able to provide competitive results on the benchmark functions.

Table 3. Benchmark functions.

	Function	Range	Parameters
F1	$F(x) = -\sum_{i=1}^{10}\left[(X-a_i)(X-a_i)^T + c_i\right]^{-1}$	$[1, 10]$	dim = 4 popsize = 100 iteration = 300
F2	$F(x) = -\sum_{i=1}^{7}\left[(X-a_i)(X-a_i)^T + c_i\right]^{-1}$	$[1, 10]$	dim = 4 popsize = 100 iteration = 300
F3	$F(x) = -\sum_{i=1}^{5}\left[(X-a_i)(X-a_i)^T + c_i\right]^{-1}$	$[1, 10]$	dim = 4 popsize = 100 iteration = 300
F4	$F(x) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij}(x_j - p_{ij})^2\right)$	$[0, 1]$	dim = 6 popsize = 100 iteration = 200
F5	$F(x) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2\right)$	$[1, 3]$	dim = 3 popsize = 100 iteration = 120
F6	$F(x) = \left(\frac{1}{500} + \sum_{j=1}^{25}\frac{1}{j+\sum_{i=1}^{2}(x_i - a_{ij})^6}\right)^{-1}$	$[-65, 65]$	dim = 2 popsize = 100 iteration = 180

Table 4. Simulation results of algorithms.

	GWO	PSO	WOA	SSA	IBES	Theoretical Value
F1	−10.5364	−105364	−10.5364	−10.5364	**−10.5364**	**−10**
F2	−10.4042	−10.4029	−10.4029	−10.4029	**−10.4029**	**−10**
F3	−10.1561	−10.1532	−10.1576	−10.1532	**−10.1532**	**−10**
F4	−3.3220	−3.3311	−3.3231	−3.3220	**−3.3220**	**−3**
F5	−3.8628	−3.8628	−3.8627	−3.8628	**−3.8628**	**−3**
F6	0.9980	0.9980	0.9980	2.9821	**0.9980**	**1**

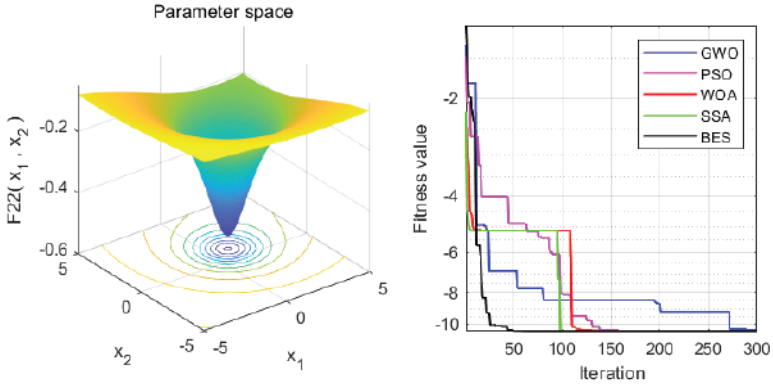


Figure 3. The result of F1.

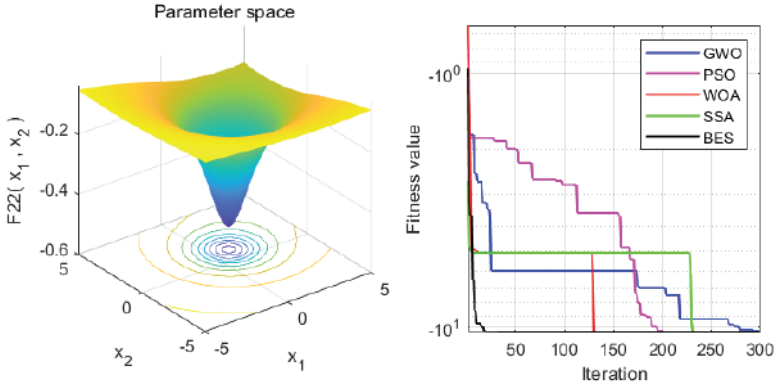


Figure 4. The result of F2.

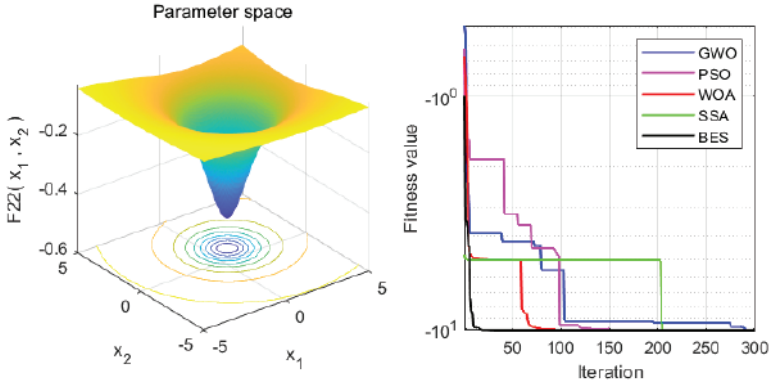


Figure 5. The result of F3.

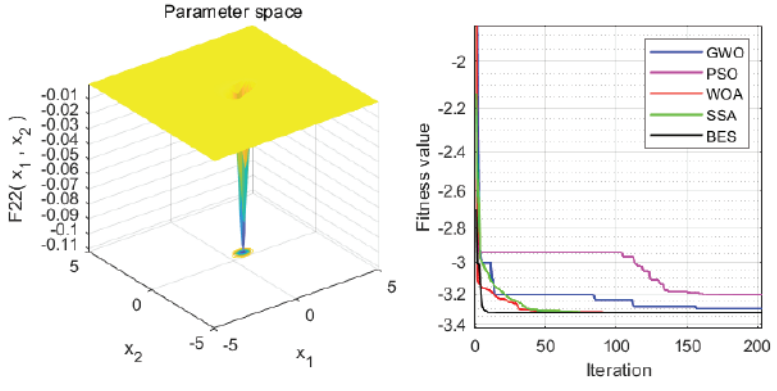


Figure 6. The result of F4.

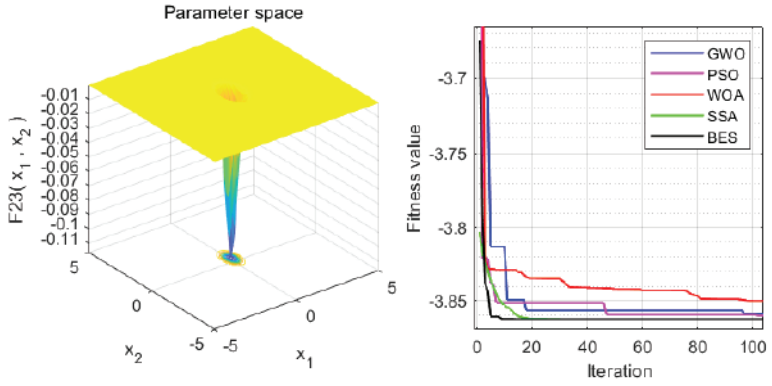


Figure 7. The result of F5.

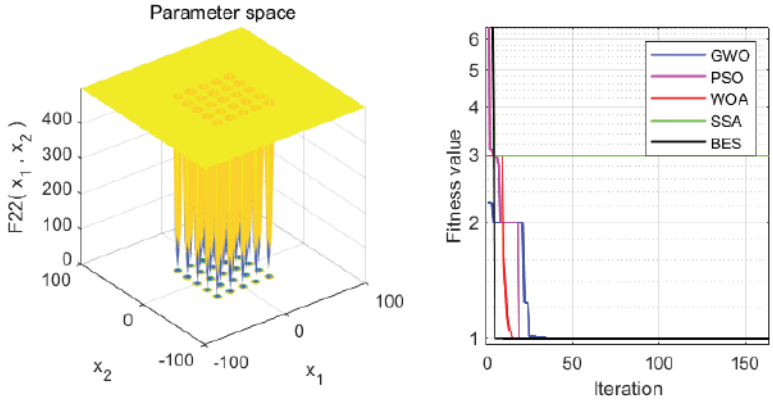


Figure 8. The result of F6.

4.2. Experiment of BOD Data

BOD is one of the most important effluent quality indexes and can reflect the water pollution situation [7]. First, the key auxiliary variables are selected for the BOD effluent data set by calculating the gray correlation degree based on (7). The threshold of the gray correlation degree is chosen as 0.8. Hence, 14 auxiliary variables (as shown in Table 5) are selected as the soft measurement model inputs. Including the output effluent BOD, there are 15 key indicators; the detailed information is shown in Figure 9. Moreover, the description of each datum is given in Figure 10.

In this paper, the BOD effluent data set has 365 sets of data; among them, 335 sets of data are randomly selected as training samples, and the remaining 30 sets of data are treated as the prediction samples. In order to demonstrate the superiority of the proposed IBES-LSSVM method, it is compared with some existing results, i.e., CNN, LSTM, ELMAN, WOA-LSSVM, GWO-LSSVM, PSO-LSSVM and SSA-LSSVM. In the experiments, the initialization conditions are set as: *iter* is 50, $n = 30$, $\omega_{max} = 10$, $\omega_{min} = 0$, $R_1 = 1.8$, $R_2 = 1$, $R_3 = 1.5$, $R_4 = 1.5$.

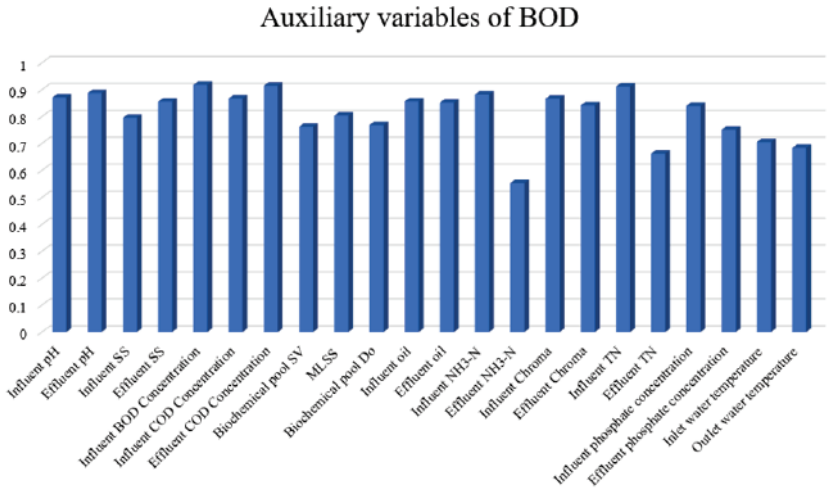


Figure 9. Auxiliary variables of BOD.

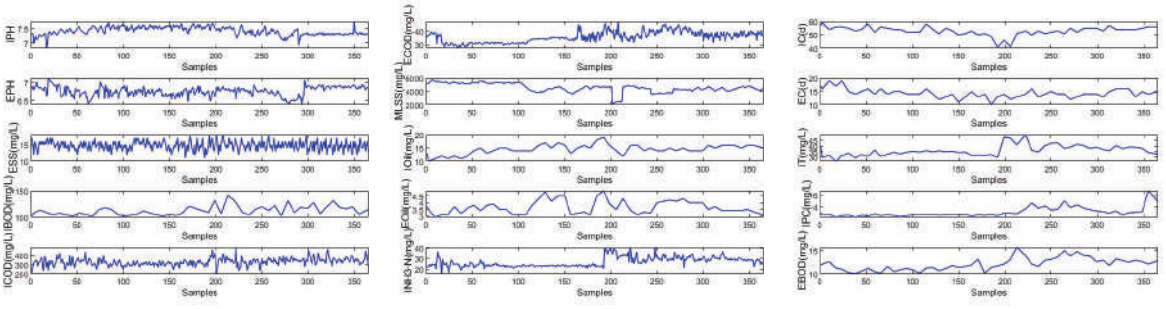
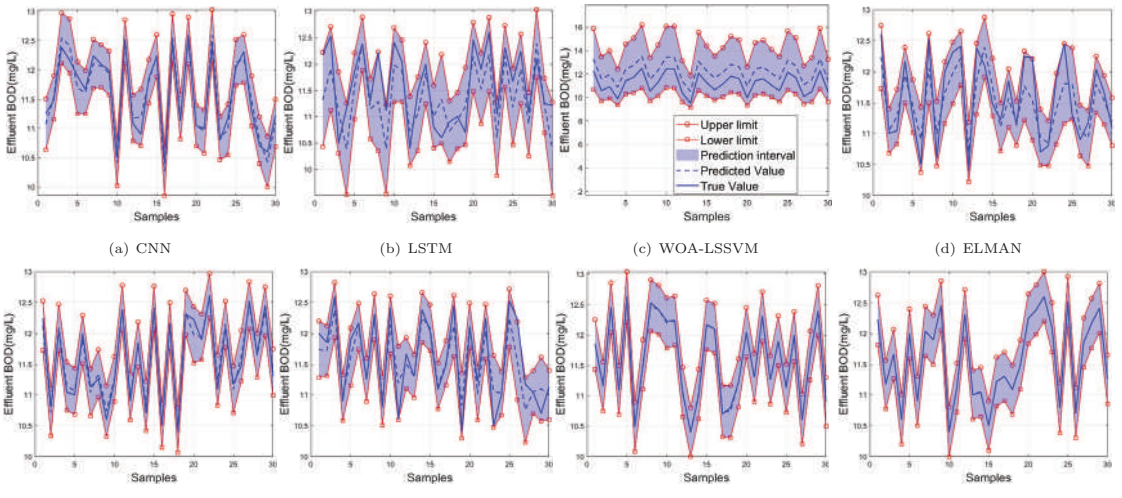


Figure 10. Original data of BOD.

Table 5. Data after processing.

Number of Coefficient	Auxiliary Variable	Correlation
1	Influent BOD	0.9179
2	Effluent COD	0.9151
3	Influent TN	0.9119
4	Effluent pH	0.8878
5	Influent NH3-N	0.8826
6	Influent pH	0.8716
7	Influent COD	0.8676
8	Influent Chroma	0.8669
9	Influent oil	0.8562
10	Effluent SS	0.8556
11	Effluent oil	0.8519
12	Effluent Chroma	0.8415
13	Influent phosphate	0.8397
14	MLSS in biochemical tank	0.8037

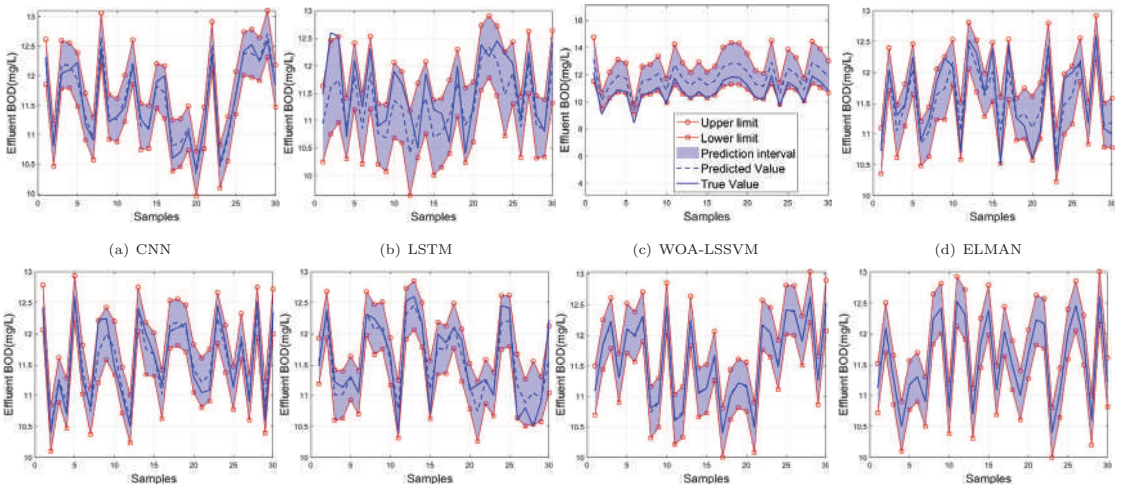
From Tables 6 and 7 and Figures 11–13, we can see that, compared with the existing CNN model, LSTM model, ELMAN model, WOA-LSSVM model, GWO-LSSVM model, PSO-LSSVM model and SSA-LSSVM model, the prediction accuracy of the proposed method is better, demonstrating its effectiveness.



(a) CNN (b) LSTM (c) WOA-LSSVM (d) ELMAN

(e) PSO-LSSVM (f) SSA-LSSVM (g) GWO-LSSVM (h) IBES-LSSVM

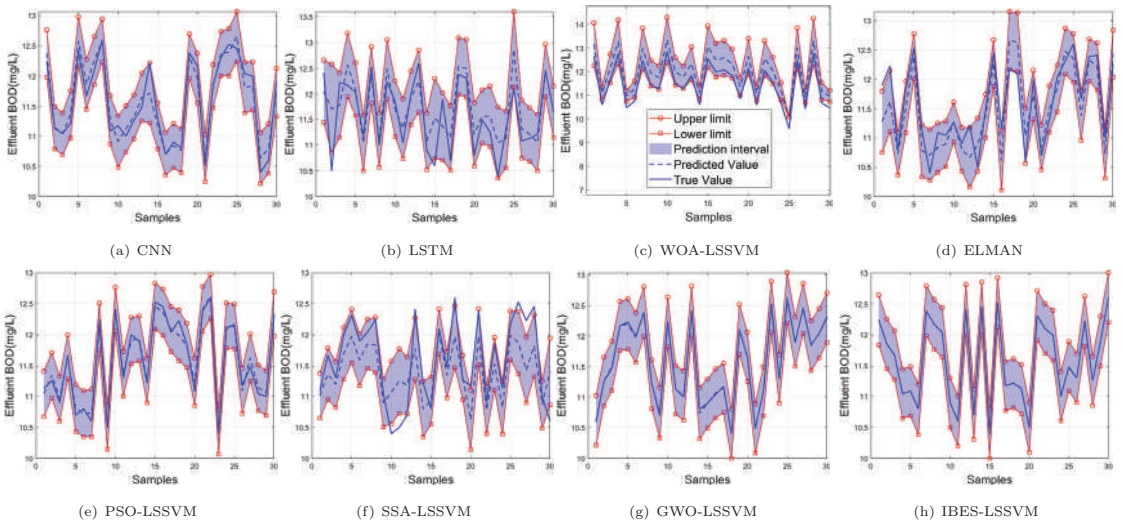
Figure 11. 99% of BOD.



(a) CNN (b) LSTM (c) WOA-LSSVM (d) ELMAN

(e) PSO-LSSVM (f) SSA-LSSVM (g) GWO-LSSVM (h) IBES-LSSVM

Figure 12. 95% of BOD.



(a) CNN (b) LSTM (c) WOA-LSSVM (d) ELMAN

(e) PSO-LSSVM (f) SSA-LSSVM (g) GWO-LSSVM (h) IBES-LSSVM

Figure 13. 90% of BOD.

Table 6. Predictive index of BOD.

Model	MSE	RMSE	MAE	R^2
CNN	0.0847	0.1500	0.1115	0.9503
LSTM	0.1310	0.2985	0.2330	0.8132
ELMAN	0.2425	0.3120	0.2523	0.7849
GWO-LSSVM	0.0659	0.0217	0.0182	0.9889
WOA-LSSVM	0.0711	0.1831	0.1521	0.9693
PSO-LSSVM	0.0587	0.1049	0.0851	0.9757
SSA-LSSVM	0.0726	0.2371	0.1707	0.9758
IBES-LSSVM	**0.0201**	**0.0104**	**0.0103**	**0.9911**

Table 7. PI of BOD.

	μ = 90%					μ = 95%					μ = 99%				
	PICP	**PINRW**	**CWC**	**PINAW**	**Time**	**PICP**	**PINRW**	**CWC**	**PINAW**	**Time**	**PICP**	**PINRW**	**CWC**	**PINAW**	**Time**
CNN	0.9298	0.2731	0.2731	0.2348	41.489	0.9617	0.3848	0.3848	0.3325	42.940	0.9911	0.2841	0.2841	0.2413	46.076
LSTM	0.9124	0.3632	0.3632	0.3112	27.486	0.9609	0.3796	0.3796	0.3254	27.731	0.9913	0.3554	0.3554	0.3020	27.821
ELMAN	0.9073	0.2978	0.2978	0.2474	316.316	0.9549	0.2573	0.2573	0.2202	241.446	0.9909	0.2571	0.2571	0.2132	90.582
WOA-LSSVM	0.9104	0.2663	0.2663	0.2325	1.686	0.9633	0.2697	0.2697	0.2346	1.873	0.9909	0.2673	0.2673	0.2245	1.654
GWO-LSSVM	0.9099	0.2557	0.2557	0.2241	1.396	0.9587	0.2668	0.2668	0.2355	1.389	0.9911	0.2689	0.2689	0.2254	2.012
PSO-LSSVM	0.9111	0.2519	0.2519	0.2198	1.029	0.9544	0.2596	0.2596	0.2155	0.967	0.9908	0.2773	0.2773	0.2277	0.963
SSA-LSSVM	0.9072	0.2901	0.2901	0.2543	1.428	0.9563	0.3178	0.3178	0.2613	1.410	0.9907	0.2961	0.2691	0.2245	1.599
IBES-LSSVM	**0.9053**	**0.2468**	**0.2468**	**0.2007**	**1.406**	**0.9531**	**0.2569**	**0.2569**	**0.2064**	**1.432**	**0.9907**	**0.2569**	**0.2569**	**0.2111**	**1.207**

4.3. Experiment of NH3-N Data

In this experiment, the NH3-N effluent data set is considered, which has been described in [44]. First, the gray correlation degree is calculated from (7), and the results are presented in Figure 14. In addition, each selected auxiliary datum of the NH3-N data set is shown in Figure 15.

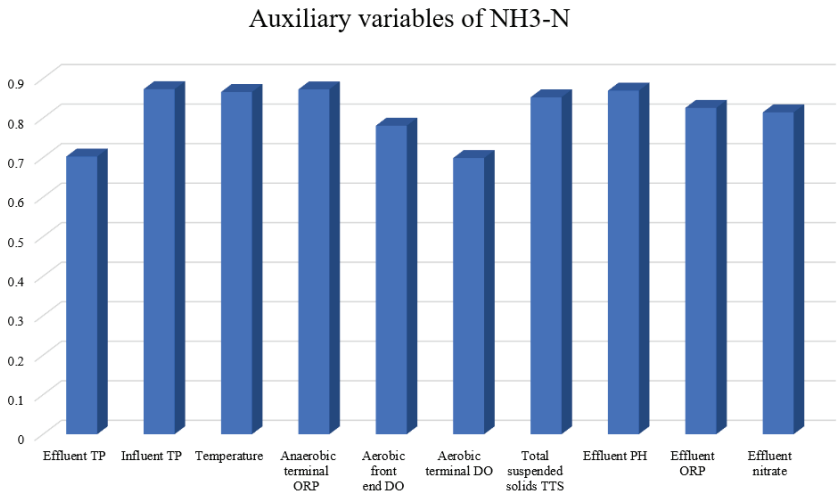


Figure 14. Auxiliary variables of NH3-N.

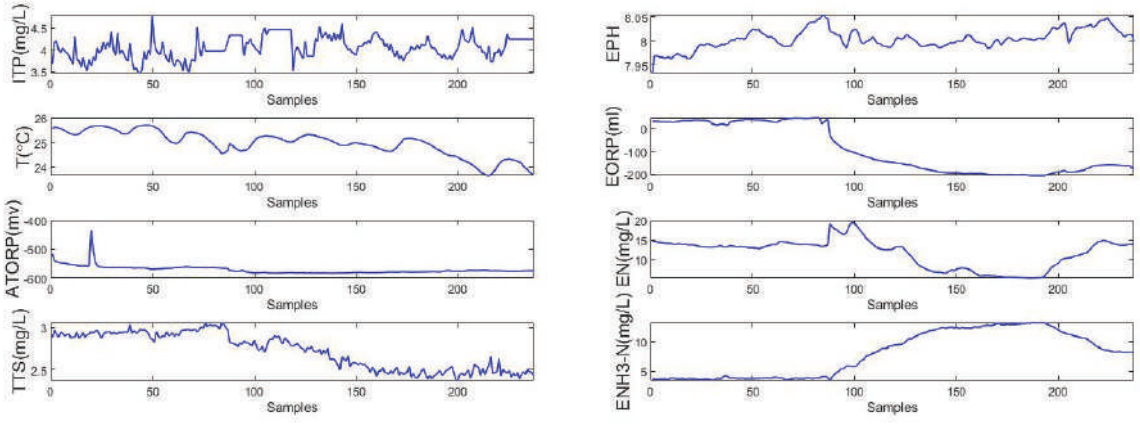


Figure 15. Original data of NH3-N.

In this example, the threshold of the gray correlation degree is also chosen as 0.8; hence, 7 auxiliary variables (as shown in Table 8) are selected as the soft measurement model input. The experimental data of effluent NH3-N used in this paper are from a sewage treatment plant in Beijing. In total, 237 sets of data were obtained, including 200 sets of data that were randomly selected as training samples, and the remaining 37 sets of data were treated as the prediction samples.

Table 8. Data after processing.

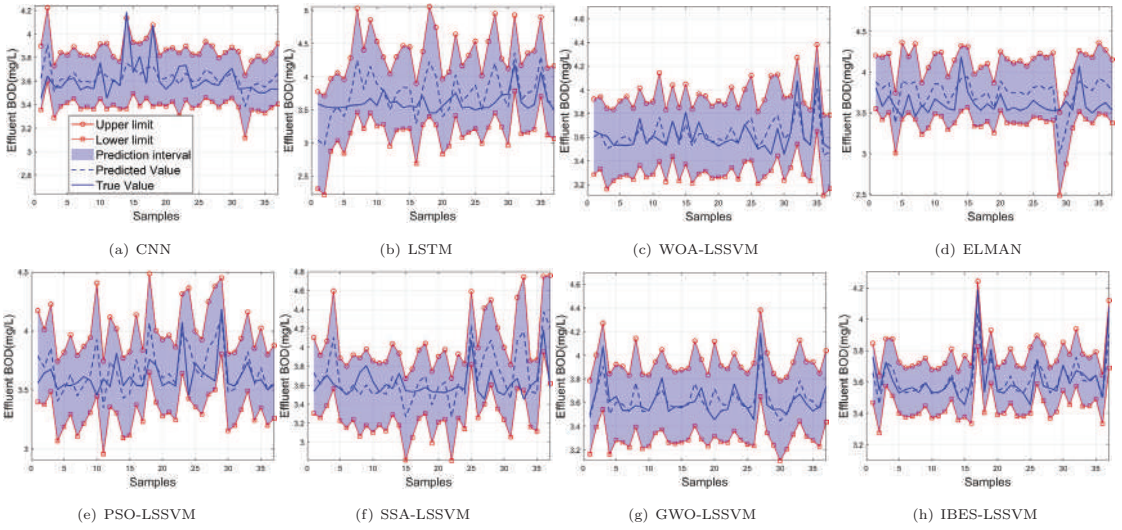
Number of Coefficient	Auxiliary Variable	Correlation
1	Influent TP	0.8730
2	Anaerobic terminal ORP	0.8726
3	Effluent PH	0.8693
4	Temperature	0.8659
5	Total suspended solids TTS	0.8525
6	Effluent ORP	0.8257
7	Effluent nitrate	0.8143

In order to demonstrate the superiority of the proposed BES-LSSVM method, it is compared with some existing approaches, i.e., CNN, LSTM, ELMAN, WOA-LSSVM, GWO-LSSVM, PSO-LSSVM and SSA-LSSVM. In the experiments, the parameters are set as follows: *iter* is 50, $n = 30$, $\omega_{max} = 10$, $\omega_{min} = 0$, $R_1 = 1.8$, $R_2 = 1.2$, $R_3 = 1.8$, $R_4 = 1.8$.

From Tables 9 and 10 and Figures 16–18, we can see that, compared with the existing CNN model, LSTM model, ELMAN model, WOA-LSSVM model, GWO-LSSVM model, PSO-LSSVM model and SSA-LSSVM model, the prediction accuracy of the proposed method is the best, demonstrating its effectiveness.

Table 9. PI of NH3-N.

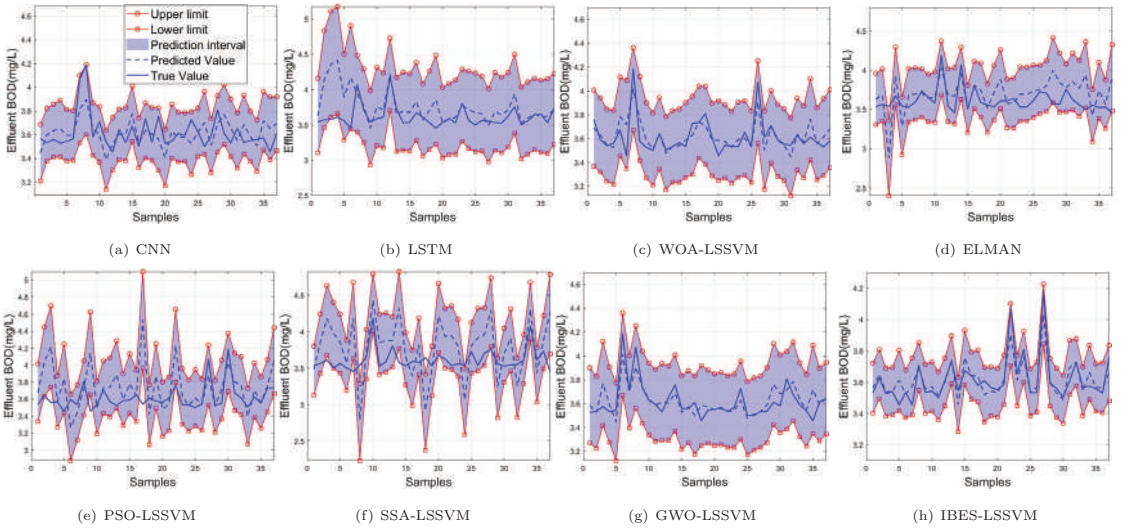
	μ = 90%					μ = 95%					μ = 99%				
	PICP	**PINRW**	**CWC**	**PINAW**	**Time**	**PICP**	**PINRW**	**CWC**	**PINAW**	**Time**	**PICP**	**PINRW**	**CWC**	**PINAW**	**Time**
CNN	0.9231	0.53951	0.53951	0.50111	29.991	0.9619	0.49776	0.49776	0.46854	32.446	0.9919	0.52063	0.52063	0.48445	31.703
LSTM	0.9182	0.49437	0.49437	0.44235	22.176	0.9588	0.42320	0.42320	0.37824	22.637	0.9921	0.53185	0.53185	0.50111	21.272
ELMAN	0.9066	0.38637	0.38637	0.34255	6.661	0.9580	0.37625	0.37625	0.32142	3.175	0.9912	0.42032	0.42032	0.38764	3.120
WOA-LSSVM	0.9197	0.49711	0.49711	0.45739	1.547	0.9581	0.46106	0.46106	0.42131	1.711	0.9913	0.47562	0.47562	0.41121	1.584
GWO-LSSVM	0.9227	0.51067	0.51067	0.46174	1.346	0.9601	0.51117	0.51117	0.47894	1.166	0.9913	0.51776	0.51776	0.45669	1.163
PSO-LSSVM	0.9241	0.48209	0.48209	0.45394	0.959	0.9604	0.47815	0.47815	0.42756	0.797	0.9917	0.49209	0.49209	0.46401	0.801
SSA-LSSVM	0.9112	0.40579	0.40579	0.35752	1.363	0.9574	0.38947	0.38947	0.34556	1.184	0.9909	0.38777	0.38777	0.36454	1.142
IBES-LSSVM	**0.9037**	**0.34531**	**0.34531**	**0.30989**	**1.354**	**0.9556**	**0.34906**	**0.34906**	**0.31128**	**1.181**	**0.9907**	**0.34677**	**0.34677**	**0.31001**	**1.366**



(a) CNN (b) LSTM (c) WOA-LSSVM (d) ELMAN

(e) PSO-LSSVM (f) SSA-LSSVM (g) GWO-LSSVM (h) IBES-LSSVM

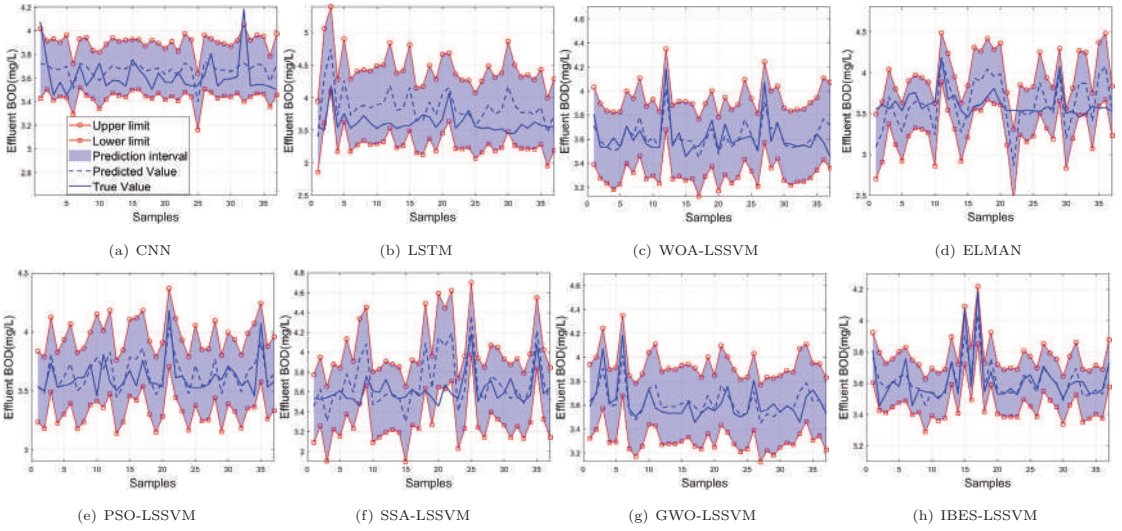
Figure 16. 99% of NH3-N.



(a) CNN (b) LSTM (c) WOA-LSSVM (d) ELMAN

(e) PSO-LSSVM (f) SSA-LSSVM (g) GWO-LSSVM (h) IBES-LSSVM

Figure 17. 95% of NH3-N.



(a) CNN (b) LSTM (c) WOA-LSSVM (d) ELMAN

(e) PSO-LSSVM (f) SSA-LSSVM (g) GWO-LSSVM (h) IBES-LSSVM

Figure 18. 90% of NH3-N.

Table 10. Predictive index of NH3-N.

Model	MSE	RMSE	MAE	R^2
CNN	0.1874	0.1711	0.1450	0.8932
LSTM	0.1138	0.2131	0.1663	0.7666
ELMAN	0.0954	0.1846	0.1564	0.7872
GWO-LSSVM	0.0997	0.0895	0.0628	0.7280
WOA-LSSVM	0.1929	0.2371	0.1709	0.8959
PSO-LSSVM	0.1312	0.1722	0.1247	0.8922
SSA-LSSVM	0.1196	0.1958	0.2037	0.8117
IBES-LSSVM	**0.0917**	**0.0645**	**0.0450**	**0.8967**

5. Conclusions

This paper investigates an improved IBES-LSSVM algorithm to predict the effluent water quality indicators of a wastewater treatment plant, in which an improved BES method is proposed to find the optimal LSSVM parameters. To deal with the uncertainties of the data, the prediction interval is generated within a certain confidence level, which could provide the upper and lower bounds of the prediction results. Compared with other existing methods, the proposed approach demonstrates high prediction accuracy, with reduced computational time and an easy calculation process, in predicting effluent water quality parameters. Note that the proposed results can only predict the water quality indicators, but this is not the end work for a wastewater treatment plant process. The application of this work to reliable decision-making and the generation of a suitable control strategy will be our future work.

Author Contributions: Conceptualization M.Z., V.P.; methodology M.Z., J.W., Y.S., V.P.; resources Y.S.; writing-review and editing M.Z., V.P., Y.Z.; supervision M.Z., J.W.; investigation J.W.; formal analysis M.Z.; software and data curation Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Key Research and Development Program (Grant No. 2018YFC1602704), National Natural Science Foundation of China (Grant No. 51805021,61973023), the Fundamental Research Funds for Beijing (No. 110052972027/015) and the Research foundation for Talents of NCUT (No. 213051360020XN173/017), "Science and Technology Innovation" Special Fund Project in Shijingshan District: 401053762117.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available upon request.

Acknowledgments: This work was carried out with the help of Dingyuan Chen at Beijing University of Technology.

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

References

1. Julio, N.; Figueroa, R.; Ponce Oliva, R.D. Water resources and governance approaches: Insights for achieving water security. *Water* **2021**, *13*, 3063. [CrossRef]
2. Abrams, A.L.; Carden, K.; Teta, C.; Wågsæther, K. Water, sanitation, and hygiene vulnerability among rural areas and small towns in south Africa: Exploring the role of climate change, marginalization, and inequality. *Water* **2021**, *13*, 2810. [CrossRef]
3. Shannon, M.A.; Bohn, P.W.; Elimelech, M.; Georgiadis, J.G.; Mariñas, B.J.; Mayes, A.M. Science and technology for water purification in the coming decades. *Nature* **2008**, *452*, 301–310. [CrossRef]
4. Olsson, G. ICA and me—A subjective review. *Water Res.* **2012**, *46*, 1585–1624. [CrossRef]
5. Thürlimann, C.M.; Dürrenmatt, D.J.; Villez, K. Soft-sensing with qualitative trend analysis for wastewater treatment plant control. *Control Eng. Pract.* **2018**, *70*, 121–133. [CrossRef]
6. Vorosmarty, C.J.; Hoekstra, A.Y.; Bunn, S.E.; Conway, D.; Gupta, J. Fresh water goes global. *Science* **2015**, *349*, 478–479. [CrossRef] [PubMed]
7. Qiao, J.; Wang, L.; Yang, C.; Gu, K. Adaptive Levenberg-Marquardt algorithm based echo state network for Chaotic time series prediction. *IEEE Access* **2018**, *6*, 10720–10732. [CrossRef]
8. Chen, Y.; Han, D. Water quality monitoring in smart city: A pilot project. *Autom. Constr.* **2018**, *89*, 307–316. [CrossRef]
9. Lu, H.; Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* **2020**, *249*, 126169. [CrossRef]
10. Zhou, P.; Li, Z.; Snowling, S.; Baetz, B.W.; Na, D.; Boyd, G. A random forest model for inflow prediction at wastewater treatment plants. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1781–1792. [CrossRef]
11. Liu, H.; Zhang, H.; Zhang, Y.; Zhang, F.; Huang, M. Modeling of Wastewater Treatment Processes Using Dynamic Bayesian Networks Based on Fuzzy PLS. *IEEE Access* **2020**, *8*, 92129–92140. [CrossRef]
12. Manandhar, A.; Fischer, A.; Bradley, D.J.; Salehin, M.; Islam, M.S.; Hope, R.; Clifton, D.A. Machine learning to evaluate impacts of flood protection in Bangladesh, 1983–2014. *Water* **2020**, *12*, 483. [CrossRef]

13. de Canete, J.F.; Orozco, P.D.S.; Baratti, R.; Mulas, M.; Ruano, A.; Garcia-Cerezo, A. Soft-sensing estimation of plant effluent concentrations in a biological wastewater treatment plant using an optimal neural network. *Expert Syst. Appl.* **2016**, *63*, 8–19. [CrossRef]
14. Cong, Q.; Yu, W. Integrated soft sensor with wavelet neural network and adaptive weighted fusion for water quality estimation in wastewater treatment process. *Measurement* **2018**, *124*, 436–446. [CrossRef]
15. Hamada, M.; Adel Zaqoot, H.; Abu Jreiban, A. Application of artificial neural networks for the prediction of Gaza wastewater treatment plant performance-Gaza strip. *J. Appl. Res. Water Wastewater* **2018**, *5*, 399–406. [CrossRef]
16. Zeinolabedini, M.; Najafzadeh, M. Comparative study of different wavelet-based neural network models to predict sewage sludge quantity in wastewater treatment plant. *Environ. Monit. Assess.* **2019**, *191*, 1–25. [CrossRef]
17. Kadam, A.; Wagh, V.; Muley, A.; Umrikar, B.; Sankhua, R. Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Model. Earth Syst. Environ.* **2019**, *5*, 951–962. [CrossRef]
18. Heddam, S.; Lamda, H.; Filali, S. Predicting effluent biochemical oxygen demand in a wastewater treatment plant using generalized regression neural network based approach: A comparative study. *Environ. Process.* **2016**, *3*, 153–165. [CrossRef]
19. Tan, T.; Yang, Z.; Chang, F.; Zhao, K. Prediction of the First Weighting from the Working Face Roof in a Coal Mine Based on a GA-BP Neural Network. *Appl. Sci.* **2019**, *9*, 4159. [CrossRef]
20. Nourani, V.; Elkiran, G.; Abba, S.I. Wastewater treatment plant performance analysis using artificial intelligence—An ensemble approach. *Water Sci. Technol.* **2018**, *78*, 2064–2076. [CrossRef]
21. Cheng, T.; Dairi, A.; Harrou, F.; Sun, Y.; Leiknes, T. Monitoring influent conditions of wastewater treatment plants by nonlinear data-based techniques. *IEEE Access* **2019**, *7*, 108827–108837. [CrossRef]
22. Han, H.; Liu, H.; Liu, Z.; Qiao, J. Fault detection of sludge bulking using a self-organizing type-2 fuzzy-neural-network. *Control Eng. Pract.* **2019**, *90*, 27–37. [CrossRef]
23. Wu, J.; Cheng, H.; Liu, Y.; Liu, B.; Huang, D. Modeling of adaptive multi-output soft-sensors with applications in wastewater treatments. *IEEE Access* **2019**, *7*, 161887–161898. [CrossRef]
24. Lotfi, K.; Bonakdari, H.; Ebtehaj, I.; Mjalli, F.S.; Zeynoddin, M.; Delatolla, R.; Gharabaghi, B. Predicting wastewater treatment plant quality parameters using a novel hybrid linear-nonlinear methodology. *J. Environ. Manag.* **2019**, *240*, 463–474. [CrossRef]
25. Han, H.; Zhang, L.; Qiao, J. Data-based predictive control for wastewater treatment process. *IEEE Access* **2017**, *6*, 1498–1512. [CrossRef]
26. Mateo Pérez, V.; Mesa Fernández, J.M.; Ortega Fernández, F.; Villanueva Balsera, J. Gross Solids Content Prediction in Urban WWTPs Using SVM. *Water* **2021**, *13*, 442. [CrossRef]
27. Ribeiro, D.; Sanfins, A.; Belo, O. Wastewater Treatment Plant Performance Prediction with Support Vector Machines. In *Advances in Data Mining. Applications and Theoretical Aspects*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 99–111. [CrossRef]
28. Baghban, A.; Sasanipour, J.; Habibzadeh, S.; Zhang, Z. Sulfur dioxide solubility prediction in ionic liquids by a group contribution—LSSVM model. *Chem. Eng. Res. Des.* **2019**, *142*, 44–52. [CrossRef]
29. Zhang, N.; Kamaha, R.; Behera, P. Prediction of Surface Water Supply Sources for the District of Columbia Using Least Squares Support Vector Machines (LS-SVM) Method. *Adv. Comput. Sci. Int. J.* **2015**, *4*, 1–9.
30. Luo, F.; Qiao, X.; Liao, W. Soft-sensing modeling based on GK-LSSVM method for online predictions of BOD in activated sludge process. In Proceedings of the 2017 International Conference on Robotics and Automation Sciences (ICRAS), Hong Kong, China, 26–29 August 2017; pp. 134–138.
31. Manu, D.; Thalla, A.K. Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl Nitrogen from wastewater. *Appl. Water Sci.* **2017**, *7*, 3783–3791. [CrossRef]
32. Liu, Z.; Wan, J.; Ma, Y.; Wang, Y. Online prediction of effluent COD in the anaerobic wastewater treatment system based on PCA-LSSVM algorithm. *Environ. Sci. Pollut. Res.* **2019**, *26*, 12828–12841. [CrossRef] [PubMed]
33. Nieto, P.G.; García Gonzalo, E.; Arbat, G.; Duran Ros, M.; de Cartagena, F.R.; Puig Bargués, J. A new predictive model for the filtered volume and outlet parameters in micro-irrigation sand filters fed with effluents using the hybrid PSO-SVM-based approach. *Comput. Electron. Agric.* **2016**, *125*, 74–80. [CrossRef]
34. Han, H.; Lu, W.; Hou, Y.; Qiao, J. An adaptive-PSO-based self-organizing RBF neural network. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *29*, 104–117. [CrossRef] [PubMed]
35. Chen, S.; Fang, G.; Huang, X.; Zhang, Y. Water quality prediction model of a water diversion project based on the improved artificial bee colony–backpropagation neural network. *Water* **2018**, *10*, 806. [CrossRef]
36. Zeng, F.; Amar, M.N.; Mohammed, A.S.; Motahari, M.R.; Hasanipanah, M. Improving the performance of LSSVM model in predicting the safety factor for circular failure slope through optimization algorithms. *Eng. Comput.* **2021**, 1–12. [CrossRef]
37. Shariati, M.; Mafipour, M.S.; Ghahremani, B.; Azarhomayun, F.; Ahmadi, M.; Trung, N.T.; Shariati, A. A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement. *Eng. Comput.* **2020**, 1–23. [CrossRef]
38. Yao, W.; Zeng, Z.; Lian, C. Generating probabilistic predictions using mean-variance estimation and echo state network. *Neurocomputing* **2017**, *219*, 536–547. [CrossRef]
39. Yuan, X.; Chen, C.; Jiang, M.; Yuan, Y. Prediction interval of wind power using parameter optimized Beta distribution based LSTM model. *Appl. Soft Comput.* **2019**, *82*, 105550. [CrossRef]

40. Liao, Y.; Zhang, L.; Liu, C. Uncertainty prediction of remaining useful life using long short-term memory network based on bootstrap method. In Proceedings of the 2018 IEEE International Conference on Prognostics and Health Management (ICPHM), Seattle, WA, USA, 11–13 June 2018; pp. 1–8.
41. Marín, L.G.; Cruz, N.; Sáez, D.; Sumner, M.; Núñez, A. Prediction interval methodology based on fuzzy numbers and its extension to fuzzy systems and neural networks. *Expert Syst. Appl.* **2019**, *119*, 128–141. [CrossRef]
42. Sun, X.; Wang, Z.; Hu, J. Prediction interval construction for byproduct gas flow forecasting using optimized twin extreme learning machine. *Math. Probl. Eng.* **2017**, *2017*, 5120704. [CrossRef]
43. Wu, D.; Wang, H.; Seidu, R. Smart data driven quality prediction for urban water source management. *Future Gener. Comput. Syst.* **2020**, *107*, 418–432. [CrossRef]
44. Chen, Z.; Yang, C.; Qiao, J. The optimal design and application of LSTM neural network based on the hybrid coding PSO algorithm. *J. Supercomput.* **2021**, *5*. [CrossRef]
45. Su, Y.; Han, L.; Wang, J.; Wang, H. Quantum-behaved RS-PSO-LSSVM method for quality prediction in parts production processes. In *Concurrency and Computation Practice and Experience*; John Wiley & Sons: Hoboken, NJ, USA, 2019; pp. 1–15.
46. Li, R.; Jin, Y. A wind speed interval prediction system based on multi-objective optimization for machine learning method. *Appl. Energy* **2018**, *228*, 2207–2220. [CrossRef]

sensors

MDPI

Article

Estimation of Infiltration Volumes and Rates in Seasonally Water-Filled Topographic Depressions Based on Remote-Sensing Time Series

Pavel P Fil [1,2,*], Alla Yu Yurova [1], Alexey Dobrokhotov [3] and Daniil Kozlov [1]

[1] V. Dokuchaev Soil Science Institute, 109017 Moscow, Russia; alla.yurova@gmail.com (A.Y.Y.); daniilkozlov@gmail.com (D.K.)
[2] Department of Geography, Lomonosov Moscow State University, 119991 Moscow, Russia
[3] Agrophysical Research Institute, 195220 St. Petersburg, Russia; dobralexey@gmail.com
* Correspondence: philpaulp@gmail.com

Abstract: In semi-arid ecoregions of temperate zones, focused snowmelt water infiltration in topographic depressions is a key, but imperfectly understood, groundwater recharge mechanism. Routine monitoring is precluded by the abundance of depressions. We have used remote-sensing data to construct mass balances and estimate volumes of temporary ponds in the Tambov area of Russia. First, small water bodies were automatically recognized in each of a time series of high-resolution Planet Labs images taken in April and May 2021 by object-oriented supervised classification. A training set of water pixels defined in one of the latest images using a small unmanned aerial vehicle enabled high-confidence predictions of water pixels in the earlier images (Cohen's K = 0.99). A digital elevation model was used to estimate the ponds' water volumes, which decreased with time following a negative exponential equation. The power of the exponent did not systematically depend on the pond size. With adjustment for estimates of daily Penman evaporation, function-based interpolation of the water bodies' areas and volumes allowed calculation of daily infiltration into the depression beds. The infiltration was maximal (5–40 mm/day) at onset of spring and decreased with time during the study period. Use of the spatially variable infiltration rates improved steady-state shallow groundwater simulations.

Keywords: closed depressions; temporary water bodies; remote sensing; infiltration

Citation: Fil, P.P.; Yurova, A.Y.; Dobrokhotov, A.; Kozlov, D. Estimation of Infiltration Volumes and Rates in Seasonally Water-Filled Topographic Depressions Based on Remote-Sensing Time Series. *Sensors* **2021**, *21*, 7403. https://doi.org/10.3390/s21217403

Academic Editors: Miquel Àngel Cugueró-Escofet and Vicenç Puig

Received: 14 September 2021
Accepted: 3 November 2021
Published: 7 November 2021

1. Introduction

Shallow groundwater is present in many semi-arid landscapes across the world either intermittently or permanently, depending on the lithological profile, topography, and water balance. Unlike in wetter environments with diffuse groundwater recharge, recharge in these environments is primarily focused (local) in areas of excess water input [1]. In such environments, where moisture deficits in upland soils are high, groundwater recharge will only occur if there is sufficient infiltration of converging flow to overcome the deficits. One mechanism involved is a localized recharge process that routes surface water runoff within the landscape to topographically low areas (depressions), allowing infiltration of water through ephemeral seasonal ponds [2–4]. Moreover, depression-focused recharge driven by snowmelt is a major annual hydrological event in cold semi-arid regions such as the Pothole Prairie Region of North America. In recent decades, there has been an accelerated increase in process understanding of the contributions of prairie potholes to surface runoff [5,6] and depression-focused groundwater recharge [3] in this part of North America. The knowledge has been acquired through studies involving conceptual and mathematical modeling of hydrological processes of surface flows [5–7], subsurface flows and combinations of the two [8,9], applications of isotopic and environmental tracers [3], digital elevation model (DEM)-based delineations of depressions and their watersheds [10–14], assessments of

hydrologic connectivity [6,10,15–17], and remote sensing with high and intermediate resolution [18,19]. Studies in various catchments have shown that both horizontal and vertical connectivity in pothole hydrological systems are very site-specific and no model can be applied to a new system without validation.

In Russia, areas rich in pothole-like systems of depressions ("zapadiny") in interfluves of forest-steppe catchments cover a larger region than in North America, extending across much of European Russia and into Siberia. However, after an initial period of intensive hydrological research in the 1960s to the 1980s there was very little study of depression-focused groundwater recharge despite advances in GIS-facilitated simulation and remote sensing. Moreover, there is increasing societal need for such studies to enhance the understanding of key landscape functions related to water storage or movement, e.g., water capacitance, carbon sequestration, and both nutrient retention and cycling [17,20] and precision agricultural management. With some justification, early studies noted similarities between prairie potholes and forest-steppe zapadiny. However, before applying tools developed in North American research to Russian systems, there is a need for quantitative evaluation of concepts that emerged in earlier local studies.

One of the key hypotheses developed during the 1960s is that the major source of recharge for shallow groundwater in areas such as the Oka-Don Lowland of the Tambov region in European Russia is depression-focused infiltration during snowmelt [21]. In a very recent study an indirect method was used to calibrate the groundwater recharge to hydraulic conductivity ratio for application in an analytical steady-state solution of the 2D shallow groundwater flow equation using soil redoximorphic features of typical classified catenas of the Samovetc catchment in this lowland [22]. In the cited study, the same recharge rate was prescribed for all points along a topographical transect. In contrast, in the study presented here, the spatial variability of depression-focused groundwater recharge along the transect was studied in a field campaign in spring 2021 during, immediately after snowmelt, and several weeks later.

There is no single method for classifying remote-sensing data for the ponds' retrieval. The methods and materials used vary greatly depending on the region of study, season of the year, image resolution or type of the pond. In terms of wavelengths used in the electromagnetic spectrum, they are visible (RGB), near infrared (NIR), shortwave infrared (SWIR) and thermal infrared (TIR) [23,24]. In addition to optical methods, data from RADAR and LIDAR are also used [25]. Methods for extraction of small water bodies are divided into four groups. The first group is the threshold methods, the essence of which is the discretization of individual spectral channels or spectral indicators based on expert or experimental threshold values [26,27]. The second group of methods are statistical methods, such as those using multivariate regression or discriminant analysis. Classification methods (the third group) are a matrix of combinations of different methods—this is a pixel or object-oriented approach, classifications with or without training, various classification machines; for example: a random forest or support vector machine, neural algorithms [28–32]. There are also various special techniques (group four) such as entropy-based computer vision techniques [33].

In this work, remote-sensing data were used to construct a mass balance and estimate volumes of ephemeral ponds by object-oriented supervised classification of high-resolution Planet Labs images of the Tambov area acquired from April to May 2021. The data acquired on dynamic changes in delineated ponds, in combination with a DEM, observations using an unmanned aerial vehicle (UAV), a widely accepted method for calculating evaporation, and visual hydrological observations were used to estimate infiltration volumes and rates through the depression bottoms and account for their spatial and temporal variability. Considering that groundwater recharge from the depressions' bottom is very area-focused and occurs episodically during the snowmelt, the process is not usually accounted for in the regional-scale evaluation of the groundwater resources in Tambov region. To include the impacts of spatial heterogeneity and dynamic fluctuation the depression-focused infiltration may be modeled numerically [8]. To examine the early hypothesis [21] on a

critical role of depressions in ground-water recharge in a forest-steppe region through the simplified approach for estimating recharge, this paper aims: (1) to determine the variations of the pond recession and infiltration rate in time and between the depressions due to systematic (vertically-varying hydraulic conductivity) and random factors (presence of clogging or frozen layers, pond- surface drainage network connection); (2) to determine the role of such spatial variations through numerical analysis of shallow groundwater model for the simplified 2D case; and (3) to determine a method for calculating the volume of recharge through depression in other catchments both with and without the requirement of numerical modeling and data assimilation. The result allowed identification of the volume of intercepted water during snowmelt and calculation of the rate of water recession and infiltration rates in closed depressions for the first time for the study region. Use of the horizontal variation in parameters obtained along the studied transect substantially improved results of the shallow ground water model developed in the cited study [22].

2. Materials and Methods

Remote-sensing data were used to construct a mass balance and estimate volumes of ephemeral ponds by object-oriented supervised classification of high-resolution Planet Labs images of the Tambov area acquired from April to May 2021. The data acquired on dynamic changes in delineated ponds, in combination with a DEM, observations using an unmanned aerial vehicle (UAV), a widely accepted method for calculating evaporation, and visual hydrological observations were used to estimate infiltration volumes and rates through the depression bottoms and account for their spatial and temporal variability. The acquired time series of changes in the volume of nine temporary ponds enabled parametrization with a negative exponential curve. A time series of the infiltration rate, calculated from the water balance, was used to estimate the total amount accumulated during the event, and both the initial (maximum) and saturated (minimum) infiltration rates per unit area.

2.1. Study Area

The study area covers approximately 560 ha in the center of the Oka-Don lowland (52°37′ N, 40°2′ E) in the Petrovsky district of the Tambov region, Russia (Figure 1). The lowland is the largest in the forest-steppe biome. With elevation ranging from 120 to 180 m above sea level, on average it is 100 m lower than adjacent territories. The lowland has a semi-arid climate with long winters, pronounced spring snowmelt events and relatively dry summers with an annual precipitation to potential evapotranspiration ratio of 0.8. According to data recorded at a meteorological station 10 km north of the study site, during the period 2005–2020 the annual temperature was 6.9 °C, and average monthly temperatures in January and July were −8.6 °C and 21.0 °C, respectively [34]. Mean annual precipitation during this period amounted to 550 mm (of which 113 mm fell during periods with sub-zero temperatures), and the mean snow height before onset of snowmelt was 320 mm, very similar to the recorded historical climatic norm for 1961–1990 (290 mm).

The soils are mainly chernozems and the area is mainly used for cultivating crops (typically wheat, corn, sunflower, soy, sugar beet), despite hindrance by water shortages. Clay and loamy deposits, generally 5–15 m (but sometimes up to 40 m) thick, with boulders of glacial origin, underlie a layer of loess-like loam with thickness ranging from 2 m in the lower parts of slopes to 30 m in the interfluve. The upper layer is porous and can both accumulate and retain moisture, while the glacial clays and loams form a local aquiclude for infiltrated surface waters. Shallow groundwater above this aquiclude is permanent and forms a continuous layer in the focal catchment. Evidence of stagnic condition in topsoil is restricted to the presence of albic material in the lower part of the humus horizon in grey gleysols in the depression bottom. There is clear evidence of gleyic conditions in the soil morphology (Fe-Mn concretions, Fe masses, pore lining, reduced matrix) in the catchment and continuous presence of water saturation below 2–3 m depth in the poorly drained soils and 1 m depth in the waterlogged soils. The latter was confirmed by a few cases of drilling

in different years and seasons (the WTD is consistently highest after the springflood) and automated measurements in 2019 (a year with extremely low snow accumulation). Physical properties of the surface loams have contributed to development of closed depressions of multifactorial genesis, which are widely spread throughout the Oka-Don lowland. These depressions delay the runoff of surface waters into rivers [21] and transfer surface runoff to groundwater, thereby replenishing the groundwater and moistening the surrounding soil. The closed depressions are filled with water in the spring when the snow melts. Snow located in the catchment area of each basin melts and replenishes it, usually in mid-March to early April. At the end of spring, surface water only remains in small parts of the depressions, and in summer they usually dry up completely, in contrast to the closed depressions of the Pothole Prairie Region. Monitoring the dynamics of water volume and its filtration enables estimation of amounts of valuable additional moisture entering the soil in this semi-arid region.

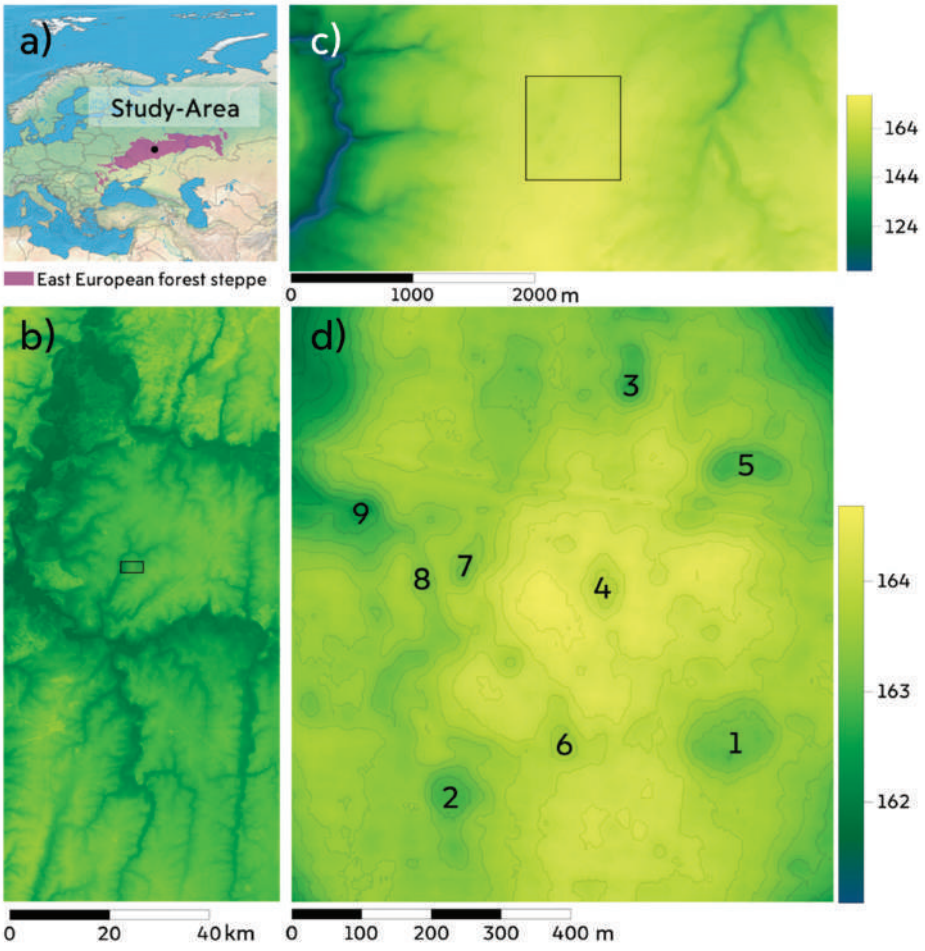


Figure 1. The study area. (**a**) Location of the study area in the forest-steppe biome of Eurasia. (**b**) Location of the study area in the catchment of the Matyr River-Oka-Don Lowland (SRTM). (**c**) Digital elevation model (DEM) of the interfluve of the Samovets brook; the studied territory of the depressions is marked with a rectangle. (**d**) Unmanned aerial vehicle (UAV) DEM of the study area, the numbers indicate numbers of closed depressions filled with pond water in the spring of 2021 (for details, see Table 2).

2.2. Input Data

Water dynamics in a closed depression in spring 2021 was tracked and modeled using the following three types of data (Figure 2): ultra-high resolution (25 cm) digital terrain and elevation models obtained using a small UAV—DJI Mavic 2 Pro, orthophotomaps of terrain in the visible range with a ultra-high resolution (25 cm) from UAV, orthophotomaps of high resolution (3 m) in the visible range from the sensors of the RapidEye and SkySat mini-satellites of the Planet Labs system [25]. Precipitation data were obtained from the nearest weather station with daily resolution. Evaporation data from the water surface was obtained using Penman's equation and meteorological input from the same station.

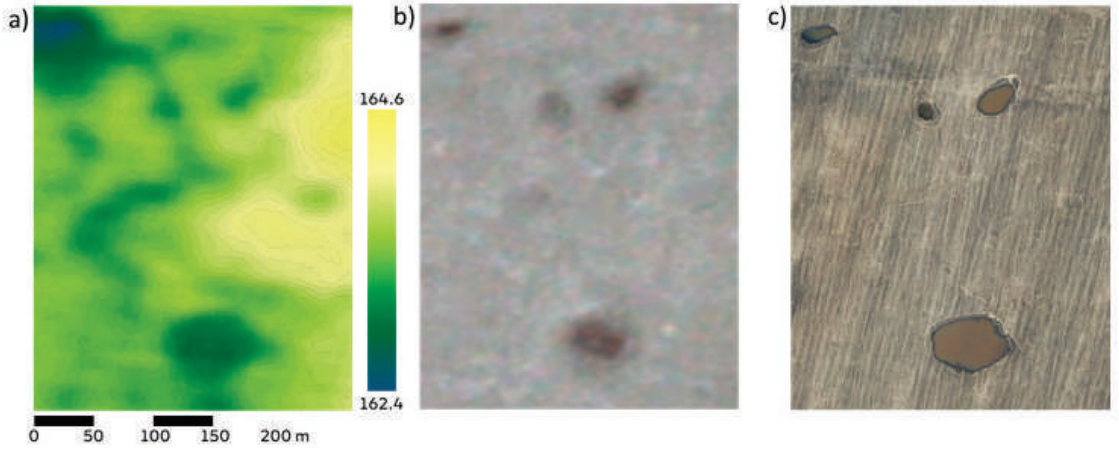


Figure 2. Sources of spatial data. (**a**) High-resolution DEM obtained photogrammetrically with colors indicating heights. (**b**) Planet Labs' digital images of terrain in the visible spectrum with 3 m resolution. (**c**) High-resolution digital image obtained using the UAV in the visible spectral range.

In this study we used stereophotogrammetry, i.e., estimation of three-dimensional coordinates of points on an object from two or more photographic images taken from different positions by the small UAV. In this article, we used a standard method for constructing a digital terrain model using a small UAV with an accuracy of 0.03 m (hereafter, the UAV DEM).

For this, we used the Mavic 2 Pro routing app (DroneDeploy.com). Geolocation markers were located on the ground, and their positions were determined using the STONEX GNSS system (flight altitude, 150 m; image overlap, 75%). We processed the data using Agisoft Metashape and created a dense point cloud to generate a digital terrain model. We manually filtered points associated with agroforestry areas within the fields in ArcGis Pro using the field mask and the vegetation mask. The masks were obtained by manual decoding the UAV materials. Orthophoto maps generated from free satellite photos obtained via Bing were used to identify trees. Points related to heights of the trees were removed. A digital model of the territory with 25 cm resolution was created from the remaining point cloud using the kriging interpolation tool in ArcMap. An orthomosaic was created in Agisoft Metashape and exported with 25 cm resolution.

A time series of high-resolution visible orthomosaics (with 1–3 m resolution) at times when there was no cloud cover, from the beginning of spring snowmelt to the drying up of temporary water bodies in early summer were downloaded from Planet Labs Inc. Images of the scenes were downloaded when there was no cloud cover, from the beginning of spring snowmelt to the drying up of temporary water bodies in early summer.

2.2.1. Delineation of Water Bodies

Orthophotomaps generated using an UAV allow correct interpretation of water surfaces, as they can be visually inspected to delineate water/dry surface boundaries accurately. Orthomosaic maps from Planet Labs have lower resolution and higher atmospheric noise. Therefore, we used the Interactive Supervised Classification tool in the ArcGIS Pro desktop application to delineate water bodies in them. For this we created a water feature training set from the ultra-high resolution orthomosaic, and used it to enable automatic recognition of water bodies in the Planet Labs orthomosaics via object-oriented supervised classification, as implemented in the ArcGis Pro raster classification tool [35]. It is well established that object-oriented classification is superior to pixel-based classification for high-resolution images [36], and it has been previously used to delineate similar depression-shaped natural systems [37,38].

The classification involved the following steps. First, the analyzed raster layer was constrained by a field cadastral border [39] buffered 15 m on each side to prevent inclusion of objects rather than a bare soil surface without water (e.g., an agricultural field with no vegetation in early spring) and surfaces that may be flooded with water. Masking was applied to avoid possible classification errors by excluding unnecessary objects (trees, roads, buildings, etc.). The second step was imaging segmentation, based on a mean shift procedure, by criteria of the minimum segment size expressed in pixels [40,41] implemented in ArcGIS Pro, to merge adjacent pixels of relative homogeneity—preferentially based on spectral (color) characteristics—into image objects. Unitless segmentation scale parameters determining the average size of objects governing the degree of homogeneity allowed for pixel merging was set to 10 on the RGB scale. The third step was creation of a training set. As summer approaches, ponds in the depressions always shrink (Figure 3). Thus, water surfaces present on the date of a UAV flight were always present on the preceding dates, and three training samples were created for groups of dates before each UAV survey (Figure 3). Each training sample contained two categories: water and soil surface. Finally, the random forest (RF) method [42,43] and support vector machine (SVM) [44] for supervised classification of segmented images was applied, yielding a binary (water–not water) raster. The testing set from the next UAV survey was used to validate the resulting binary models. This enabled identification of the water surface areas in each period. However, in the classified images, the boundary of the ponds does not have a constant height relative to the DEM of the UAV. To avoid this unnatural variation of height the classified raster was transformed into a vector containing only water polygons. Along the outer boundary of the water polygon, the DEM values of the UAV were sampled with a frequency of 25 centimeters. The median was calculated from the extracted values. The contour was then drawn for the second time, now in accordance with the average value on the UAV DEM, thus, the outer boundaries of the ponds were forced to have a constant height value. The described procedure allowed us to avoid misclassification of the Planet Labs mixed pixels due to relatively low resolution as we operated with the vector area, not the raster area. For the comparison purely pixel-based classification was also made. The DEM and water polygons vector were used to calculate the volume of water in each depression on the days the images were taken. We used the Surface-Volume tool from ArcGIS Pro to calculate the area and volume between the surface and the reference plane (Polygon Volume (3D Analyst). This provided the water content in each depression in cubic meters on each of the days.

The maximum surface water area of a depression corresponds to the volume of water up to its overflow point, defined here as the minimum value of the height in the UAV DEM along its drainage basin vector boundary [45]. The watershed boundary was defined by the Basin tool in ArcGis based on the raster of the flow direction, derived from the UAV DEM using the "Direction of flow" tool in ArcGis Pro. The water layer (mm) in the catchment area of each depression required for its maximum volume is equal to its total maximum volume of water divided by its entire catchment area.

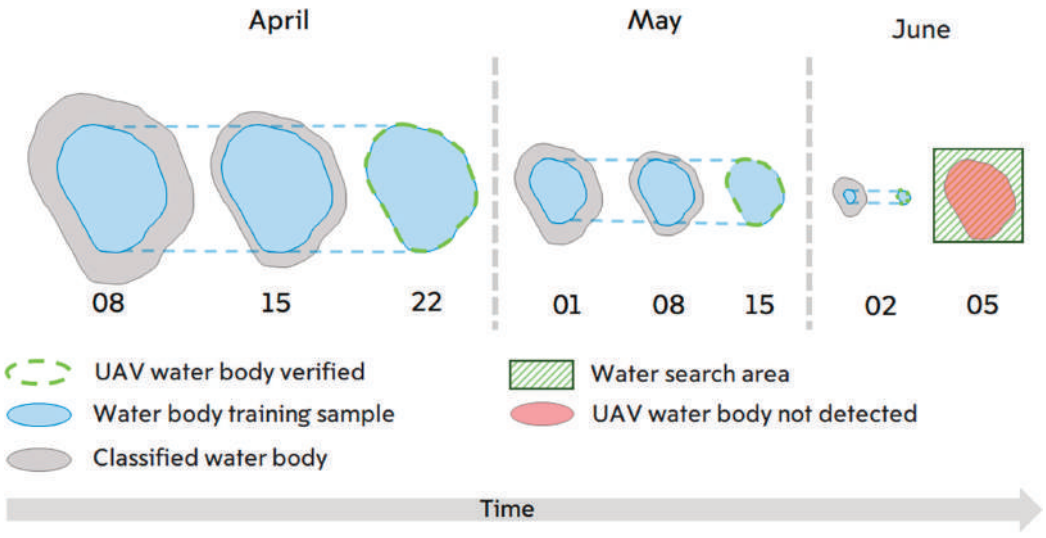


Figure 3. Schematic illustration of the water surface classification method. The arrow at the bottom indicates the general trend of depressions drying (from left to right) in spring. The green dashed lines around the blue areas show the extent of the water as photographed by the UAV on the control dates (22 April, 15 May and 2 June). The gray color indicates the water surface resulting from object-oriented image classification trained on a subset of points from the later control surveys by the UAV. Three time intervals were used for the training. The green shading at the end is the area searched for a water mirror surface by the UAV on a day when the pond had already disappeared (light red fill).

2.2.2. Evaporation

Results of a previous comparison suggest that all of three conventional methods for estimating evapotranspiration from water-filled and vegetated depressions have acceptable applicability for estimating evaporation from open water [46]. The most convenient of these methods, the classical form of the Penman equation [47,48], was used in this study to estimate potential evaporation:

$$E_{PEN} = \frac{\Delta}{\Delta+\gamma}\,\frac{R_n}{\lambda}\,\frac{\gamma}{\gamma+\Delta}\,\frac{6.43E_A}{\lambda}, \quad (1a)$$

Here: E_{PEN} is potential (open water) evaporation (mm/d); R_n is net radiation at the surface (MJ/m^2/d); Δ is the slope of the saturation vapor pressure curve (kPa/C); γ is a psychrometric coefficient (kPa/°C); λ is the latent heat of vaporization (MJ/kg); and E_A is the drying power of the air, which can be found using the following Dalton-type formulation:

$$E_A = f(U)D = (1 + 0.536U)(e_s - e_a) \quad (1b)$$

Here: $f(U)$ is a wind function with linear coefficients for the original Penman equation (1948, 1963); u is the wind speed at 2 m height (m/s), $D = (e_s - e_a)$ is vapor pressure deficit (kPa); e_S is saturation vapor pressure (kPa); and e_a is actual vapor pressure (kPa).

Open-water evaporation was computed from readily available data as previously described [49] and implemented in the Evaplib Python library [50]. Input data for this were air temperature (T, °C), solar radiation (RS, MJ/m^2/d), relative humidity (RH, %), and wind velocity (u, m/s).

In the absence of actinometric measurements of net radiation at the surface, this was calculated from amounts of cloud cover recorded at the weather station and a previous regional calibration [51].

$$R_a = \frac{24(60)}{\pi} G_{sc} d_r (\omega_s \sin(\phi) \sin(\delta) + \cos(\phi) \cos(\delta) \sin(\omega_s)) \quad (2a)$$

Here: R_a is extraterrestrial radiation (MJ/m^2/day**), G_{sc} is the solar constant (0.0820 MJ/m^2/min), d_r is the inverse of the relative distance between the Earth and Sun, ω_s is the sunset hour angle (rad), ϕ is latitude (rad), and δ is solar declination (rad).

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi}{365} J\right) \quad (2b)$$

where J is the day of the year; $\delta = 0.409 \sin\left(\frac{2\pi}{365} J - 1.39\right)$;

$$\omega_s = \arccos(-\tan(\phi) \tan(\delta)) \; ; N = \frac{24}{\pi} \omega_s$$

Solar radiation, R_s, can be calculated from the amount of cloud:

$$R_s = (a_s + b_s(1 - N)) R_a \quad (3)$$

where N is the amount of cloud (ranging from 0 for clear sky to 1 for full cloud cover), while a_s and b_s are Angstrom values, and without regional calibration values of 0.25 and 0.50, respectively, are recommended [52].

Net longwave radiation (R_{nl}) can be estimated from the air temperature, actual vapor pressure, and solar radiation. Net longwave radiation is expressed by the Stefan–Boltzmann law:

$$R_{nl} = \sigma \left(\frac{T}{max^2_{4^4_{min}}} (0.34 - 0.14\sqrt{e_a}) \left(1.35 \frac{R_s}{R_{so}} - 0.35\right) \right) \quad (4)$$

where T_{max} is daily maximum air temperature (K), T_{min} is daily minimum air temperature (K), and R_{so} is clear-sky radiation (MJ/m^2/day) according to:

$$R_{so} = (a_s + b_s) R_a \quad (5)$$

We applied a constant albedo of 7% (0.07) for water surfaces in the calculations, based on the latitude and published mean reference values [51].

$$R_n = (1 - \alpha) R_s - R_{nl} \quad (6)$$

We calculated daily evaporation values. Input data for Equations (1a), (1b) and (4) and daily precipitation were obtained from the nearest meteorological station (at Lipetsk city).

2.2.3. Water Balance and Groundwater Model Recalibration

Infiltration rates (mm/day) were calculated from the daily water balance equation:

$$F = 1000 \cdot [-\Delta V - A_t (E_{PEN} - P)] / A_t \quad (7)$$

where $-\Delta V$ is the daily rate of reduction in pond volume (m^3), A_t is the current pond area (m^2), and P is the daily precipitation (mm/d).

The volume of a pond on a given Julian day (*dayT*) was derived from the volume on the first Julian day in a series (*dayF*) and the following negative exponential equation:

$$V = a \cdot e^{-c \cdot (dayT - dayF)} \quad (8)$$

The scaling coefficient a and the power of the exponent c (the pond's approximate initial volume and decay rate, respectively) were obtained by the least square method, which has given a fit with $R^2 > 0.9$ for each of studied lakes (Table 2, Figure 4d). The estimated volume on each day was used to calculate the rate of reduction in pond volume in Equation (7) with a daily time step.

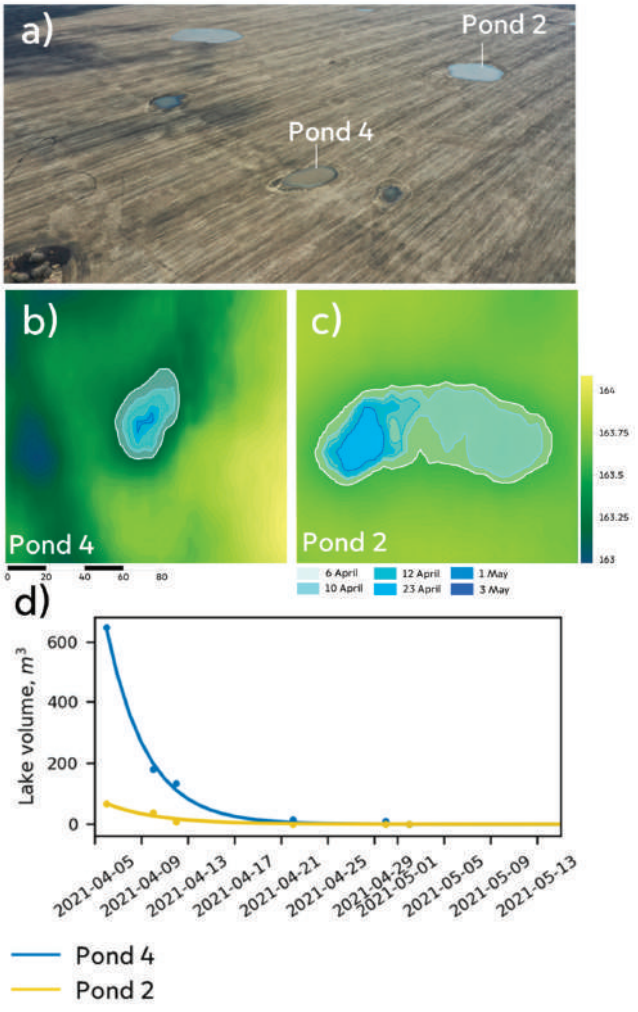


Figure 4. Temporal dynamics of selected ponds in the depressions. (**a**) Positions of Ponds 4 and 2 in a UAV photo image. (**b**,**c**) Groundwater levels in basins of the ponds on indicated dates based on the classification of images (highlighted in color by day). (**d**) Water volumes in the basins of ponds 4 and 2 on indicated dates (points) and negative exponential fits derived by Equation (8) (lines). Pond numbering as in Figure 1 and Table 2.

A limitation of the method lies in the choice of the first day of infiltration, because it is impossible to determine the water boundary in depressions when they are covered with snow. Water infiltrates the soil when the temperature is already above 0 degrees Celsius, but classification of images with partial snow cover is problematic. Thus, the first Planet Labs image that was subjected to classification was the first when there was no snow cover according to the nearest (Lipetsk) weather station.

A 2D profile of the steady-state shallow water table depth was obtained by the analytical form of continuity equation with calibration based on soil redoximorphic features. In [22] the hypothesized relationship between archived morphological properties (redoximorphic features as indicators of gleyic conditions) of soils and a current hydrological process indicator (WTD) were established based on the expert knowledge of soil types, WTD co-occurrence, then verified under a hillslope flow continuity constraint expressed mathematically as a steady-state solution with two free parameters: hydraulic conductivity and recharge rate. Here the input horizontal transect of groundwater recharge rate was taken as a time integral of Equation (7). Spatially, it varied along the transect according to positions of the depressions in the landscape. Infiltration into the soil is not equal to the groundwater recharge rate, so relative values in the $[0, 1]$ interval were used to describe the variability along the transect while the formal calibration of the absolute values of recharge rate to hydraulic conductivity (N/k) was preserved in the method. Time-averaged infiltration was calculated based on the volume of water that infiltrated in closed depressions. We established 10 regular topographic profiles representing the generalized transect, 3 km long and crossing the interfluve along the main slope with regularly (5 m) spaced points. At each standard point, the value of the water layer (mm) was extracted, which was filtered out in a closed depression. The 10 topographic lines were combined into a single profile by averaging values corresponding to the order of the points of the water layer. The regular placement of topographic profiles and sampling points was intended to optimize the two-dimensional characterization of additional moisture infiltration along the studied transect.

3. Results

The proposed combination of object-oriented image classification based on a time series of Planet Labs images and an orthomosaic derived from UAV surveys to verify the satellite data enabled highly accurate identification of the water mirrors of closed depressions during their drying (Cohen's kappa = 0.99). Moreover, high-precision digital terrain models obtained using UAVs can be used to calculate volumes of water in closed depressions.

We compared different methods of pond extraction for the scene on 9 April 2021, when the reference UAV image was obtained. Two supervised pixel-based classification methods were compared: random forest (RF) and support vector machines (SVM) providing results as a raster. Then the ponds boundaries were brought to a constant median value of the DEM to obtain vector pond polygons (also both from RF and SVM classification). Root mean square error (RMSE) and mean absolute percentage error (MAPE) of pond volume and area were the lowest for the vector approach and notably higher for raster approach (Table 1). SVM and RF errors were almost the same within the vector approach (Table 1), and it was decided to use RF as the most common in such studies. Contrary to [52], the novel way to vectorize the polygons based on idea of flat pond mirror with constant height brought a very notable increase in the quality of area and volume estimate.

Table 1. Root mean square error (RMSE) and mean absolute percentage error (MAPE) of the area and volume with true value taken from ultra-high resolution UAV estimate of ponds' boundaries. Random forest (RF) and support vector machine (SVM) methods are compared for the pixel-based (raster) and median DEM height-based (vector) delineation of the image taken on 9 April 2021.

	Errors of Area Estimate		**Errors of Volume Estimate**	
	RMSE, m^2	**MAPE, %**	**RMSE, m^3**	**MAPE, %**
RF vector	7.8	1.4	2.5	5.9
RF raster	43.2	3.1	16.0	21.9
SVM vector	8.7	1.6	2.1	5.1
SVM raster	44.9	3.6	16.3	22.2

Results obtained using the described procedure show that the drainage process of the focal depressions follows an exponential equation (Figure 4, Table 2), with coefficients (Table 2) that presumably depend on various factors (e.g., the depressions' source rocks and filtration areas), but we found no systematic quantitative relationships between the coefficients and considered parameters.

Table 2. Derived characteristics of the ponds during the decreasing volume phase after snowmelt from the maximum (starting day) to zero (final day). Coefficients a and c are from Equation (8) and R^2 is the coefficient of determination for the negative exponential fit of the lake volume by the least square method.

Pond No. [1]	Maximum Volume, m^3	Maximum Area, m^2	Coefficient *a*	Coefficient *c*	R^2	Total Infiltration, m^3	Initial Infiltration Rate, m^3/d	Soil Refill Capacity, mm/dd [2]	ksat mm/d	Total Evaporation, m^3	Total Precipitation, m^3
1	640	6123	604	0.11	0.91	475	47	54	3	320	155
2	647	4275	645	0.29	1.00	594	153	163	3	113	59
3	173	1756	173	0.38	0.98	167	50	172	12	18	12
4	66	1283	67	0.23	0.95	56	10	53	4	24	14
5	416	3546	428	0.19	0.94	366	64	103	6	103	54
6	16	540	16	0.53	1.00	14	5	37	1	6	4
7	103	978	104	0.28	1.00	92	23	113	2	24	13
8	36	543	37	0.34	0.97	34	9	93	4	7	5
9	12	295	13	0.19	0.99	9	1	34	2	7	4

[1] The numbering follows Figure 1. [2] dd—melt water peak event duration (days).

During the initial phase the rate of pond recession is much higher than later in the season (Figure 5 top). Notably less water is evaporated than infiltrates (Figure 5, middle and bottom), so the depression-focused replenishment of the groundwater is consistent with the previously mentioned hypothesis that the major source of recharge for shallow groundwater in the study area (and similar areas) is depression-focused infiltration during snowmelt [21]. There are two phases of infiltration—fast and slow (Figure 5, bottom). Measurements during the fast phase enable estimation of the unsaturated soil's refill rate and capacity (Table 2). During the slow phase the change in infiltration rate from day to day is much smaller. The saturated hydraulic conductivity decreases strongly with depth under a depression [8], reflecting the effects of the decreasing frequency of fractures with depth, and the flow is presumably limited by the lowest layer with the smallest frequency. Thus, the infiltration rate estimated during the slow phase provides an approximation of the hydraulic conductivity (Table 2), corresponding to the maximum possible flux out of the soil column.

Overflow can occur from any closed depression (Figure 6). The probability of spillage depends on multiple factors, including elevations of the lowest point in the catchment area and the depression's overflow point. In 2018, water reached the overflow point in almost all the depressions considered here (Figure 6). Thus, the initial volumes (*a* coefficients) obtained for the nine studied ponds can be used in speculation regarding the effects of the landscape morphometry and meltwater input on initial volumes of ponds after snowmelt.

The results also indicate that the hypothesis of a quantitative linear relationship between the volume of water accumulated in a depression and the catchment area of the basin is only partly correct. The water volumes do not appear to be linearly related to the depressions' catchment, because the amount of water in a depression depends on the catchment area and maximum volume that can be stored in it. Excess water will flow through the overflow point without replenishing the water table. Limits of the possible volume and layer of water intercepted by the focal depressions, which limit their ability to converge surface runoff into underground flows, were identified. The maximal layer depends on the catchment area of the depression and height of its overflow (Figure 6, right).

The water layer filling closed depressions during snowmelt in the forest-steppe zone is highly dynamic. From 2005 to 2021, the snowmelt water layer (snow water equivalent, SWE), reconstructed from the statistically corrected snow height and snow density data series, varied from 50 to 300 mm. The derived snowmelt water layer during this period has a binomial distribution with two maxima, at 50 and 200 mm SWE. Analysis of the

meteorological data showed that closed depressions did not overflow during snowmelt in 60% of cases, on average, from 2005 to 2021. This corroborates the finding that in most cases closed depressions intercept the surface runoff and transfer it to groundwater. Frequencies of overflow were lowest for Ponds 2 and 4 (around 10%) and highest for Pond 9 (90%).

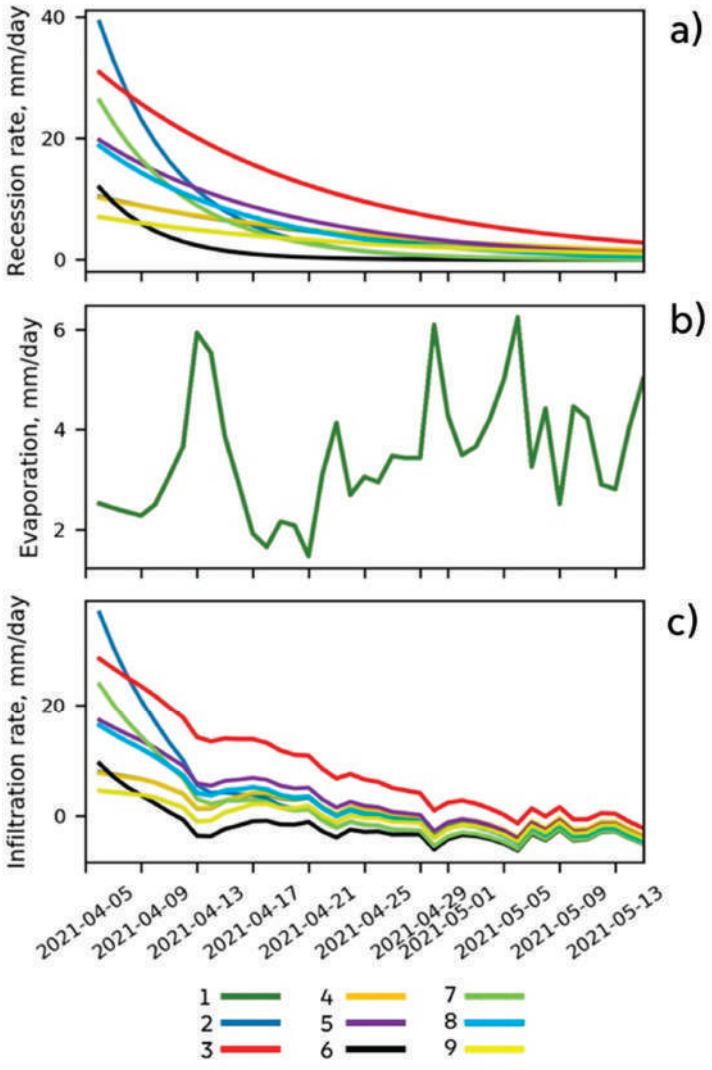


Figure 5. Rate of recession of the indicated ponds' water levels (**a**), daily evaporation rate (**b**) and infiltration rates of the ponds estimated from the mass balance (bottom panel) expressed in mm of the water layer (**c**). The numbering of the ponds in the color legend follows Figure 1 and Table 2.

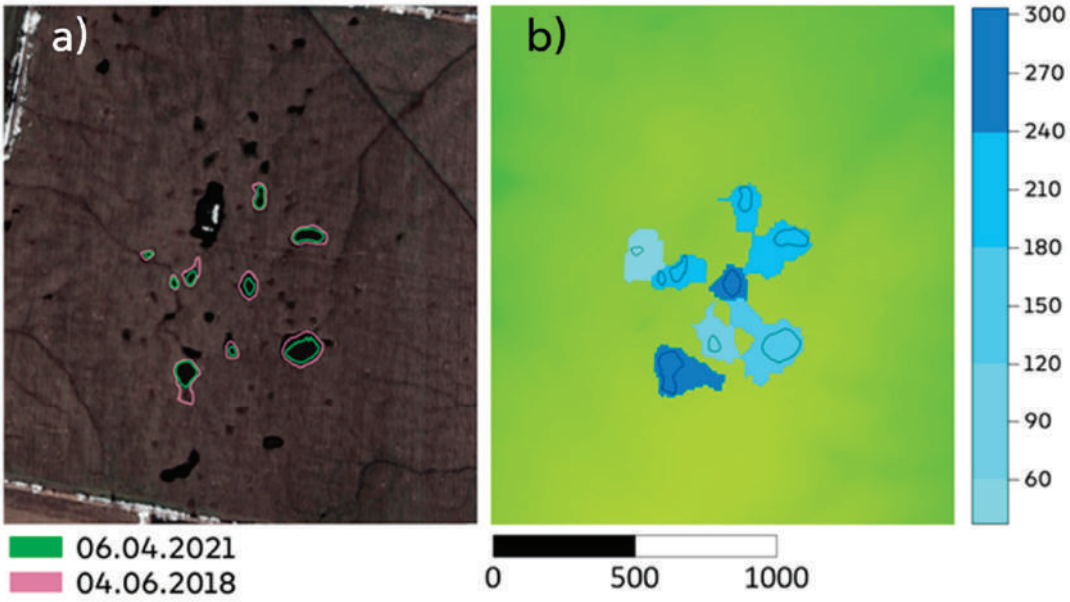


Figure 6. (**a**) Boundaries of the maximum potential filling of closed depressions (up to their overflow heights) shown in white in a Planet Labs image from 4 April 2018, with color shading and magenta contours, and the maximum pond boundaries in 2021 from 6 April 2021 (green contours). In 2018, the investigated closed depressions were overflowing. (**b**) Shades of blue indicating the layers of water (in mm) that must enter the depressions from their catchment areas to completely fill them.

Figure 7 illustrates the simulations of the shallow groundwater level for cases with the recharge rate either constant or spatially varied along the transect. The parameter N/k was restricted by the requirement for correspondence between the simulated WTD and range of WTD for soils of each type from expert knowledge (Table 2 in [22]) in distance intervals across the catena's whole toposequence. For example, if very poorly drained soils (under depression bed) are present in M unit intervals, those in which WTD > 3 m (too deep) were counted with 0 weight and the others with 1 weight. The same procedure was then applied for each of the intervals with the other soil types, then the sums were added for all groups and scaled to the total number of unit intervals in the catena toposequence to obtain the accuracy in percent. Simulation of WTD was successful for the generalized transect in terms of correspondence between the simulated WTD and ranges of WTD obtained from the indirect soil indicators (redoximorphic features) and expert knowledge both in the cases of constant and spatially variable recharge. However, the required accuracy threshold was set at 97%, and was met for the spatially variable recharge. A significantly lower threshold (80%) was satisfied for the constant recharge case. Therefore, the method to estimate depression-focused infiltration proposed here can make the shape of the water table profile more realistic.

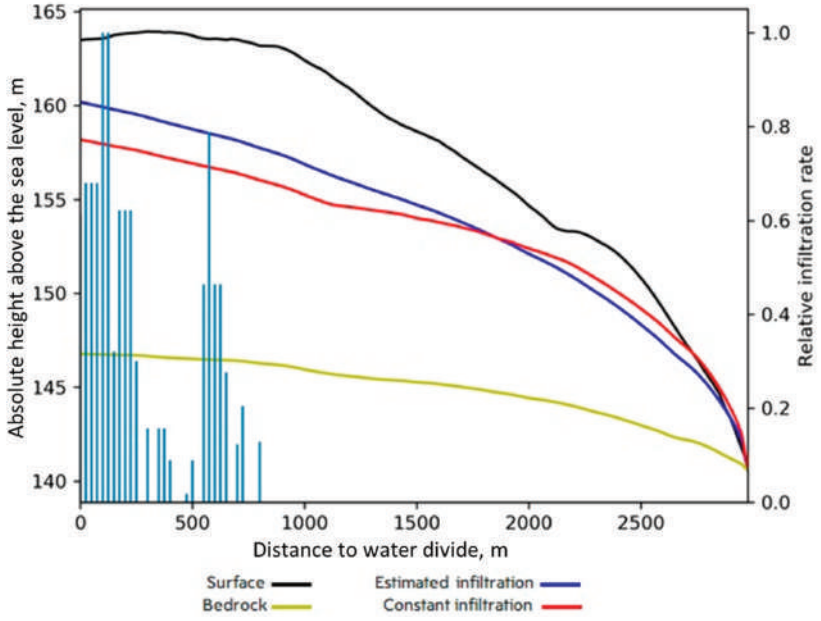


Figure 7. Cross-section of the catena with water table depth (WTD, blue and red lines) adjusted to correspond to soil group [22]. The yellow and black lines indicate the position of the bedrock and DEM profile, respectively. The blue and red lines respectively indicate WTD obtained with infiltration along the transects estimated from the depressions' positions and specific infiltration rates (as indicated by the blue bars and right y axis) and constant infiltration along the transect.

4. Discussion

The word steppe is usually associated with the Russian plains, but the northern part of this ecoregion has notable similarity to North American prairies. The lithological and geomorphological similarity of the Tambov region to the Saskatchewan and Alberta provinces in Canada enables direct comparison of the depression-focused infiltration into their soils through temporal ponds that are very similar in size distribution and shape. The recession rate of the ponds after snowmelt obtained in this study is similar to that derived from an artificial flooding experiment in the C24 depression, northwest of Calgary, Alberta, Canada, in 2004 [8]. As in the cited study [8] and another previous investigation [4], we found that evaporation accounts for a much smaller proportion of the pond water balance loss term than infiltration into the soil. The pre-event pore space available for filling with infiltration water was not directly measured in this study. However, data from a depression monitoring site in the study region in the years 2003–2005 show a spread of 30–400 mm of water deficit to saturation. An assumption underlying our two-stage infiltration conceptual model is that pores of the soils below the bottom of a pond are all filled to saturation during the first stage down to the shallow groundwater depth (approximately 2 m). Thus, the inflow is restricted by the bottleneck hydraulic conductivity below this point, which is an order of magnitude lower than in the upper soil layers [8] and also by the gradual rise of the water table when two fronts of water are jointing. The soil refill amount of 34 to 172 mm recorded in Table 2 fits well into this range. There is also similarity with the refill amount (148 mm) obtained in the cited Canadian study [8]. A strength of our study is that the infiltration rate was estimated for nine ponds, not just one pond such as the well-studied experimental pond C24. The variation (four-fold) in infiltration between those ponds (Table 2) could not be explained by the pond size or topographical settings. Thus, it is not sufficient to apply infiltration data from one pond to other ponds as this leads to large errors. The differences are likely due to diverse factors, inter alia physical properties

of the soil associated with their lithological and textural characteristics, the thawing rate and ice content, abundance of root channels and other pathways for preferential flow. We conclude that there is no straightforward analytical way to characterize this spatial variability, but use of data obtained by the methods proposed here in conjunction with appropriate hydrological models and high-resolution satellite images is highly promising.

Here, we used the steady-state continuity equation in kinematic wave form parameterized using expert knowledge of the links between typical water table depth (WTD) and redoximorphic features of soils with different hydromorphy degrees [22]. In this simulation, we were able to account for variation in infiltration rates in the catena using real data on depressions' positions within the transect. However, calibration was still necessary because the infiltration and recharge are split in time by unsaturated zone processes. In future research, we plan to develop a model conceptually similar to the VSMB Depression-Upland System (VSMB-DUS) model [8] using data acquired in investigations of the surface water–groundwater interaction in individual depressions and their catchments. The planned model will be based on the watershed hydrological WASA-SED model [53], which already discretizes focal watersheds into hierarchical levels (sub basins, land units, terrestrial components, soil-vegetation components). Land units are representative catenas and terrestrial components can be easily supplemented with depressions and uplands providing surface flow to them by an already incorporated horizontal flow mechanism. For the terrestrial components prescribed as depressions, the temporally varying fluxes obtained by the method developed here will be used as upper boundary conditions. Collection of field data is planned to obtain saturated hydraulic conductivity values for the vertical levels besides the bottom soil layer. Groundwater depth measurements will provide calibration for the drainage rates from the deepest soil layer and validation for the dynamic version of the WASA-SED shallow groundwater flow sub-model. In this manner, groundwater recharge will fully account for the spatial variability of depression density, such as prevailing areas of numerous depressions at the water divide.

In this study, we derived the saturated hydraulic conductivity, not for a single point, but aggregated for the area of depressions. Most grid data used represent points, but landscape-level data are essential inputs for a hydrological model. A hypothesis under test is that soil hydraulic properties are related to landscape position and topography [54]. If so, elucidation of these relationships could greatly enhance pedotransfer functions for estimating saturated hydraulic conductivities at the level of land units and terrestrial components, not just points. Our study, based on remote sensing, provides an example of such derivation because the hydraulic conductivity is based on the depressions' water balance accounting for their positions in the landscape.

A limitation of this study lies in the assumption that all snowmelt runoff from the upland was routed to the depressions before the initial day of the study, and water volume within each depression exceeding its maximum storage capacity overflowed directly into surface runoff with no contribution to infiltration into the soil. However, it is widely acknowledged that depressions tend to form fill-spill networks, where overflow from one depression feeds an adjacent depression [5,6]. This process can be modeled [6,10,15–17], but studies of fill-spill processes have primarily focused on effects of depression storage on surface flow to streams rather than depression-focused groundwater recharge. Visual observations during the hydrological phase after the most active snowmelt showed no signs of connectivity between depressions at our study site, but that was typical for the active snowmelt phase itself of about a week duration. We justify the restriction of our approach with the hypothesis that non-stationary volumes of the depression ponds when snow is still present contribute little to total infiltration, partly due to the frozen state of the soil.

5. Conclusions

Estimation of infiltration through ponds is an important step toward the challenging goal to estimate depression-focused recharge of groundwater, and thus evaluate this

important resource, in the forest-steppe zone of Russia. Using high-resolution Planet Labs images and widely evaluated tools for object-based image recognition, we have developed a relatively simple method to reconstruct a time series of infiltration into the soil under ponds and estimate landscape-scale saturated hydraulic conductivity. The simulation of the steady-state groundwater profile for the topographical transect fed with data on relative water supplies through depressions along the transect was more consistent with observations (based on soil redoximorphic indicators of water level) than the simulation fed with a uniform recharge function. Further development is needed to assimilate the data generated with consideration of the spatial variability of pond infiltration into a process-based model of groundwater recharge that accounts for interactions between depressions and their catchments.

Author Contributions: Conceptualization, A.Y.Y. and P.P.F.; methodology, P.P.F. and A.Y.Y.; software, P.P.F., A.D. and A.Y.Y.; validation, P.P.F. and A.D., formal analysis, A.Y.Y.; investigation, P.P.F. and A.Y.Y.; resources, P.P.F.; data curation, P.P.F.; writing—original draft preparation, P.P.F., A.D. and A.Y.Y.; writing—review and editing, A.Y.Y. and D.K.; visualization, P.P.F.; supervision, D.K.; project administration, D.K.; funding acquisition, D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the RFBR, grant number 19-29-05277.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in Fil, Pavel; Yurova, Alla (2021), "Water bodies recognition for depression-focused recharge. Tambov region, Russia", Mendeley Data, V1, doi:10.17632/pn3gdzhdy4.1.

Acknowledgments: We are grateful to three anonymous reviewers for the useful comments that have helped to improve the manuscript.

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

References

1. de Vries, J.J.; Simmers, I. Groundwater recharge: An overview of process and challenges. *Hydrogeol. J.* **2002**, *10*, 5–17. [CrossRef]
2. Hayashi, M.; Van Der Kamp, G.; Schmidt, R. Focused infiltration of snowmelt water in partially frozen soil under small depressions. *J. Hydrol.* **2003**, *270*, 214–229. [CrossRef]
3. Bam, E.K.P.; Ireson, A.M.; van der Kamp, G.; Hendry, J.M. Ephemeral Ponds: Are They the Dominant Source of Depression-Focused Groundwater Recharge? *Water Resour. Res.* **2020**, *56*, e2019WR026640. [CrossRef]
4. Pavlovskii, I. *Groundwater Recharge in the Canadian Prairies: Mechanisms, Constraints, and Rates*; University of Calgary: Calgary, AB, Canada, 2019.
5. Shook, K.; Pomeroy, J.W.; Spence, C.; Boychuk, L. Storage dynamics simulations in prairie wetland hydrology models: Evaluation and parameterization. *Hydrol. Process.* **2013**, *27*, 1875–1889. [CrossRef]
6. Shook, K.; Papalexiou, S.; Pomeroy, J.W. Quantifying the effects of Prairie depressional storage complexes on drainage basin connectivity. *J. Hydrol.* **2021**, *593*, 125846. [CrossRef]
7. Ahmed, M.I.; Elshorbagy, A.; Pietroniro, A. A novel model for storage dynamics simulation and inundation mapping in the prairies. *Environ. Model. Softw.* **2020**, *133*, 104850. [CrossRef]
8. Noorduijn, S.L.; Hayashi, M.; Mohammed, G.A.; Mohammed, A.A. A coupled soil water balance model for simulating depression-focused groundwater recharge. *Vadose Zo. J.* **2018**, *17*, 170176. [CrossRef]
9. Muhammad, A.; Evenson, G.R.; Stadnyk, T.A.; Boluwade, A.; Jha, S.K.; Coulibaly, P. Impact of model structure on the accuracy of hydrological modeling of a Canadian Prairie watershed. *J. Hydrol. Reg. Stud.* **2019**, *21*, 40–56. [CrossRef]
10. Kiss, J. *Predictive Mapping of Wetland Types and Associated Soils through Digital Elevation Model Analyses in the Canadian Prairie Pothole Region*; University of Saskatchewan: Saskatoon, SK, Canada, 2018.
11. Shaw, D.A.; Pietroniro, A.; Martz, L.W. Topographic analysis for the prairie pothole region of Western Canada. *Hydrol. Process.* **2013**, *27*, 3105–3114. [CrossRef]
12. Wang, N.; Chu, X. A new algorithm for delineation of surface depressions and channels. *Water* **2020**, *12*, 7. [CrossRef]
13. Wu, Q.; Lane, C.R. Delineating wetland catchments and modeling hydrologic connectivity using lidar data and aerial imagery. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 3579–3595. [CrossRef]

14. Wu, Q.; Lane, C.R.; Wang, L.; Vanderhoof, M.K.; Christensen, J.R.; Liu, H. Efficient Delineation of Nested Depression Hierarchy in Digital Elevation Models for Hydrological Analysis Using Level-Set Method. *J. Am. Water Resour. Assoc.* **2019**, *55*, 354–368. [CrossRef]
15. Brooks, J.R.; Mushet, D.M.; Vanderhoof, M.K.; Leibowitz, S.G.; Christensen, J.R.; Neff, B.P.; Rosenberry, D.O.; Rugh, W.D.; Alexander, L.C. Estimating Wetland Connectivity to Streams in the Prairie Pothole Region: An Isotopic and Remote Sensing Approach. *Water Resour. Res.* **2018**, *54*, 955–977. [CrossRef] [PubMed]
16. McDonough, O.T.; Lang, M.W.; Hosen, J.D.; Palmer, M.A. Surface Hydrologic Connectivity Between Delmarva Bay Wetlands and Nearby Streams Along a Gradient of Agricultural Alteration. *Wetlands* **2015**, *35*, 41–53. [CrossRef]
17. McLaughlin, D.L.; Kaplan, D.A.; Cohen, M.J. A significant nexus: Geographically isolated wetlands influence landscape hydrology. *Water Resour. Res.* **2014**, *50*, 7153–7166. [CrossRef]
18. Rover, J.; Wright, C.K.; Euliss, N.H.; Mushet, D.M.; Wylie, B.K. Classifying the hydrologic function of prairie potholes with remote sensing and GIS. *Wetlands* **2011**, *31*, 319–327. [CrossRef]
19. Vanderhoof, M.K.; Lane, C.R. The potential role of very high-resolution imagery to characterise lake, wetland and stream systems across the Prairie Pothole Region, United States. *Int. J. Remote Sens.* **2019**, *40*, 5768–5798. [CrossRef]
20. Tiner, R.W. Geographically isolated wetlands of the United States. *Wetlands* **2003**, *23*, 494–516. [CrossRef]
21. Rode, A.A. *The Water Regime of Soils and Its Regulation*; Academy of USSR Publishing: Moscow, Russia, 1963.
22. Yurova, A.Y.; Smirnova, M.A.; Lozbenev, N.I.; Fil, P.P.; Kozlov, D.N. Using soil hydromorphy degree for adjusting steady-state water table simulations along catenas in semiarid Russia. *Catena* **2021**, *199*, 105109. [CrossRef]
23. Jawak, S.D.; Kulkarni, K.; Luis, A.J. A Review on Extraction of Lakes from Remotely Sensed Optical Satellite Data with a Special Focus on Cryospheric Lakes. *Adv. Remote Sens.* **2015**, *4*, 196. [CrossRef]
24. Montgomery, J.; Mahoney, C.; Brisco, B.; Boychuk, L.; Cobbaert, D.; Hopkinson, C. Remote Sensing of Wetlands in the Prairie Pothole Region of North America. *Remote Sens.* **2021**, *13*, 3878. [CrossRef]
25. Yuan, C.; Gong, P.; Bai, Y. Performance Assessment of ICESat-2 Laser Altimeter Data for Water-Level Measurement over Lakes and Reservoirs in China. *Remote Sens.* **2020**, *12*, 770. [CrossRef]
26. Zhang, Y.; Pulliainen, J.T.; Koponen, S.S.; Hallikainen, M.T. Water quality retrievals from combined Landsat TM data and ERS-2 SAR data in the Gulf of Finland. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 622–629. [CrossRef]
27. Verpoorter, C.; Kutser, T.; Tranvik, L. Automated mapping of water bodies using Landsat multispectral data. *Limnol. Oceanogr. Methods* **2012**, *10*, 1037–1050. [CrossRef]
28. Huang, W.; DeVries, B.; Huang, C.; Lang, M.; Jones, J.; Creed, I.; Carroll, M. Automated Extraction of Surface Water Extent from Sentinel-1 Data. *Remote Sens.* **2018**, *10*, 797. [CrossRef]
29. Martin, R.; Brabyn, L.; Beard, C. Effects of class granularity and cofactors on the performance of unsupervised classification of wetlands using multi-spectral aerial photography. *J. Spat. Sci.* **2014**, *59*, 269–282. [CrossRef]
30. Talukdar, S.; Mankotia, S.; Shamimuzzaman, M.; Mahato, S. Wetland-Inundated Area Modeling and Monitoring Using Supervised and Machine Learning Classifiers. In *Advances in Remote Sensing for Natural Resource Monitoring*; Wiley: Hoboken, NJ, USA, 2021.
31. Tulbure, M.G.; Broich, M. Spatiotemporal dynamic of surface water bodies using Landsat time-series data from 1999 to 2011. *ISPRS J. Photogramm. Remote Sens.* **2013**, *79*, 44–52. [CrossRef]
32. Mahdianpari, M.; Rezaee, M.; Zhang, Y.; Salehi, B. Wetland Classification Using Deep Convolutional Neural Network. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018.
33. Zhaohui, Z.; Prinet, V.S.M. Water body extraction from multi-source satellite images. In Proceedings of the IGARSS 2003–2003 IEEE International Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003.
34. RIHMI-WDC (Roshydromet) The Database of the Main Meteorological Parameter at the Russian Weather Stations. Available online: http://aisori-m.meteo.ru/ (accessed on 10 September 2021).
35. ArcGis Pro The Image Classification Wizard. Available online: https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/the-image-classification-wizard.htm (accessed on 10 September 2021).
36. Li, M.; Zang, S.; Zhang, B.; Li, S.; Wu, C. A review of remote sensing image classification techniques: The role of Spatio-contextual information. *Eur. J. Remote Sens.* **2014**, *47*, 389–411. [CrossRef]
37. Waz, A.; Creed, I.F. Automated Techniques to Identify Lost and Restorable Wetlands in the Prairie Pothole Region. *Wetlands* **2017**, *37*, 1079–1091. [CrossRef]
38. Gadal, S.; Zakharov, M.; Kamicaityte, J.; Danilov, Y. Alas landscape modeling by remote sensing image analysis and geographic ontology: Study case of central yakutia (Russia). In Proceedings of the GISTAM 2020—6th International Conference on Geographical Information Systems Theory, Applications and Management, Prague, Czechia, 7–9 May 2020; pp. 112–118. [CrossRef]
39. Public Cadastral Map of Russia. Available online: https://pkk.rosreestr.ru (accessed on 10 September 2021).
40. Fukunaga, K.; Hostetler, L.D. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inf. Theory* **1975**, *21*, 32–40. [CrossRef]
41. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]
42. Breiman, L. Random forest. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

43. Belgiu, M.; Drăgu, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
44. Vapnik, V.N. *The Nature of Statistical Learning*; Springer: New York, NY, USA, 1995.
45. ArcGIS Pro Polygon Volume (3D Analyst). Available online: https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/polygon-volume.htm (accessed on 10 September 2021).
46. Armstrong, R.N.; Pomeroy, J.W.; Martz, L.W. Evaluation of three evaporation estimation methods in a Canadian prairie landscape. *Hydrol. Process. Int. J.* **2008**, *22*, 2801–2815. [CrossRef]
47. Penman, H.L. Natural evaporation from open water, hare soil and grass. *Proc. R. Soc. Lond. A. Math. Phys. Sci.* **1948**, *193*, 120–145. [CrossRef] [PubMed]
48. Penman, H.L. Evaporation: An introductory survey. *Netherlands J. Agric. Sci.* **1956**, *4*, 9–29. [CrossRef]
49. Valiantzas, J.D. Simplified versions for the Penman evaporation equation using routine weather data. *J. Hydrol.* **2006**, *331*, 690–702. [CrossRef]
50. Evaplib Python Library. Available online: http://python.hydrology-amsterdam.nl/moduledoc/_modules/evaplib.html (accessed on 10 September 2021).
51. Allen, R.G. FAO Irrigation and Drainage. *Irrig. Drain.* **1998**, *300*, 300.
52. Wade, T.G.; Wickham, J.D.; Nash, M.S.; Neale, A.C.; Riitters, K.H.; Jones, K.B. A Comparison of Vector and Raster GIS Methods for Calculating Landscape Metrics Used in Environmental Assessments. *Photogramm. Eng. Remote Sensing* **2003**, *69*, 1399–1405. [CrossRef]
53. Bronstert, A.; de Araújo, J.C.; Batalla, R.J.; Costa, A.C.; Delgado, J.M.; Francke, T.; Foerster, S.; Guentner, A.; López-Tarazón, J.A.; Mamede, G.L.; et al. Process-based modelling of erosion, sediment transport and reservoir siltation in mesoscale semi-arid catchments. *J. Soils Sediments* **2014**, *14*, 2001–2018. [CrossRef]
54. Van Looy, K.; Bouma, J.; Herbst, M.; Koestel, J.; Minasny, B.; Mishra, U.; Montzka, C.; Nemes, A.; Pachepsky, Y.A.; Padarian, J.; et al. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. *Rev. Geophys.* **2017**, *55*, 1199–1256. [CrossRef]

sensors

MDPI

Article

A Wireless Sensor Network Deployment for Soil Moisture Monitoring in Precision Agriculture

Jaime Lloret *, Sandra Sendra, Laura Garcia and Jose M. Jimenez

Instituto de Investigación Para la Gestión Integrada de Zonas Costeras (IGIC), Universitat Politècnica de València, Paraninf 1, 46730 Valencia, Spain; sansenco@upv.es (S.S.); laugarg2@teleco.upv.es (L.G.); jojiher@dcom.upv.es (J.M.J.)

* Correspondence: jlloret@dcom.upv.es; Tel.: +34-6-0954-9043

Abstract: The use of precision agriculture is becoming more and more necessary to provide food for the world's growing population, as well as to reduce environmental impact and enhance the usage of limited natural resources. One of the main drawbacks that hinder the use of precision agriculture is the cost of technological immersion in the sector. For farmers, it is necessary to provide low-cost and robust systems as well as reliability. Toward this end, this paper presents a wireless sensor network of low-cost sensor nodes for soil moisture that can help farmers optimize the irrigation processes in precision agriculture. Each wireless node is composed of four soil moisture sensors that are able to measure the moisture at different depths. Each sensor is composed of two coils wound onto a plastic pipe. The sensor operation is based on mutual induction between coils that allow monitoring the percentage of water content in the soil. Several prototypes with different features have been tested. The prototype that has offered better results has a winding ratio of 1:2 with 15 and 30 spires working at 93 kHz. We also have developed a specific communication protocol to improve the performance of the whole system. Finally, the wireless network was tested, in a real, cultivated plot of citrus trees, in terms of coverage and received signal strength indicator (RSSI) to check losses due to vegetation.

Keywords: electromagnetic induction; soil moisture; precision agriculture; low cost; water management; Internet of Things (IoT); wireless sensor network

Citation: Lloret, J.; Sendra, S.; Garcia, L.; Jimenez, J.M. A Wireless Sensor Network Deployment for Soil Moisture Monitoring in Precision Agriculture. *Sensors* **2021**, *21*, 7243. https://doi.org/10.3390/s21217243

Academic Editors: Vicenç Puig and Miquel À. Cugueró-Escofet

Received: 26 September 2021
Accepted: 27 October 2021
Published: 30 October 2021

1. Introduction

Given the basic need to provide food to the world's population, it is necessary to introduce technology to the agriculture sector to reduce the environmental impact caused by the crops and to increase the conservation of natural resources, among others [1]. Efficient Irrigation is one of the essential factors to increase the development of sustainable agriculture, especially in arid and semi-arid regions where there are the greatest limitations. Irrigation methods can be classified into three generic categories; these are (1) gravity irrigation, (2) sprinkler irrigation, and (3) drip irrigation. The gravity irrigation system is the oldest method and the least efficient for the conservation of natural resources. However, in order to determine the specific irrigation needs of crops, sensing devices must be deployed to obtain data such as soil moisture.

Precision agriculture is a concept that appeared in the USA in the 1980s. It is a management strategy that allows making decisions to improve farming productivity and to achieve more sustainable activity. It is based on the management of crops by observing, measuring, and acting against the variability of the many factors that affect them. Using Internet of Things (IoT) solutions, the soil where the crops are planted can be monitored to make decisions and perform more effective irrigation. These solutions may include not only the electronic devices deployed in the fields but also the use of vehicles such as drones to support the network [2] and to manage the use of pesticides on the crops [3,4]. However, in crop monitoring tasks, especially in those where fruit trees are grown, it is important to be able to control soil moisture levels accurately. For the correct progress

of a fruit tree, it is necessary to ensure that the roots have the right levels of moisture. High humidity levels can facilitate the proliferation of fungi in the roots and leaves, thus affecting production. However, an extremely low soil moisture level can provoke the soil to crack. causing broken roots and the tree to die. This fact negatively affects the growth of plants and consequently their production.

One of the main drawbacks that hinder the use of precision agriculture is the cost of the sensors and the utilized technology. For farmers that want to use technology on a massive scale, it is necessary to provide low-cost systems to make easier deployments.

The available commercial sensors for soil monitoring use different methods to assess the water content of the soil. The most relevant existing methods for obtaining moisture values from the soil [5] are the gravimetric method, tensiometric method, neutronic method, gamma-ray attenuation method, dielectric method, Wenner or resistive method, and light method infrared [6]. Generally, when low-cost sensors are used to measure soil moisture, conductivity-based sensors are based on the use of two electrodes [1]. These types of sensors have two fundamental disadvantages, lack of reliability, and durability. On the one hand, depending on the type of soil and its salt content, the conductivity measurement can vary even when the amount of water in the soil is maintained. On the other hand, the electrodes must be in contact with the ground, and consequently, they can suffer rapid deterioration. Inductive sensors are also employed to measure soil moisture. However, they do not integrate the system into a sensor node to be able to read the parameters.

The network design is an important aspect to consider as well. Usually, fields are located in remote areas. These areas may not have access to the internet infrastructure and the power grid. Therefore, PA systems should include a form of energy harvesting such as solar panels, and some characteristics of these networks should be considered when designing the deployment of sensing devices [7]. Wireless communications are a good solution because it eliminates the cost and hindrance of deploying cabled networks on extensive areas where machinery is utilized. However, the foliage of the crops affects the quality of the signal, resulting in reduced coverage between the devices. It is therefore necessary to determine the optimal deployment design for the area of interest according to the type of crop, and size of the field. Furthermore, the available protocols may not provide all the functionalities desired for a particular crop and the resources available for the area.

In this paper, we present a group-based wireless sensor network to efficiently irrigate cultivated lands. The network is composed of both actuators and sensor nodes that will collect data from the soil and will activate different irrigation systems as a function of the plot needs. Additionally, we design a new soil moisture sensor able to measure the amount of water content in the root ball of a tree. The design includes the sensor and the power circuit required to generate the bi-phase signals to power the coils. The paper presents the design of the operation algorithm and the message exchange for efficient use of water. Finally, the entire system is tested in a real environment to check the correct operation in terms of soil moisture measurements and network performance.

The rest of the paper is structured as follows. Section 2 presents some previous and related works where soil moisture systems are developed. Section 3 presents an overall description of our proposed sensor as well as the features of the different coils used to develop our soil moisture sensor and the experimental tests performed with the coils. This section also includes the power circuit in charge of generating the required signals as well as the integration of both the sensor and the power circuit with an ESP32 module. Section 4 explains the network operation algorithm and message exchange between nodes. In Section 5, the tests performed in a real environment are shown. Section 6 explains the conclusion and future works.

2. Related Work

In this section, we summarize some previous works related to our proposal. The gap in current solutions for soil moisture monitoring is also identified.

Authors such as Ojha et al. [8] present a study where they analyze the wireless sensor network (WSN) implementations for various agricultural applications. We will look at surveys such as the one presented by Garcia et al. [1], aimed at summarizing the current state of the art regarding smart irrigation systems and schemes for Internet of Things (IoT) irrigation monitoring. This survey includes the review of more than 100 scientific works. Other authors, such as Susha Lekshmi et al. [9], present a review of techniques employed for soil moisture measurement. The authors highlight the limitations of the techniques and the influence of soil parameters. Tumanski [10] describes the use of a coil to develop sensors. The work compares, summarizes, and analyzes coil design methods and frequency properties of the coil as well as the use of coil sensor applications such as magnetic antennas. Jawad et al. [11] describe applications of WSNs in agricultural research, and classify and compare wireless communication protocols, the taxonomy of energy efficiency, and energy harvesting techniques for WSNs used in agricultural monitoring systems. They also explore the challenges and limitations of WSNs in agriculture, highlighting energy reduction and agricultural management techniques for long-term monitoring. Hamami et al. [12] present a review of the application of WSNs in the field of irrigation. Mekonnen et al. [13] present a review of the application of different machine learning algorithms in the analysis of sensor data observed using WSNs in agriculture. In addition, they analyze a case study on a smart farm prototype, based on IoT data, as an integrated food, energy, and water (FEW) system. Nabi et al. [14] present a comparative study of different studies to provide a deeper insight into these implemented systems. They also present a study of apple disease prognostic systems, highlighting their key characteristics and drawbacks. The result of their study can be used to select appropriate technologies to build a WSN-based system, optimized for precision apple cultivation, which will help farmers avoid the ravages caused by disease outbreaks.

Kabashi et al. [15] present a framework to design WSNs for agricultural monitoring in developing regions, taking into account the particularities of said environments. They propose new solutions and research ideas for sensor network design, including zone-based joint topology control and power scheduling mechanism, multi-sink architecture with complementary routing associated with backlink/storage, and a task scheduling approach with parameter, energy, and environment recognition. Authors such as Kassim et al. [16] present WSNs as the best way to solve agricultural problems related to optimization of agricultural resources, decision support, and land monitoring in order to perform those functions in real time. They explain in detail the hardware architecture, network architecture, and software process control of the precision irrigation system. García et al. [7] study different WSN deployment configurations for a soil monitoring PA system, to identify the effects of the rural environment on the signal and to identify the key aspects to consider when designing a PA wireless network. The PA system is described, providing the architecture, the node design, and the algorithm that determines the irrigation requirements. The results of their testbed show high variability in densely vegetated areas. These results are analyzed to determine the theoretical maximum coverage for acceptable signal quality for each of the studied configurations. Furthermore, there are aspects of the rural environment and the deployment that affect the signal. Zervopoulos et al. [17] present the design and deployment of a WSN capable of facilitating the sensing aspects of smart and precision agriculture applications. They describe a simple synchronization scheme, which was installed in an olive grove, to provide time-correlated measurements using the receiving node's clock as a reference. The obtained results indicate the general effectiveness of the system, although they appreciate a difference in the time correlation of the acquired measurements. Bayrakdar [18] investigated an intelligent insect pest detection technique with underground wireless sensor nodes for precision agriculture using a mathematical simulation model. To evaluate performance, he examined the received signal strength and path loss parameters. He observed the need for transmission of signals with different transmission powers for depth-based communication in wireless underground sensor networks.

Other authors study the application of WSNs to monitor specific crops. Khedo et al. [19] describe the implementation of the PotatoSense application, for precision agriculture with WSNs, to monitor a potato plantation field in Mauritius. They employ different energy efficiency algorithms, to ensure that the life of the system is prolonged. Additionally, they have developed a monitoring application to process the data obtained from the simulated WSN. Rasooli et al. [20] propose using WSNs and IoT to help increase wheat and saffron production in Afghanistan in the future. Using both techniques, they predict the control of the condition and growth of the crop as well as the ability to check soil, temperature, humidity, and other environmental parameters.

Some authors propose the observation of parameters utilizing WSNs in greenhouses. Chaudhary et al. [21] propose and discuss the use of the programmable system on chip technology (PSoC) as part of the WSN to monitor and control various greenhouse parameters. Srbinovska et al. [22] propose a WSN architecture for vegetable greenhouses, in order to achieve scientific cultivation and reduce management costs from the aspect of environmental monitoring. They have designed a practical and low-cost greenhouse monitoring system based on wireless sensor network technology to monitor key environmental parameters such as temperature, humidity, and lighting.

There are authors also studying energy savings in WSNs used in monitoring agriculture. Hamouda et al. [23] study the problem of selecting the sampling interval, for precision agriculture using WSNs, due to the energy limitation that appears when deploying sensors in WSNs. They propose a Variable Sampling Interval Precision Agriculture (VSI-PA) system to measure and monitor agricultural parameters for appropriate agricultural activities, such as water irrigation. Compared to other fixed sampling interval schemes, the proposed VSI-PA system provides a significant improvement in energy consumption, while maintaining a small variation in soil moisture, regardless of soil temperature values. Qureshi et al. [24] propose Gateway Clustering Energy-Efficient Centroid (GCEEC)-based routing protocol, where a cluster head is selected from the centroid position and gateway nodes are selected from each cluster. The results obtained, after evaluating the proposed protocol in comparison to last-generation protocols, indicated a better performance of the proposed protocol, and provided a more feasible WSN-based monitoring for temperature, humidity, and lighting in the agricultural sector.

Table 1 summarizes different previous studies, carried out by other authors, regarding the use of WSNs in soil monitoring for agriculture.

Table 1. Previous studies regarding the use of WSNs in soil monitoring for agriculture.

Types	References
Surveys of WSN Implementations for Agriculture.	[1,8–14]
Frameworks, Studies, Designs and Deployments for WSN.	[7,15–18]
WSNs for the Monitoring of Specific Crops.	[19,20]
WSNs in Greenhouses.	[21,22]
Energy Savings Studies.	[23,24]

Regarding the available sensors for soil monitoring, there are works, such as [25], that study farmed podzolic soils since these types of soils are under-represented in the relevant literature. In the study, the authors established the relationship between apparent electrical conductivity (ECa) and soil moisture content (SMC). The authors also evaluated the estimated SMC with ECa measurements obtained with two electromagnetic (EMI) induction sensors. The authors concluded that ECa measurements obtained through multi-coil or multi-frequency sensors had the potential to be successfully used for field-scale SMC mapping. Others, such as [26], designed and manufactured an integrated passive wireless sensor to monitor the moisture in the sand. The sensor was made of a printed spiral inductor embedded within the sand and it contained an inductive-capacitive (LC) resonant circuit. The authors measured the level of internal moisture by monitoring the resonance frequency using a sensing coil. Kizito et al. [27] presented a study where ECH20 sensors

were used to measure soil moisture content, bulk electrical conductivity, and temperature for a range of soils, across a range of measurement frequencies between 5 and 150 MHz. The authors affirmed that the measurements carried out on soil were accurate enough to work at 70 MHz. Finally, Nor et al. [28] discussed the development of a low-cost sensor array based on planar electromagnetic sensors to determine the contamination levels of nitrate and sulfate in water sources. The authors proposed three types of sensors: parallel, star, and delta. According to their experiments, the star sensor array was the one with the highest sensitivity.

After analyzing the exhibited works and many others not included in this paper, we can conclude that our work improves the existing systems. In either very few or no cases in the other works reviewed do the authors present complete or easily integrable systems in commercial nodes, such as Arduino or similar, and many of them use working frequencies that are too high (on the MHz scale). This fact makes it difficult to develop a simple and inexpensive signal generator circuit. Our proposal aims to take a step beyond the current state of the art, proposing a complete system, consisting of a sensor based on coils whose working frequency is around 93 kHz, and a power circuit that can be easily integrated into commercial modules for the development of a more complex wireless sensor network to monitor a large-scale crop.

3. Network Nodes Description

This section describes the proposed system and the different parts that comprise our proposed system. Additionally, it presents the features of the different coils used to develop our soil moisture sensor as well as the experimental tests performed to determine the best prototype.

3.1. Overall System Description

When trying to develop complete monitoring systems for precision agriculture, it is important to take into account different aspects. On the one hand, agriculture is an essential activity for the survival and development of society; this fact is evidenced by the amount of global- and regional-scale agricultural monitoring systems [29] to assess the crop growing conditions, crop status, and agro-climatic conditions that may have an impact on global production of any type of crop. Some examples are Group on Earth Observations Global Agricultural Monitoring Initiative (GEOGLAM) [30] or CropWatch [31], among others.

On the other hand, it is necessary to know the kind of crop wanted to be developed in order to design adapted methods for monitoring activity. Considering the crop to monitor and the location of the plot, the network should use a specific wireless communication technology. Currently, it is possible to use cellular technologies by paying for subscriptions to a service or by using low-power technologies such as ZigBee, LoRa, LoRaWAN, Bluetooth BLE, or Sixfox, among others; most of these services do not require payment for using their communication network infrastructure [32]. However, the wireless technology par excellence for developing wireless sensors networks continues to be Wi-Fi. Although its energy consumption is still high, it allows transmitting any type of content without the bandwidth limitations that other technologies present. In addition, it is a widely studied standard so it is easy to develop new optimized protocols. Therefore, by making a good design of a power system based on renewable energies, it is possible to use Wi-Fi to develop a Wi-Fi-based agriculture monitoring network with very interesting properties.

In the end, the completion of the design of the system led to precisely defining the type of parameter to be monitored since this fact will indicate the type of sampling, and analysis we should do. After that, the data interpretation and the scoring curves will help us to define the correct operation of our actuator network system. Lastly, the correct processing of collected data will help us to know the soil health and its characteristics for determining if these are optimal for our crop.

Therefore, considering these previous issues, we propose the development of a group-based wireless sensor network for soil moisture monitoring in precision agriculture. The

network is composed of a set of nodes with different roles and functions. Some nodes are able to collect data from the environment, particularly data from soil moisture and other parameters required to ensure the correct progress of a tree (See Figure 1). The rest of the nodes have actuators to control the activity of ditch gates and drip irrigation elements. So, we will have 3 different sets of nodes that will communicate between them. Additionally, sensor nodes will provide data to the actuator nodes performing the required computation and decision making in the edge. Edge computing is recommended in scenarios where nodes present in the network are able to analyze the data and take decisions. Edge computing enables data produced by Internet of Things (IoT) devices to be processed closer to where it was created rather than being sent over long journeys to reach data centers and computing clouds. One of the fundamental advantages of this type of computing is that it allows analyzing important data in near real time [33]. In citrus groves, it is common to distribute them by forming rows of trees separated at a distance of approximately 6 m, being able to opt for a denser plantation, with a minimum separation of 4.5 m. The minimum depth that a citrus tree usually reaches is 45 cm. Considering these facts and taking into account that a field can have different extensions, different topologies of sensor nodes can be created. An important aspect is to ensure complete coverage between nodes to guarantee stable communication. A distributed ad-hoc network is optimal for this kind of scenario.

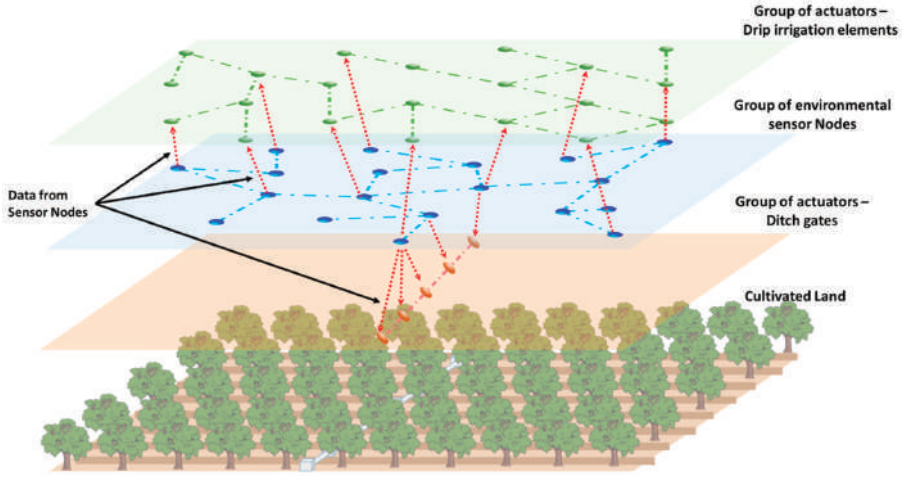


Figure 1. Proposed group-based network.

One of the main characteristic aspects of this proposal is its hierarchical structure by layers where each layer has a series of nodes that, if necessary, could change their role. That is, all sensor nodes and actuator nodes are wireless devices with the ability to act as a packet relay. In a hypothetical situation where a node falls, communications can be rerouted by other nodes of the same layer. If there is a fall of several nodes and one of them is isolated but active, it could use nodes of the upper or lower layer as an alternative way to carry out communications. However, these nodes would only forward the packet to nodes of the isolated node's layer.

To deal with the failure of a sensor or actuator node, it is convenient to establish an alarm system, based on keep-alive messages. It is a task periodically scheduled, once per day. It is possible to work with a large periodicity because the irrigation tasks of a field are not considered a critical task. If there is a node or several nodes not responding to these requests, the system will consider a node is down.

Additionally, developing a low-cost system was required to measure the moisture in the soil depths. This system consists of four coil-based sensor elements equally distributed

along 60 cm. The coils are connected to a processor module in charge of collecting the data and wirelessly share them with the rest of the nodes of its group. Finally, and considering the values of moistures collected by the sensor nodes, the actuator nodes will enable/disable the ditch gates or the drip irrigation.

When talking about moisture or soil humidity, we refer to the amount of water the soil contains. A gravimetric analysis method gives the relative comparison between the mass of dry soil and the mass of watered soil (which will always be higher). The moisture given in percentage is the result of dividing the difference between these two values by the mass of dry soil. If there is no difference, moisture will be 0%. In the opposite case, when the watered soil mass doubles the dry one, the moisture level will be 100%.

The development of our coil-based soil moisture sensor is based on the principle of electromagnetic induction of the coils and how it varies as a function of the type of core the coil has inside [34–36].

The soil moisture sensor is composed of two solenoid coils wound on the same PVC pipe support. Coil 1 receives the sinusoidal signal generated by the power circuit based on the integrated ICM7555. Coil 1 induces a current on Coil 2 which is largely affected by the content of the coil core since the magnetic field is affected by the type of soil and water content inside it. Finally, this current is measured, collected, and stored with an electronic module. In our case, a module ESP32 DevKIT [37] with an integrated Wi-Fi interface has been chosen. Figure 2 shows the diagram of the proposed soil moisture sensor.

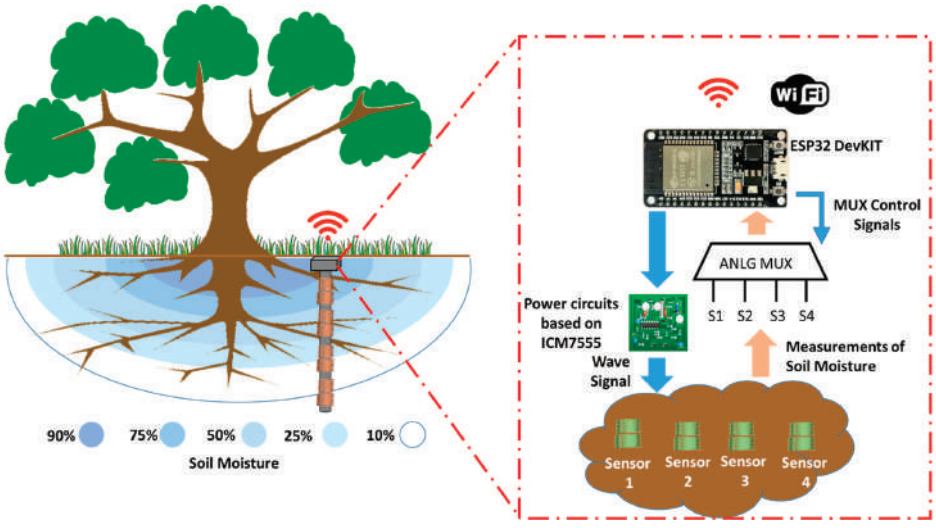


Figure 2. Diagram of proposed soil moisture sensor base on coils.

Since this kind of module usually presents one or two analog inputs to collect data, we also propose the use of an analog multiplexor of four inputs which can be controlled by using two digital outputs. With this, our system will be able to take measurements from the four soil moisture sensors.

3.2. Soil Moisture Sensor Based on Coils

As we mentioned before, it is possible to develop soil moisture sensors based on several principles and chemical processes. However, we want to use a method based on physical principles such as the variation of electromagnetic flow as a function of the nature of the coil core.

In a coil distribution such as the one shown in Figure 3, coil 1 generates a magnetic field that affects coil 2. This effect is known as mutual inductance and refers to the electromotive

force (EMF) in a coil due to the change of current in another coil attached. The induced EMF in a coil is described by Faraday's law and its direction is always opposite to the change in the magnetic field produced in it by the coupled coil (Lenz's law). The EMF in coil 1 (left) is due to its own inductance L.

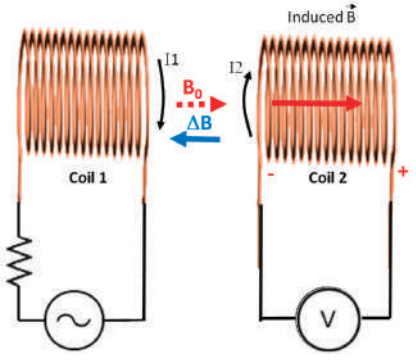


Figure 3. Principle of operation for our developed sensor.

The induced EMF in coil 2, generated by the changes of current I1, can be expressed as (see Equation (1)):

$$emf_2 = -N_2 A \frac{\Delta B}{\Delta t} = -M \frac{\Delta I_1}{\Delta t} \quad (1)$$

where N_2 is the number of spires of coil 2, M the coefficient of mutual self-induction, A is the cross-sectional area of the coil, $\frac{\Delta B}{\Delta t}$ the variation of the magnetic field as a function of the time, and $\frac{\Delta I_1}{\Delta t}$ the variation of current in coil 1 as a function of time. Mutual inductance (M) can be defined as the ratio between the electromagnetic force (EMF) generated in coil 2, and the changes in current in coil 1 that causes that EMF. Likewise, M is highly affected by the characteristics of the medium that surrounds the coils, usually expressed by its magnetic permeability.

Since it is difficult to measure the value of the magnetic permeability of the earth core as a function of the moisture level, two theoretical approximations of the air core are introduced [38]. Based on Equation (2) (which presents the coil inductance), we can state Equation (3) where l is the length of the coil and r is the radius to the center of the coil of the innermost layer of the conductor while R is the radio for the outermost layer.

$$L = \frac{\Phi N}{I} = \frac{\mu N^2 A}{l} = \frac{\mu_0 \mu_r N^2 \pi r^2}{l} \text{ (H)} \quad (2)$$

$$L_{layer} = \frac{N^2 r^2}{2.54(9r + 10l)} \text{ (μH)} \quad (3)$$

where L is the inductance of our coil (in H), Φ is the magnetic flow (in Wb), N represents the number of turns (dimensionless), l is the length of the coil (in m), r expresses the radius of the inner coil's layer (in m), R is the radius of the outer coil's layer (in m), A is the area of the coil's surface (in m^2), μ_0 is the magnetic permeability (free space) (H/m) and finally, μ_r is the relative magnetic permeability (medium) (dimensionless).

This approximation allows estimating the components of the circuit for an air core, which would be similar to those obtained with a large amount of pure water; so this value will vary depending on the type of soil, its composition, and the level of soil moisture presented by the soil that contains coil. The resonance peak of our coils can be calculated by Equation (4).

$$f_r = \frac{1}{2\pi} \sqrt{\frac{1}{LC_d} - \frac{R_S^2}{L^2}} \approx \frac{1}{2\pi\sqrt{LC_d}} \text{ (Hz)} \quad (4)$$

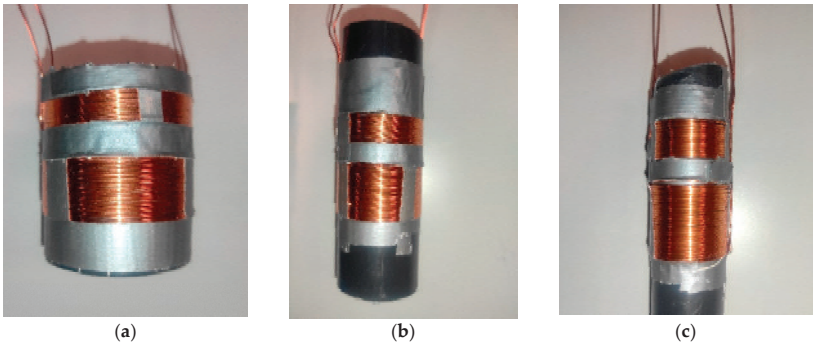
where f_r is resonance frequency (in Hz), C_d is coil's parasite capacity (in F), L is the coil's inductance (in H) and, R_s is the coil's resistance (in Ω).

We should take into account that the primary coil and secondary coil will have different resonance frequencies because the secondary coil has a different number of coils. However, our sensor only intends to detect changes in the induced current due to the presence of a changeable medium and, finally, we want to relate this value of current with the amount of water content in the soil.

Equations from (1) to (3) are theoretical approaches to explain how important it is to know the relationship between the physical and electrical characteristics of the coil. Equations (1)–(3) explain how the coil inductance, and hence mutual inductance, depend on its geometry (length, radius, and the number of turns) for single-layer coils. Equation (4) helps us to design a resonant circuit to obtain the maximum power transfer. It is highly important to consider the appearance of a possible parasite capacity due to coil geometry and working frequency.

We previously performed several tests with different combinations of coils, varying the number of spires, the ratio of spires between coils, and the diameter [39]. In these previous works, we performed many experiments with combinations of spires and the best results were determined for a ratio of 1:2 with a medium value of spires and larger diameter. For a fixed diameter, if we reduced the number of spires, the working frequency increased for a fixed number of spires; if we increased the coil diameter, the working frequency decreased. Additionally, developing a simple and cheap electronic system to generate the signals was required. For such a system, it is highly recommended to have a sensor that requires the lowest working frequency. Therefore, we chose to set some parameters such as the type of copper and number of spires, we only varied the diameter of coils.

In developing our coils, 0.6 mm enameled copper wire was used. The process entails winding copper wire along a cylinder, forming two solenoids. The distance between the primary coil and the secondary coil is five mm. Figure 4 shows the developed coils with a single layer of spires. Table 2 shows the physical features of each prototype.



(**a**) (**b**) (**c**)

Figure 4. Coils used in our developed sensor: (**a**) P1, coil of 50 mm;(**b**) P2, coil of 32 mm; (**c**) P3, coil of 20 mm.

Table 2. Prototypes to measure soil moisture.

Prototype	Caliber	N° Layers	N° Spires 1st Coil	N° Spires 2nd Coil	N:n	Diameter
P1 (Figure 4a)	0.6 mm	1	15	30	1:2	50 mm
P2 (Figure 4b)	0.6 mm	1	15	30	1:2	32 mm
P3 (Figure 4c)	0.6 mm	1	15	30	1:2	20 mm

The procedure to perform these tests with the coils consists of introducing each model into a container filled with dry and compacted soil to observe the behavior of the output voltage as a function of the amount of water. Therefore, for the same moisture level, a

frequency sweep will be carried out to find the frequency that shows a peak in the induced voltage. This value will be the sought-out resonance frequency. After that, the linearity of each model will be analyzed. For each test, 4000 g of soil is used, with increments of 250 mL of water, for each moisture level up to 1000 mL of water. When water is added to the soil, the sample is reposed for an hour to obtain a homogeneous sample inside the coil. Specifically, five levels of content of water in soil will be measured: 0%, 6.25%, 15.5%, 18.75%, and 25% (see Table 3). For this type of soil, 25% of the content of water in soil implies a land completely flooded. Measurements have been taken at 25 °C.

Table 3. Samples used during the tests.

Sample	Mass of Dry Soil (g)	Mass of Wet Soil (g)	Mass of Water (g)	% Volumetric Water Content
1	4000	4000	0	0
2	4000	4250	250	6.25
3	4000	4500	500	12.5
4	4000	4750	750	18.75
5	4000	5000	1000	25

In order to determine in which type of soil our sensor can be used, we endeavored to determine which one presents the biggest linearity. The idea of this concept is that an increase in the percentage of moisture is equivalent to an increase in output voltage without instabilities.

3.3. Experimental Results with the Developed Coils

This subsection presents the test performed to determine the most suitable prototype selected to develop the system, followed by the testing of the selected coil with different types of soils and different levels of moisture.

Considering the types of soil, we can conclude that soils are usually made up of different proportions of sand, silt, and clay. Each of them has morphological characteristics:

- Sandy soils, coarse texture (sand and clay sand).
- Silty soils, moderately coarse texture (sandy clay and fine sandy clay), medium texture (very fine sandy clay, silt, silt loam, and sediments), and moderately fine texture (clay silt, clay sand loam, and sand loam soil silty).
- Clay soils, fine texture (sandy clay, silty clay, and clay).

In order to perform our tests, we have selected three different types of soils, i.e., sand from the beach, soil from cultivated land, and commercial universal substrate.

The sand on the beaches is formed by sediments from rocks and other marine debris such as shells, corals, animals, algae, and even sand that travels through the rivers until flowing into the sea. Due to the erosion of water and wind, due to rain and waves, or temperature differences, the grain size of the sand tends to be reduced.

The soil from cultivated land is usually made up of an organic fraction, organic matter more or less degraded into humus and humic and fulvic acids. These elements provide the fertile part of the earth. The rest of the soil is considered as physical support. Some farmlands have a high degree of clay which also intervenes in ion exchange and water retention, facilitating the release of fertile elements according to the needs of the plants.

The raw materials used in the manufacture of a commercial universal substrate are usually blonde peat from sphagnum moss, coconut fiber, compost, perlite, organic fertilizer, mineral fertilizer, algae extract, etc. In addition, this type of soil usually contains a high level of aeration.

Performing the measurements entails three identically constructed sensors simultaneously placed in three samples of each soil. The results shown in our graphs are the average value of the three measurements collected, which in all cases were identical.

In order to perform the test, the primary coil is powered by using a wave of 7 Vpp with positive and negative values. For example, it is possible to use sine or square waves,

such as the one shown in Figure 5. In Figure 5, we can see, in blue, the signal used to power the primary coil while the result of the induced current is shown in yellow.

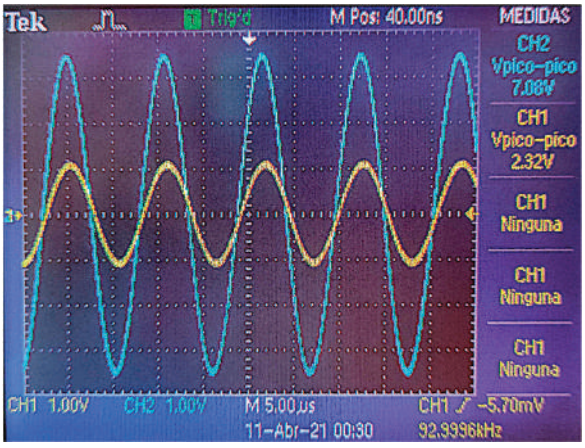


Figure 5. Example of generated and obtained signals.

Figure 6 shows the preliminary results obtained with the 3 coils. Figure 6 shows the value of the resonance frequency and the maximum voltage value.

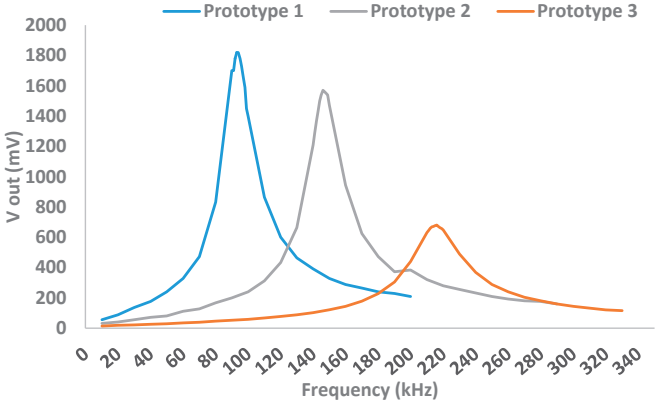


Figure 6. Resonance frequency obtained for each prototype.

Table 4 shows the resonance frequency values (in kHz) of the developed prototypes and the maximum voltage value (in mV) obtained in the induced coil.

Table 4. Prototypes to measure soil moisture.

Prototype	Working Frequency (kHz)	Maximum Voltage (mV)
1	93	1820
2	146	1570
3	216	680

After analyzing the results obtained in Figure 7, we can conclude that the prototype that gives the best results is Prototype 1, with a working frequency of 93 kHz and a maximum output voltage of 1.82 V.

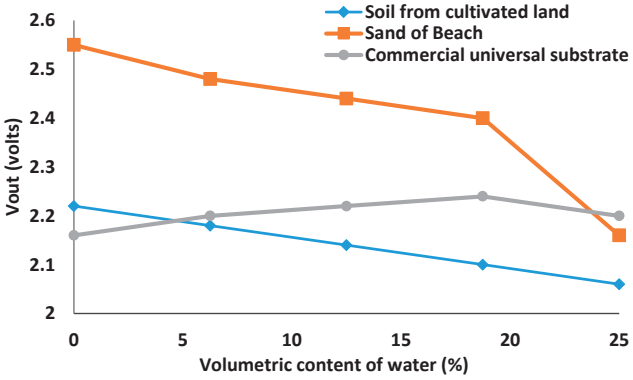


Figure 7. Behavior for prototype 1 in several types of soils.

Once the most suitable sensor for further development has been determined, this sensor is tested on different types of soil to determine its versatility. In all cases, we will look for the maximum linearity in the sensor response.

As we can see in the previous figure, the selected model has a linear behavior for all three cases up to a humidity degree of 18.75%, i.e., a total volume of water of 750 mL for 4000 g of sand.

Another important aspect to highlight is that the behavior of the sensor for a universal substrate is inverse to the behavior shown in the case of beach sand or cultivated soil. This aspect should be considered when the results are processed in a real environment.

3.4. Power Circuit Design

Locating the resonance frequencies for the selected prototype of coil systems requires a power supply and excitation circuit able to generate an alternating signal. To do this, a 555 series oscillator integrated circuit [40] will be used. A series of components will be used to obtain the desired output signal with a resonance frequency of 93 kHz. Our circuit is based on the ICM7555 [41]. According to the manufacturers' specifications, this integrated circuit can generate signals up to 3 MHz. Figure 8 shows the schema of our entire circuit. This kind of integrated circuit has been conceived to be customized regarding the duty cycle of signal and the required frequency. In this case, we can use R2 and C1 to change the working frequency while C3, C4, and R1 are used to control the ripple signal and its form. Modifying C4 and R1, it is possible to obtain both a sine wave and square signal, such as the one shown in Figure 9.

To directly read a value of voltage proportional to the amount of water content in the soil, we include a Graetz bridge or double wave rectifier bridge followed by an RC filter that has been connected to the terminals of the secondary coil.

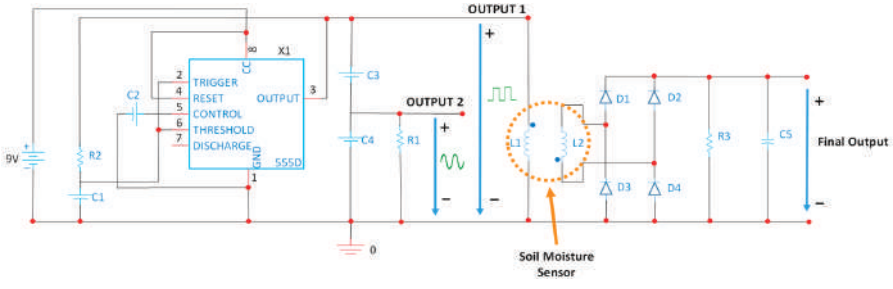


Figure 8. Enhanced power circuit schematic.

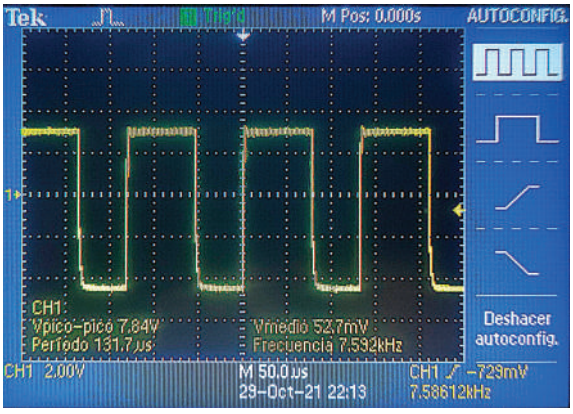


Figure 9. Example of signal obtained in Output 1.

3.5. Comparision and Discussion with Existing Published Systems

We have compared our sensor model with existing and commercial soil moisture sensors. Table 5 shows this analysis. It is important to consider that this table only contains the price of the sensor, with exception of references [42–45] which include an electronic module. In the rest of the cases, a microprocessor module must be included similar to the one used in this paper that can cost approximately $10–$15.

Table 5. Comparison of soil moisture sensors with our proposal.

Ref.	Model	Sensitivity	Power	Size	Cost
[42]	RK520-02 Soil Moisture Sensor, Temperature Probe & EC Sensor	0–100%	5 VDC, 12–24 VDC	136 × 45 mm	$58.00–$72.00
[43]	S-Soil MT-02A	±3% (0–53%) ±5% (53–100%)	3.6–30 VDC	149 × 45 mm	$79.00
[44]	S-Temp&VWC&EC-02A	±2% (0–50%) ±3% (50–100%)	3.6–30 VDC	149 × 45 mm	$99.00
[45]	SenseCAP Wireless Soil Moisture & Temperature Sensor	±2% (0–50%); ±3% (50–100%)	3.6 V	149 × 45 mm	$219.00
[46]	Sensor YL-69	0–95%	3.3–5 VCD	60 × 30 mm	$2.65
[47]	Keyes Brick Soil Moisture Sensor Module	-	3.3–5V VDC	63 × 22 × 8 mm	$1.34
[48]	KeeYees	-	3.3–5 VDC	3.858 × 0.906 mm	$7.99
[49]	Seeed Studio Grove—Capacitive Moisture Sensor—101020614	-	3.3–5 V	-	$5.95
[50]	Grove—Soil Moisture Sensor	0–95%	3.3–5 V	60 × 20 × 6.35 mm	$2.99
[51]	Seeed Studio Moisture Sensor—101020008	-	3.3 V CC y 5 VDC	-	$3.99
[52]	SEN0308	0–57%	3.3–5.5 VDC	175 × 30 mm	$15.51
[52]	MSE020SMS	-	3.3V–12 VDC	-	$5.80
-	Our proposal	-	9 V	50 m × 30 m	$2.15

It is evident that our proposed sensor, based on 2 coils, is one of the models that presents the lowest prices. The price includes the pipe and wire, because it can be added to any electronic platform to gather the data.

When this type of system is designed and developed, it is extremely important to consider practical implementation problems and challenges.

One of the main problems in the outdoors is how to protect the electronics from adverse conditions. A waterproof protection is highly recommended to protect the dif-

ferent devices since the places where the devices are deployed can be highly changeable. Additionally, the time during which the sensor nodes should work and the exposure to environmental temperature and humidity can cause some variations in the measurements. This issue should be controlled since a wrong reading would cause anomalous values and consequently wrong behavior. To shore up this problem, it is possible to use artificial intelligence and redundancy mechanisms.

Furthermore, there is another important issue regarding manufacturing techniques of certain sensors and probes. In several cases, they are manufactured with copper. An improvement in the system implementation could be the replacement of these sensors for ones protected with the process of gold plating which help to combat the corrosion of probes.

Coverage estimations do not usually match practical experimentation because the emulation of environmental conditions is difficult. For this reason, we highly recommend performing practical experiments and test benching, as presented in this paper.

4. Network Protocol Design and System Procedure

This section presents the network protocol used in our topology. In addition, it also presents the algorithm designed to collect data from sensors and control the different actuators as well as the messages exchanged between devices and the algorithm designed for the system procedure.

As we presented before, our network is composed of three different types of nodes, which can be classified as sensor nodes and actuator nodes. Additionally, we consider an additional node that is placed in the engine to provide water to the plot. This node will be in charge of starting the process of monitoring the entire network. Sensor nodes collect data from the soil and provide the required warning alarms to the actuators for enabling or disabling the irrigation systems. Figure 10 shows the diagram of our entire network deployed in the plot.

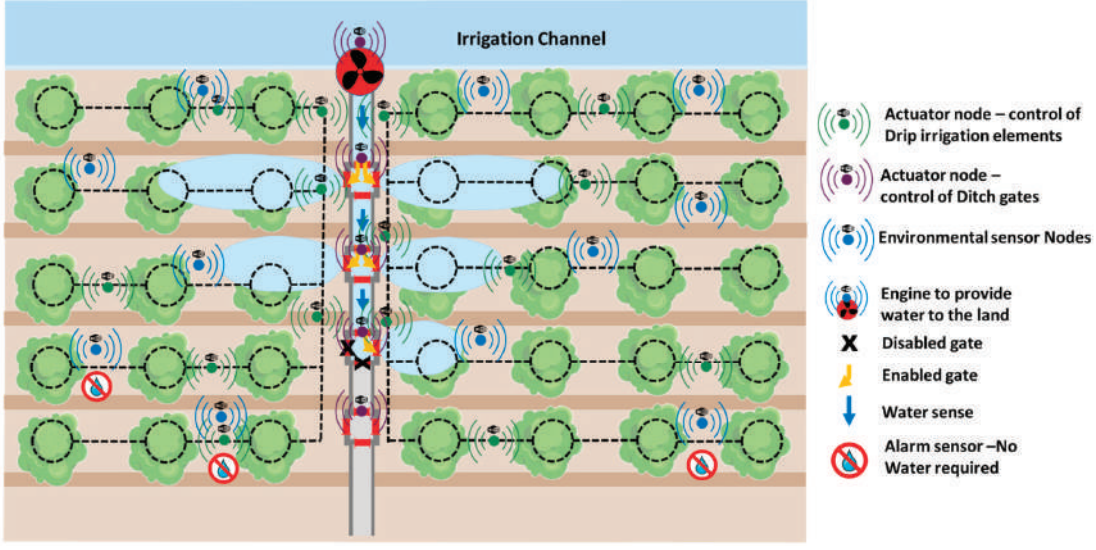


Figure 10. Diagram of our entire network in the plot.

In order to obtain high performance, we have developed a specific network protocol. This section presents the message exchange between devices, the fields of the messages exchanged between devices, and the algorithm designed for the system procedure.

The designed network is a distributed network made of sensor nodes (each one has one or several physical moisture sensors), and one or several actuators which activate the engine, for the drip irrigation system and/or the ditch gates, depending on the case.

Then the number of nodes (see Equation (5)) of the whole system (N) is:

$$N = n_s + n_a \tag{5}$$

where n_s is the number of sensor nodes and n_a is the number of actuator nodes.

Our network will use Ad Hoc On-Demand Distance Vector Routing (AODV) since it is one of the ad hoc routing protocols that presents the best performance [53].

4.1. Algorithm of the System

In order to determine when the irrigation process should be carried out, we need to collect the data from the different n_s which are placed and identified by zones (i). For each zone, we defined the maximum number of nodes comprised in the zone as counter. Carrying out this automation process of irrigation requires the design of an operation algorithm. Figure 11 shows the operation algorithm of our soil moisture monitoring system.

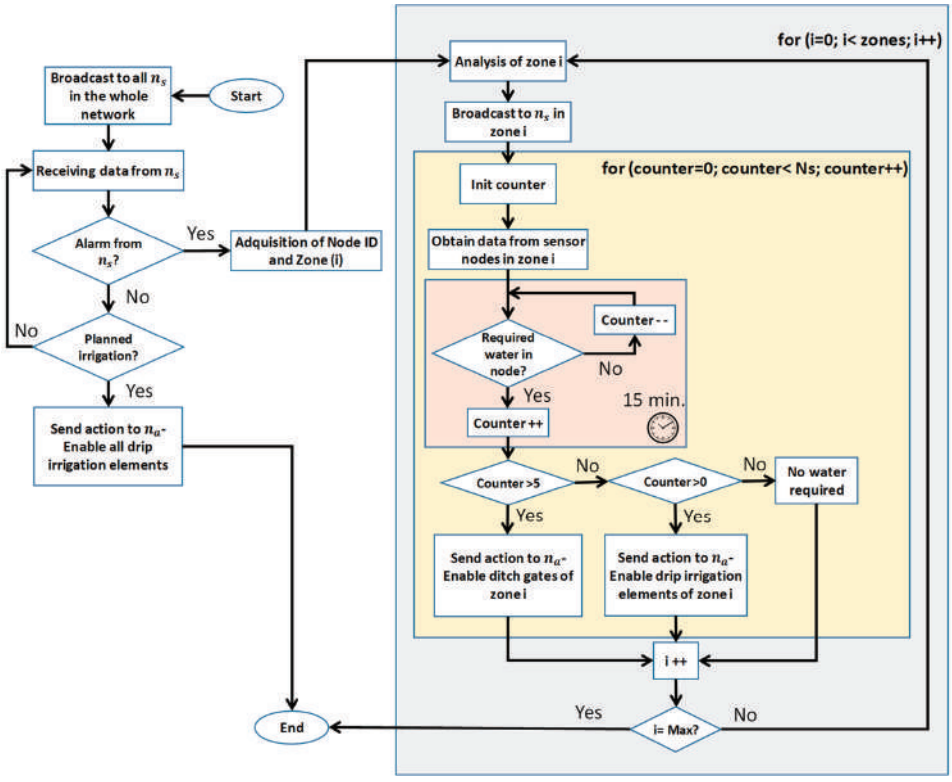


Figure 11. Operation algorithm.

As commonly done in agriculture, there are periodic planned irrigations that should be performed. In this case, the system of drip irrigation elements is enabled and it covers the entire extension of trees. If an alarm is registered from a sensor, the system will request the data from all nodes of that zone. If the number of nodes that register the need for water is higher than 5, the system will enable the ditch gate of this zone. Even if only some sensors warn about the need for water, the system will enable the drip irrigation

elements of this zone. The rest of the zones in the plot will be analyzed to check if it is required to proceed with irrigation. The different orders will be sent to the sink node by the nearest node of that area to the sink node which will be in charge of enabling/disabling the irrigations systems.

Finally, if the plot does not require any action, the system will remain in idle mode waiting for new information.

4.2. Message Flow between Nodes

Finally, in order to send the required actions to the correct actuator nodes, it is important to design the message exchange between nodes. In this sense, we should consider three different situations (see Figure 12). Firstly, the most frequent situation is the one in which the plot does not require any type of irrigation. In this case, if the sensor nodes do not send any message in the next 30 min, the system will consider that no irrigation is required (1).

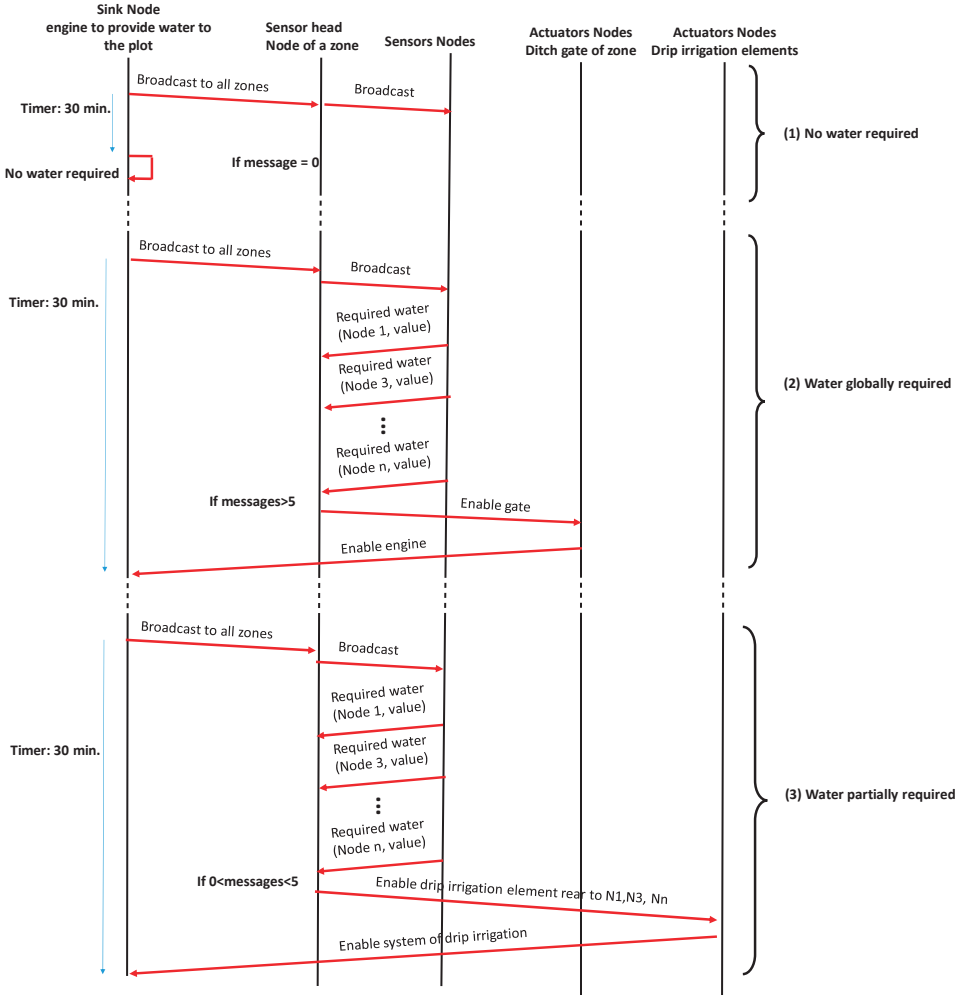


Figure 12. Message exchange between nodes.

The second situation (2) is when there is a global need for water in a zone of the plot. In this case, the sensor head node will wait 30 min for messages from sensor nodes. If more than 5 messages are received, the system will consider that global irrigation for this zone is required. Then, the sensor head node will send a message to the actuator node in charge of enabling the gates. After that, this node will inform the sink node to enable the engine to provide water to the ditch.

The third situation (3) will be done when there is a partial need for water in a zone of the plot. In this case, the sensor head node will wait 30 min for messages from sensor nodes. If less than 5 messages are received, the system will consider that partial irrigation for this zone is required. Then, the sensor head node will send a message to the actuator node in charge of enabling the drip irrigation element of the affected trees. After that, this actuator node will signal the sink node to enable the system of drip irrigation.

To make easier the process of forwarding messages from sensor nodes to actuator nodes or sink nodes, it is possible to use any node present in the network. In this sense, a node can receive several packets but if it is not the destination of this message, the node will relay the message without processing it. When the sensor nodes of a zone are communicating, an intragroup routing protocol will be utilized. When the message exchange is performed between sensor nodes and actuator nodes or between actuator nodes for ditch gates and actuator nodes for drip irrigation elements, these nodes will use an intergroup routing protocol [54].

5. Experimental Results in a Practical Deployment

In this section, the results obtained in the deployments on orange groves are presented. In order to perform the test, we have used several ESP32 DevKit nodes placed at different heights. This will allow us to study the coverage of the nodes at different heights, so it will be kept as a recommendation for practical deployments. The different deployment strategies that were tested are presented in Figure 13. As can be seen, different configurations of emitter height and receiver height were tested. The emitters were deployed at heights of 0.5 m, 1 m, 1.5 m, and 2 m. The receivers were placed at 0 m for the on-ground deployment, 0.5 m for the near-ground deployment, and 1.5 m for the above-ground deployment. The emitter and receiver were separated for each test. The trees are spaced in four-meter intervals and the field is located in an area with a Mediterranean climate. The foliage of the trees affects the wireless communication among the devices. Testing different configurations of transmitter and receiver provides us with the knowledge to design the best deployment for optimal communication with this type of crop. The tests were performed with sunny weather and temperatures of 20 °C. The measurement carried out is the received signal strength indicator (RSSI) at different measuring points. The Esp32 DevKit nodes were encapsulated on a protective box.

The results for the emitter at a height of 0.5 m and the receiver at different deployment configurations are presented in Figure 14. The positions of the trees are indicated by the bold orange numbers on the X-axis. As can be seen, the overall higher RSSI values considering multiple trees along the tested distance were obtained for the near-ground position of the receiver. This configuration of the receiver is also the most stable. Moreover, some small fluctuations occurred for tree number 1 and tree number 4. However, the foliage in the space between trees 2 to 4 presented a higher density, which lead to higher fluctuations. One of the reasons for these fluctuations may be the multipath effect. Thus, avoiding node deployments in areas of high foliage density is best to obtain more stable signals.

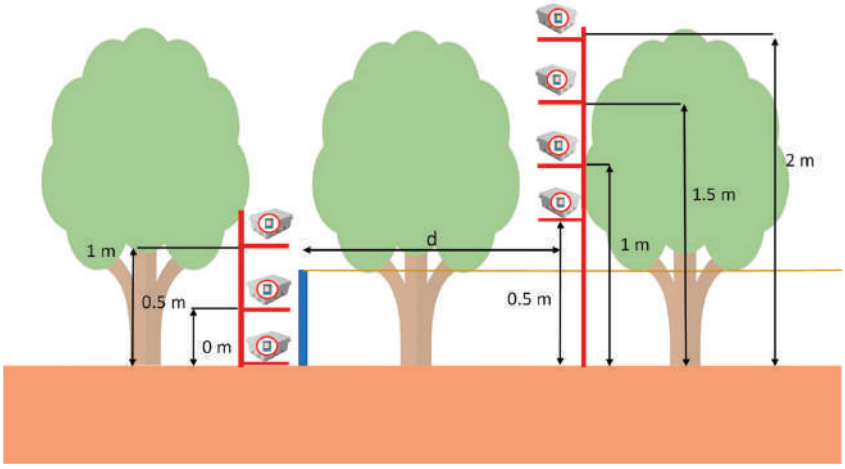


Figure 13. Testbed.

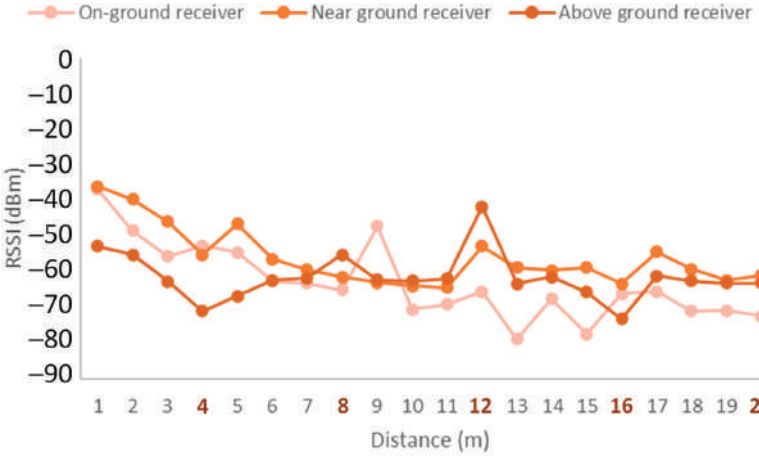


Figure 14. Emitter at height of 0.5 m.

For the case of the emitter deployed at a height of 1 m, the results are presented in Figure 15. As it can be seen, the near-ground receiver is the one with the best results. In this case, the signal quality is reduced between trees 2 and 4. However, for the near-ground receiver, the signal presents some recovery after the area with high foliage density. The above-ground deployment presents similar results for the area with high foliage density but worse signal quality for the rest of the measurement points. Lastly, the on-ground receiver deployment presents the worst results.

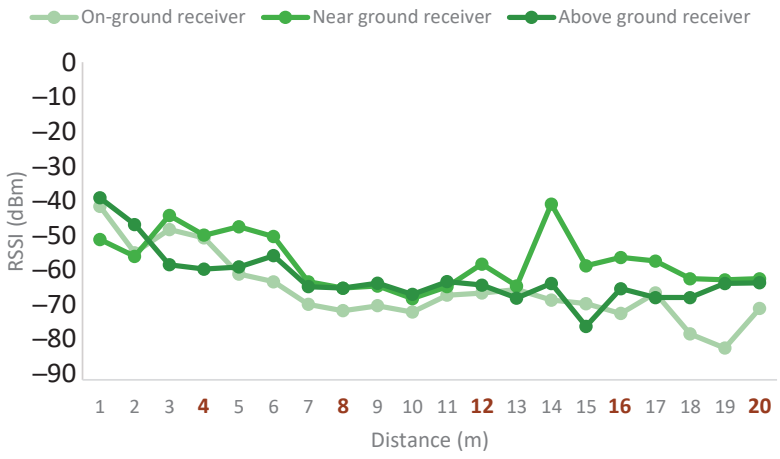


Figure 15. Emitter at height of 1 m.

Figure 16 presents the results for the emitter height of 1.5 m. The near-ground deployment has the highest signal quality values at almost all measuring points. As can be seen, it experiences some fluctuations between trees 2 and 3. However, even with the fluctuations, the signal quality is better than that of the other configurations. The next best option is the above-ground receiver. In this case, the signal is more stable while remaining below the quality levels of the near-ground receiver. Lastly, the on-ground deployment presented the worst results and the highest fluctuations. Another final aspect to consider is that the average signal quality for this emitter height was lower than the signal quality obtained for lower emitter heights.

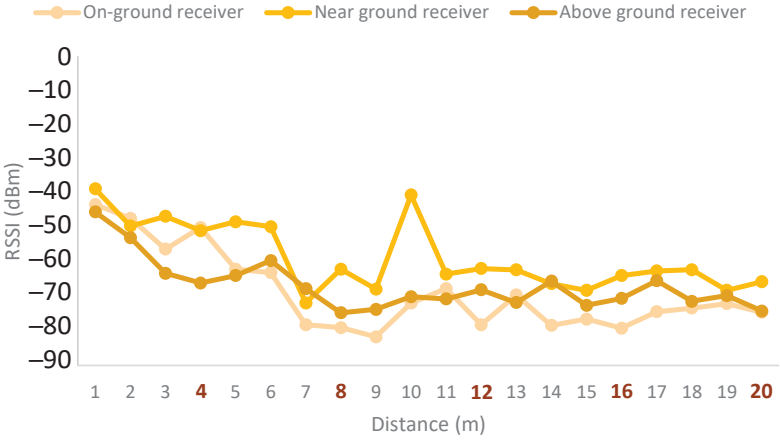


Figure 16. Emitter at height of 1.5 m.

Lastly, the results for the emitter height of 2 m are presented in Figure 17. This emitter height obtains the worst signal quality results compared to all the emitter heights. Regarding the receiver height, in this case as well, the near-ground deployment obtained the best results. However, as shown in the figure, all receiver configurations present similar results, while the results for this emitter height present the least fluctuations. As in the other cases, the on-ground configuration was the worst option.

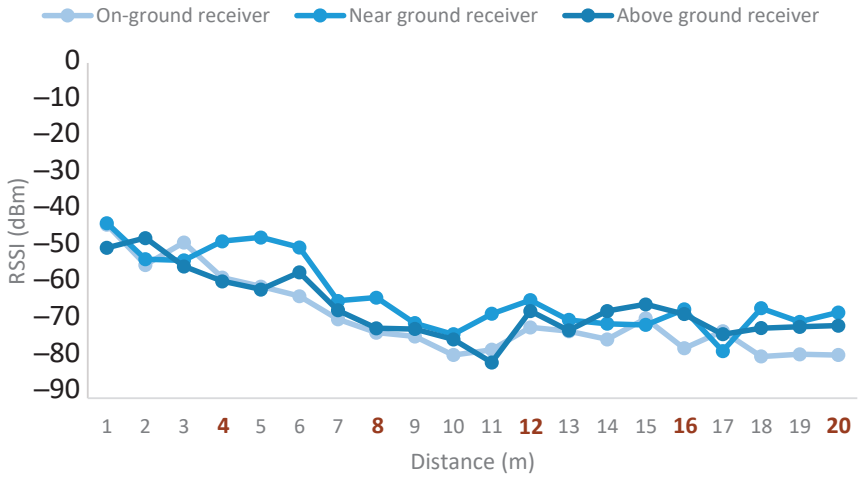


Figure 17. Emitter at height of 2 m.

Considering the results for all the emitter heights, we can conclude that in the case of orange groves, emitter heights of 0.5 and 1 m present the best signal quality and the near-ground receiver deployment is the best option for all emitter heights. Therefore, near-ground configurations are the optimal deployment style for both emitters and receivers.

The coverage results obtained from the tests performed on the orange groves have been utilized to obtain a heuristic signal attenuation model for all emitter heights as specified in [7]. The outlier values were discarded to perform this model. Equations (6)–(9) show the model for emitter heights of 0.5 m, 1 m, 1.5 m, and 2 m respectively.

$$P_{0,5\ m} = -7.182 \ln d(m) - 45.276 \quad (6)$$

$$P_{1\ m} = -7.69 \ln d(m) - 44.194 \quad (7)$$

$$P_{1,5\ m} = -9.545 \ln d(m) - 44.475 \quad (8)$$

$$P_{2\ m} = -10.34 \ln d(m) - 43.493 \quad (9)$$

Furthermore, the model, confidence intervals, and prediction intervals are presented in Figure 15, where the dots represent the values obtained from the tests on the fields. As can be seen, the model reflects that the configurations of emitter heights of 0.5 m and 1 m (See Figure 18a,b) present better signal quality. Lastly, Figure 18c shows the graphic representation for the case of the emitter height at 1.5 m and Figure 18d presents the results for the emitter height of 2 m.

Considering the results for all the emitter heights, we can conclude that for the case of orange groves, emitter heights of 0.5 and 1 m present the best signal quality and the near-ground receiver deployment was the best option for all emitter heights. Therefore, near-ground configurations are the optimal deployment style for both emitters and receivers.

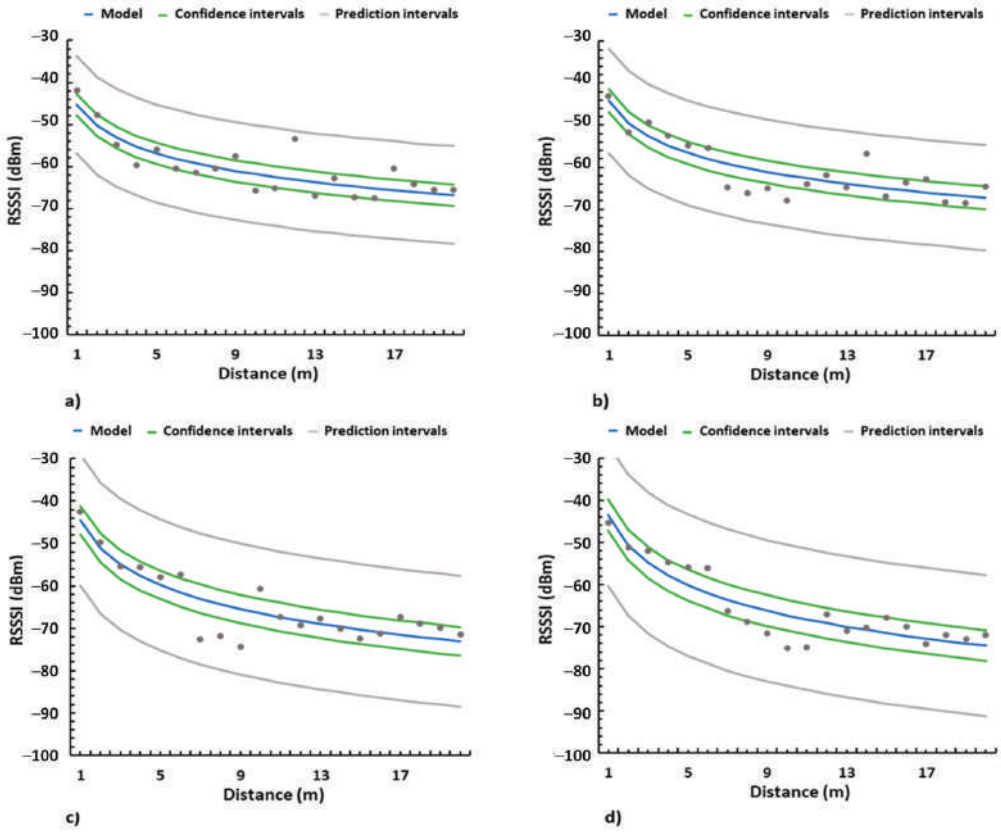


Figure 18. Heuristic model for (**a**) emitter at height of 0.5 m, (**b**) emitter at height of 1 m, (**c**) emitter at height of 1.5 m, and (**d**) emitter at height of 2 m.

6. Conclusions and Future Work

Estimating the amount of water needed to irrigate a crop is essential to carry out efficient use of a scarce resources such as water. The introduction of technology in the agricultural sector is also important to improve the sustainability and competitiveness of the sector. For this reason, this paper has presented the prototype of a low-cost sensor based on coils for measuring soil moisture. For this, three prototypes composed of two coils with different characteristics have been presented. These coils have been tested to analyze their behavior based on the humidity level of the soil. After the observed results, it has been concluded that the sensor that has had the best performance is prototype 1 working at 93 kHz. Additionally, a power circuit based on the ICM7555 has been designed to be able to generate the biphase signal to power the soil moisture sensor. This sensor is able to measure the percentage of water content in the soil at the desired depth. This fact helps us to ensure the correct irrigation of the root ball. The sensor and power supply circuit is connected to an ESP32 module for reading and storing humidity measurements. The entire system has been tested with real samples for the extraction of its mathematical behavior model. The results show that our sensor demonstrates that by using these models we can achieve accuracies close to 95%.

Additionally, the network performance has been tested in a real, cultivated plot. According to the results, and after modeling mathematically the results of the network coverage, we can conclude that for the case of orange groves, the best results are obtained when the emitter is placed at 0.5 and 1 m and the receiver is placed near the ground.

So, near-ground configurations are the optimal deployment style for both emitters and receivers.

In future work, we would like to perform more practical experiments with more models of coils and different kinds of soils to design a more versatile sensor capable of working with several sorts of soils without changing the sensor. It will also study the possibility of including a system to automatically adapt the working frequency to the type of soil. Because in our practical experiments we have included only the measurements of signal amplitude, it could be interesting to measure the quadrature component and phase of the obtained signal and trying to relate these parameters with changes of pH of water. We also want to include other sensors in a multi-parametric node to place in the crop field [55,56] to enhance the efficiency of water management in precision agriculture [57]. In this sense, we want to check if soil temperature has some effect over the soil moisture measurements and, if required, over obtaining the soil moisture values compensated with temperature. Finally, as the last step, we will study the most appropriate enclosures to protect our entire system.

Author Contributions: Conceptualization, S.S. and J.L.; methodology, S.S. and J.L.; validation, L.G. and J.M.J.; formal analysis, S.S. and J.L.; writing—review and editing, S.S., J.L., L.G. and J.M.J.; All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially supported by the European Union through the ERANETMED (Euromediterranean Cooperation through ERANET joint activities and beyond) project ERANETMED3-227 SMARTWATIR, by the "Programa Estatal de I+D+i Orientada a los Retos de la Sociedad, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2017–2020" (Project code: PID2020-114467RR-C33) and by "proyectos de innovación de interés general por grupos operativos de la Asociación Europea para la Innovación en materia de productividad y sostenibilidad agrícolas (AEI-Agri)" in the framework "Programa Nacional de Desarrollo Rural 2014–2020", GO TECNOGAR. This work has also been partially funded by the Universitat Politècnica de València through the post-doctoral PAID-10-20 program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No dataset has been created.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. García, L.; Parra, L.; Jimenez, J.M.; Lloret, J.; Lorenz, P. IoT-Based Smart Irrigation Systems: An Overview on the Recent Trends on Sensors and IoT Systems for Irrigation in Precision Agriculture. *Sensors* **2020**, *20*, 1042. [CrossRef]
2. Ueyama, J.; Freitas, H.; Faical, B.S.; Filho, G.P.R.; Fini, P.; Pessing, G.; Gomes, P.H.; Villas, L.A. Exploiting the use of unmanned aerial vehicles to provide resilience in wireless sensor networks. *IEEE Commun. Mag.* **2014**, *52*, 81–87. [CrossRef]
3. Faiçal, B.S.; Pessin, G.; Filho, G.P.R.; Carvalho AC PL, F.; Gomes, P.H.; Ueyama, J. Fine-Tuning of UAV Control Rules for Spraying Pesticides on Crop Fields: An Approach for Dynamic Environments. *Int. J. Artif. Intell. Tools* **2016**, *25*, 1660003. [CrossRef]
4. Faiçal, B.S.; Pessin, G.; Filho, G.P.R.; Furquim, G.; Carvalho, A.C.P.L.F.; Ueyama, J. Exploiting Evolution on UAV Control Rules for Spraying Pesticides on Crop Fields. In Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN 2014), Sofia, Bulgaria, 5–7 September 2014; pp. 49–58.
5. Instituto de Hidrología Meteorología y Estudios Ambientales. Métodos de Medición de Humedad del Suelo: Del Campo al Laboratorio. In PortalFruticola Web Site. 2018. Available online: https://www.portalfruticola.com/noticias/2018/08/30/metodos-de-medicion-de-humedad-del-suelo-del-campo-al-laboratorio/#:~{}:text=El%20m%C3%A9todo%20gravim%C3%A9trico%20es%20el,una%20temperatura%20de%20105\T1\textquoterightC (accessed on 20 September 2021).
6. Martini, E.; Werban, U.; Zacharias, S.; Pohle, M.; Dietrich, P.; Wollschläger, U. Repeated electromagnetic induction measurements for mapping soil moisture at the field scale: Validation with data from a wireless soil moisture monitoring network. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 495–513. [CrossRef]
7. Garcia, L.; Parra, L.; Jimenez, J.M.; Parra, M.; Lloret, J.; Mauri, P.V.; Lorenz, P. Deployment Strategies of Soil Monitoring WSN for Precision Agriculture Irrigation Scheduling in Rural Areas. *Sensors* **2021**, *21*, 1693. [CrossRef] [PubMed]

8. Ojha, T.; Misra, S.; Raghuwanshi, N.S. Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges. *Comput. Electron. Agric.* **2015**, *118*, 66–84. [CrossRef]
9. Susha Lekshmi, S.U.; Singh, D.N.; Baghini, M.S. A critical review of soil moisture measurement. *Measurement* **2014**, *54*, 105.
10. Tumanski, S. Induction coil sensors—A review. *Meas. Sci. Technol.* **2007**, *18*, R31. [CrossRef]
11. Jawad, H.M.; Nordin, R.; Gharghan, S.K.; Jawad, A.M.; Ismail, M. Energy-Efficient Wireless Sensor Networks for Precision Agriculture: A Review. *Sensors* **2017**, *17*, 1781. [CrossRef] [PubMed]
12. Hamami, L.; Nassereddine, B. Application of wireless sensor networks in the field of irrigation: A review. *Comput. Electron. Agric.* **2020**, *179*, 105782. [CrossRef]
13. Mekonnen, Y.; Namuduri, S.; Burton, L.; Sarwat, A.; Bhansali, S. Review—Machine Learning Techniques inWireless Sensor Network Based Precision Agriculture. *J. Electrochem. Soc.* **2020**, *167*, 037522. [CrossRef]
14. Nabi, F.; Jamwal, S.; Padmanbh, K. Wireless sensor network in precision farming for forecasting and monitoring of apple disease: A survey. *Int. J. Inf. Technol.* **2020**. [CrossRef]
15. Kabashi, A.H.; Elmirghani, J.M.H. Technical Framework for Designing Wireless Sensor Networks for Agricultural Monitoring in Developing Regions. In Proceedings of the Second International Conference on Next Generation Mobile Applications, Services, and Technologies, Cardiff, UK, 16–19 September 2008.
16. Kassim, M.R.M.; Mat, I.; Harun, A.N. Wireless Sensor Network in precision agriculture application. In Proceedings of the 2014 International Conference on Computer, Information and Telecommunication Systems (CITS), Jeju, Korea, 7–9 July 2014.
17. Zervopoulos, A.; Tsipis, A.; Alvanou, A.G.; Bezas, K.; Papamichail, A.; Vergis, S.; Stylidou, A.; Tsoumanis, G.; Komianos, G.; Koufoudakis, G.; et al. Wireless Sensor Network Synchronization for Precision Agriculture Applications. *Agriculture* **2020**, *10*, 89. [CrossRef]
18. Bayrakdar, M.E. A Smart Insect Pest Detection Technique with Qualified Underground Wireless Sensor Nodes for Precision Agriculture. *IEEE Sens. J.* **2019**, *19*, 10892–10897. [CrossRef]
19. Khedo, K.K.; Hosseny, M.R.; Toonah, M.Z. PotatoSense: A Wireless Sensor Network System for Precision Agriculture. In Proceedings of the 2014 IST-Africa Conference Proceedings, Pointe aux Piments, Mauritius, 7–9 May 2014.
20. Rasooli, M.W.; Bhushan, B.; Kumar, N. Applicability of Wireless Sensor Networks & IoT in Saffron & Wheat Crops: A Smart Agriculture Perspective. *Int. J. Sci. Technol. Res.* **2020**, *9*, 2456–2461.
21. Chaudhary, D.D.; Nayse, S.P.; Waghmare, L.M. Application of wireless sensor networks for greenhouse parameter control in precision agriculture. *Int. J. Wirel. Mob. Netw. IJWMN* **2011**, *3*, 140–149. [CrossRef]
22. Srbinovska, M.; Gavrovski, C.; Dimcev, V.; Krkoleva, A.; Borozan, V. Environmental parameters monitoring in precision agriculture using wireless sensor networks. *J. Clean. Prod.* **2015**, *88*, 297–307. [CrossRef]
23. Hamouda, Y.; Msallam, M. Variable sampling interval for energy-efficient heterogeneous precision agriculture using Wireless Sensor Networks. *J. King Saud Univ.—Comput. Inf. Sci.* **2020**, *32*, 88–98. [CrossRef]
24. Qureshi, K.N.; Bashir, M.U.; Lloret, J.; Leon, A. Optimized Cluster-Based Dynamic Energy-Aware Routing Protocol for Wireless Sensor Networks in Agriculture Precision. *J. Sens.* **2020**, *2020*, 9040395. [CrossRef]
25. Badewa, E.; Unc, A.; Cheema, M.; Kavanagh, V.; Galagedara, L. Soil moisture mapping using multi-frequency and multi-coil electromagnetic induction sensors on managed Podzols. *Agronomy* **2018**, *8*, 224. [CrossRef]
26. DeRouin, A.J.; You, Z.; Hansen, M.; Diab, A.; Ong, K.G. Development and application of the single-spiral inductive-capacitive resonant circuit sensor for wireless, real-time characterization of moisture in sand. *J. Sens.* **2013**, *2013*, 894512. [CrossRef]
27. Kizito, F.; Campbell, C.S.; Campbell, G.S.; Cobos, D.R.; Teare, B.L.; Carter, B.; Hopmans, J.W. Frequency, electrical conductivity and temperature analysis of a low-cost capacitance soil moisture sensor. *J. Hydrol.* **2008**, *352*, 367–378. [CrossRef]
28. Nor, A.S.M.; Yunus, M.A.M.; Nawawi, S.W.; Ibrahim, S. Low-cost sensor array design optimization based on planar electromagnetic sensor design for detecting nitrate and sulphate. In Proceedings of the 2013 Seventh International Conference on Sensing Technology (ICST), Wellington, New Zealand, 3–5 December 2013; pp. 693–698.
29. Fritz, S.; See, L.; Bayas, J.C.L.; Waldner, F.; Jacques, D.; Becker-Reshef, I.; Whitcraft, A.; Baruth, B.; Bonifacio, R.; Crutchfield, J.; et al. A comparison of global agricultural monitoring systems and current gaps. *Agric. Syst.* **2019**, *168*, 258–272. [CrossRef]
30. Group on Earth Observations Global Agricultural Monitoring Initiative (GEOGLAM). In Earth Observations Website. Available online: https://earthobservations.org/geoglam.php (accessed on 20 September 2021).
31. CropWatch. Available online: http://www.cropwatch.com.cn/ (accessed on 20 September 2021).
32. Shafi, U.; Mumtaz, R.; García-Nieto, J.; Hassan, S.A.; Zaidi, S.A.R.; Iqbal, N. Precision agriculture techniques and practices: From considerations to applications. *Sensors* **2019**, *19*, 3796. [CrossRef]
33. Khan, W.Z.; Ahmed, E.; Hakak, S.; Yaqoob, I.; Ahmed, A. Edge computing: A survey. *Future Gener. Comput. Syst.* **2019**, *97*, 219–235. [CrossRef]
34. Trumper, R.; Gelbman, M. Investigating electromagnetic induction through a microcomputer-based laboratory. *Phys. Educ.* **2000**, *35*, 90. [CrossRef]
35. Kholmetskii, A.L. The Faraday induction law and field transformations in special relativity. *Apeiron* **2003**, *10*, 118.
36. Monk, P. *Finite Element Methods for Maxwell's Equations*; Oxford University Press: Oxford, UK, 2003.
37. ESP32 DevKit Features. In Espressif Website. Available online: https://www.espressif.com/en/products/devkits (accessed on 20 September 2021).

38. The national association for amateur Radio. *The ARRL Handbook for Radio Communications 2017*, 94th ed.; Amer Radio Relay League: Newington, CT, USA, 2017.
39. Parra, L.; Sendra, S.; Lloret, J.; Bosch, I. Development of a conductivity sensor for monitoring groundwater resources to optimize water management in smart city environments. *Sensors* **2015**, *15*, 20990–21015. [CrossRef]
40. Texas Instruments. 555 IC Datasheet. Available online: http://www.ti.com/lit/ds/symlink/ne555.pdf (accessed on 20 September 2021).
41. ICM7555 Datasheet. Available online: https://www.nxp.com/docs/en/data-sheet/ICM7555.pdf (accessed on 20 September 2021).
42. RK520-02 Soil Moisture Sensor, Temperature Probe& EC Sensor. Available online: https://www.rikasensor.com/rk520-02-soil-moisture-sensor-temperature-probe-ec-sensor.html (accessed on 20 September 2021).
43. RS485 Soil Moisture & Temperature & EC Sensor. Available online: https://files.seeedstudio.com/products/101990667/res/RS485%20Soil%20Moisture%20&%20Temperature%20&%20EC%20Sensor%20(S-Temp&VWC&EC-02A).pdf (accessed on 20 September 2021).
44. SenseCAP LoRaWAN Sensor User Manual. Available online: https://sensecap-docs.seeed.cc/pdf/SenseCAP%20LoRaWAN%20Sensor%20User%20Manual-V1.1.pdf (accessed on 20 September 2021).
45. Waterproof_Capacitive_Soil_Moisture_Sensor_SKU_SEN0308-DFRobot. Available online: https://docs.rs-online.com/8389/A700000007238462.pdf (accessed on 20 September 2021).
46. YL-69. Available online: https://www.electronicoscaldas.com/datasheet/OBSoil-01_ElecFreaks.pdf (accessed on 20 September 2021).
47. Keyes Brick Soil Moisture Sensor Module. Available online: https://www.fasttech.com/product/9720737-keyes-brick-soil-moisture-sensor-module (accessed on 20 September 2021).
48. KeeYees moisture sensor. Available online: https://www.amazon.com/KeeYees-Capacitive-Moisture-Sensor-Module/dp/B07R174TM1 (accessed on 20 September 2021).
49. Grove—Capacitive Soil Moisture Sensor (Corrosion Resistant). Available online: https://www.seeedstudio.com/Grove-Capacitive-Moisture-Sensor-Corrosion-Resistant.html (accessed on 20 September 2021).
50. Grove—Soil Moisture Sensor. Available online: https://www.seeedstudio.com/Grove-Moisture-Sensor.html (accessed on 20 September 2021).
51. RS485 Soil Moisture & Temperature Sensor (S-Soil MT-02A). Available online: https://files.seeedstudio.com/products/101990668/res/RS485%20Soil%20Moisture%20&%20Temperature%20Sensor%20(S-Soil%20MT-02)-Datasheet.pdf (accessed on 20 September 2021).
52. Soil Moisture Sensor. Available online: https://www.makerfabs.com/soil-moisture-sensor.html (accessed on 20 September 2021).
53. Xin, H.M.; Yang, K. Routing protocols analysis for Internet of Things. In Proceedings of the 2nd International Conference on Information Science and Control Engineering (CISCE 2015), Shanghai, China, 24–26 April 2015; pp. 447–450.
54. Garcia, M.; Sendra, S.; Lloret, J.; Canovas, A. Saving energy and improving communications using cooperative group-based wireless sensor networks. *Telecommun. Syst.* **2013**, *52*, 2489–2502. [CrossRef]
55. Sendra, S.; García, L.; Lloret, J.; Bosch, I.; Vega-Rodríguez, R. LoRaWAN Network for Fire Monitoring in Rural Environments. *Electronics* **2020**, *9*, 531. [CrossRef]
56. Nikolidakis, S.A.; Kandris, D.; Vergados, D.D.; Douligeris, C. Energy efficient automated control of irrigation in agriculture by using wireless sensor networks. *Comput. Electron. Agric.* **2015**, *113*, 154–163. [CrossRef]
57. Navarro-Hellín, H.; Torres-Sánchez, R.; Soto-Valles, F.; Albaladejo-Pérez, C.; López-Riquelme, J.A.; Domingo-Miguel, R. A wireless sensors architecture for efficient irrigation water management. *Agric. Water Manag.* **2015**, *151*, 64–74. [CrossRef]

sensors

MDPI

Review

Towards Synoptic Water Monitoring Systems: A Review of AI Methods for Automating Water Body Detection and Water Quality Monitoring Using Remote Sensing

Liping Yang [1,2,3,*], Joshua Driscol [1,2], Sarigai Sarigai [1,2], Qiusheng Wu [4], Christopher D. Lippitt [1,2] and Melinda Morgan [1]

[1] Department of Geography and Environmental Studies, University of New Mexico, Albuquerque, NM 87131, USA; joshuadr@unm.edu (J.D.); rsa@unm.edu (S.S.); clippitt@unm.edu (C.D.L.); mhbenson@unm.edu (M.M.)
[2] Center for the Advancement of Spatial Informatics Research and Education (ASPIRE), University of New Mexico, Albuquerque, NM 87131, USA
[3] Department of Computer Science, University of New Mexico, Albuquerque, NM 87106, USA
[4] Department of Geography, University of Tennessee, Knoxville, TN 37996, USA; qwu18@utk.edu
* Correspondence: lipingyang@unm.edu

Abstract: Water features (e.g., water quantity and water quality) are one of the most important environmental factors essential to improving climate-change resilience. Remote sensing (RS) technologies empowered by artificial intelligence (AI) have become one of the most demanded strategies to automating water information extraction and thus intelligent monitoring. In this article, we provide a systematic review of the literature that incorporates artificial intelligence and computer vision methods in the water resources sector with a focus on intelligent water body extraction and water quality detection and monitoring through remote sensing. Based on this review, the main challenges of leveraging AI and RS for intelligent water information extraction are discussed, and research priorities are identified. An interactive web application designed to allow readers to intuitively and dynamically review the relevant literature was also developed.

Keywords: surface water; water body detection; surface water extraction; water quality monitoring; remote sensing; artificial intelligence; computer vision; machine learning; deep learning; convolutional neural networks

Citation: Yang, L.; Driscol, J.; Sarigai, S.; Wu, Q.; Lippitt, C.D.; Morgan, M. Towards Synoptic Water Monitoring Systems: A Review of AI Methods for Automating Water Body Detection and Water Quality Monitoring Using Remote Sensing. *Sensors* **2022**, *22*, 2416. https://doi.org/10.3390/s22062416

Academic Editors: Miquel À. Cugueró-Escofet and Vicenç Puig

Received: 2 January 2022
Accepted: 15 March 2022
Published: 21 March 2022

1. Introduction and Motivation

Water is fundamentally necessary to all forms of life, and it is also the primary medium through which climate change impacts Earth's ecosystem and thus the livelihood and wellbeing of societies [1]. While water covers about 71% of the Earth's surface, only approximately 3% of the Earth's water bodies are freshwater [2]. Climate change will bring unique challenges to these water bodies. Many rivers and streams are heavily dependent on winter snowpack, which is declining with rising temperatures and changing precipitation patterns [3]. Sea level rise is also impacting the continued quality and quantity of water supplies [4]. Both the quantity and the quality of freshwater systems are critical environmental features essential to increasing resilience in the face of climate change [5,6]. Resilience is defined here as the capacity of a system to absorb disturbance and still retain its basic function and structure [7]. Climate change will bring new disturbances in many forms, including increased pollution from wildfires, saltwater intrusion, and deteriorated water quantity resulting from prolonged drought [1,8]. It is critical that we gather, ideally automatically, as much information as possible about freshwater bodies and how they function in order to increase our capacity to respond to a changing climate. Rockström [5,6] and his colleagues conceptualize freshwater use and biogeochemical flows that threaten the integrity of freshwater (via pollution) as two of seven variables key to overall Earth

system function. Each of these variables, they argue, can be thought of as having "planetary boundary", a threshold that should not be crossed if we are to maintain the Earth in its current system state [5]. In this sense, the integrity and functioning of freshwater systems are essential not only in the local scale in which they provide critical ecosystem services; they also create a "safe operating space" for humanity as a whole, as we seek to achieve global solutions to the larger environmental challenges we face with climate change and associated stressors [6].

Responding to climate change challenge impacts on water resources requires adaptation strategies at the local, regional, national, and global scales. Countries are urged to improve their water resources management systems and to identify and implement "no regrets" strategies in order to be resilient to climate change [1]. The changing spatial and temporal patterns of surface water are important, in both practical and scientific terms, for water resources management, biodiversity, emergency response, and climate change [9]. More specifically, automated monitoring of water bodies is critical for adapting to climate change, water resources, ecosystem services, and the hydrological cycle, as well as for urban hydrology, which can facilitate timely flood protection planning and water quality control for public safety and health [10–12]. Accurate water quality monitoring is essential for developing sustainable water resource management strategies and ensuring the health of communities, ecosystems, and economies [13]. However, current knowledge of water quality is often disconnected in time and space across different measurement techniques and platforms that may fail to capture dynamic ecosystem changes. This disconnection indicates an inefficiency and redundancy in research and monitoring activities. A major challenge for water resource management is how to integrate multiple sources of water quality data and indices into usable and actionable information of environmental, social, economic, and infrastructural value [13,14].

Geospatial big data are leading to transformative changes in science (with the advent of data-driven and community science) and in society (with the potential to support the economy, public health, and other advances). Artificial intelligence (AI), especially its branches machine learning (ML), deep learning (DL), and computer vision (CV), are central to leveraging geospatial big data for applications in both domains. Remote sensing (RS) is the single largest source of geospatial big data and has increased dramatically in terms of both spatial and temporal resolution. This poses serious challenges for effective and efficient processing and analysis [15]. Meanwhile, recent advances in DL and CV have significantly improved research in RS and geosciences [16–18]. These advances, if integrated in creative and appropriate ways, host potential to enable the automated identification and monitoring of large-scale water bodies and water quality effectively and efficiently.

In this article, we argue specifically that bridging research into extracting important water information (e.g., water body extent, water quality) from RS imagery will provide an important computational foundation for the development of smart, RS-enabled water resource management systems. We review a range of recent developments in the relevant fields that can be leveraged to support intelligent automation of water body extraction and water quality detection and monitoring through RS imagery. An accompanying interactive web application allows our readers to intuitively track scholars and publications covered in this review (the web app tool URL and its brief demo video link are provided in Appendix A).

1.1. Selection Criterion for Reviewed Papers and Brief Graphic Summary

In the literature review process, we performed a systematic search on Google Scholar with the keywords and search strategy detailed in Table 1. In addition, our search was restricted to research articles published in English and in peer-reviewed journals or conference proceedings. For water body detection, we combined the water body keywords with some combination of the general keywords. The process for finding publications related to water quality was the same, only with the water quality keywords list. Beyond the keywords listed in this table, references (those cited in the papers we reviewed) cited by

the keyword-identified papers were also retained. A total of 90 papers relevant to the topic of water body and/or water quality from RS imagery using AI/ML/DL/CV algorithms were identified. A total number of 56 highly relevant articles were identified by applying the following exclusion criteria: (1) papers related to plastic pollution and sewage/water treatment plants, (2) precipitation forecasting or groundwater detection (as it is not intuitive to detect groundwater from RS imagery), and (3) general land use classification. Figure 1 shows the spatial distribution and a simple statistics summary of the papers covered in this review, where (d) shows the number of published papers by year in the reviewed topics from 2011 to early 2022.

Table 1. Keywords used for article search.

Keyword Category	Search Strategy
General keywords [1]	"remote sensing" OR "satellite data" OR "UAV" AND "computer vision" OR "machine learning" OR "deep learning" OR "neural networks" OR "AI"
Water body	"water body" AND "detection" OR "extraction"
Water quality	"water quality" AND "sensing" OR "monitoring"

[1] A list of general keywords were combined with either the category of water body or water quality, respectively, to perform our search.

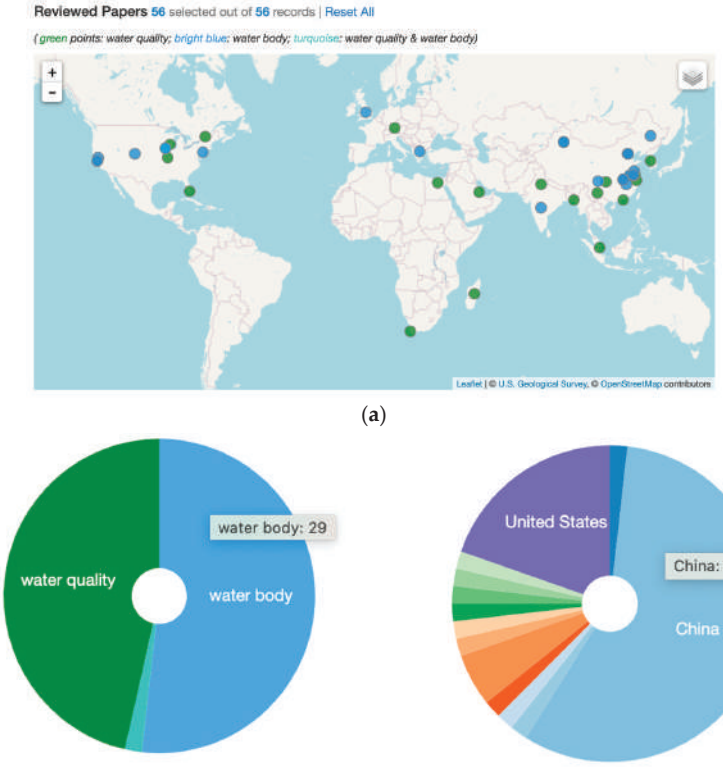
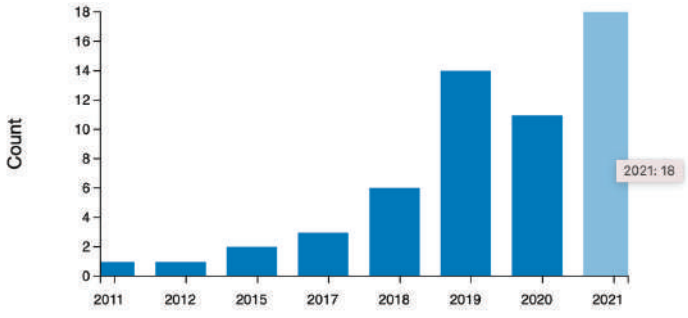


Figure 1. *Cont.*



(**d**)

Figure 1. Geospatial distribution and simple statistics of the reviewed papers. Note that a freely accessible interactive version of the charts can be accessed via our web app tool (the web app tool URL and its brief demo video are provided in Appendix A). We can easily see that the major countries are China and the United States and that the number of published papers by year (2011 to 2021) has dramatically increased since 2018 and 2019. (**a**) Spatial distribution of reviewed papers based on the first author's institution location. (**b**) Topic distribution (water body, water quality, both). (**c**) Country distribution. (**d**) Number of published papers by year from 2011 to 2021 on the relevant topics.

1.2. Roadmap

Here, we provide a roadmap for the rest of the paper. Section 2 outlines the scope of this review and our intended audience. Section 3 is the core of the paper, focused on identifying important and recent developments and their implications to water body detection and water quality monitoring from RS imagery through the leverage of AI/ML/DL/CV. Here, we highlight recent advances in several subfields of AI that water domains and RS can leverage. Specifically, we provide general characteristics of the reviewed studies using word clouds (Section 3.1). We then examine and appraise key components of influential work in water body detection (Section 3.2) and water quality monitoring (Section 3.3). Section 4 starts with a brief summary (Section 4.1), followed with a discussion of key technical challenges (Section 4.2) and opportunities (Section 4.3). The paper concludes in Section 5.

To allow our readers to intuitively and dynamically review the relevant literature, we have developed a free-of-charge interactive web app tool (the web app URL and its brief demo video are provided in Appendix A). To provide background for readers (particularly those from water resources and RS) who are new to AI/ML/DL/CV, we introduce essential ML terms in Appendix B. As evaluation metrics are essential for measuring the performance of AI/ML/DL/CV models, we also provide an introduction to a set of commonly used evaluation metrics in Appendix C. In addition, as there are plenty of acronyms in this paper, we provide a full list of abbreviations right before the appendices.

2. Audience and Scope

It is important to know where water is and how its extent and quality are changing over time in a quick and accurate manner. Water quality is a key issue in water supply, agriculture, human and animal health, and many other areas [19]. Impaired water quality can be caused by natural disasters, but the most common cause is anthropogenic pollution. Pollutants, excessive nutrients from fertilizers, and sediment (e.g., from soil erosion) are carried into local lakes and rivers via runoff from urban or agricultural areas [19,20]. The quality of water varies from places and from time to time [19]. Affected surface waters are present in RS imagery and can be identified with the help of computational techniques such as ML. *To make near real-time intelligent water body detection and water quality monitoring possible, we need to first detect extent of water bodies from RS imagery, from which volume can be*

computed, and then recognize their corresponding water quality, eventually linking the two to allow water quality monitoring.

Environmental nonprofits, government agencies, and water managers need access to this type of integrated spatial–time series of water body and water quality information to see how local water resources are changing and plan for future drought conditions. Collective detection and monitoring of water bodies and their associated water quality has applications for human health, as well as to private-sector industries including timber, agriculture, recreation, and tourism. Public policy planners need to be better informed as they make environmental preservation and restoration decisions based on changing water availability, and with this data we can be better equipped to monitor water quality that can quickly change due to floods, hurricanes, or human-caused pollution, *and yet, to date, water body detection and water quality monitoring research has been historically separate and does not focus enough on producing intuitive, operational products*.

Building on the long-term interest in ML and CV within the RS community, the main goals of this review paper are to (1) survey recent advances in water body detection and water quality monitoring from RS data using AI to identify commonly cited challenges in order to provide suggestions for new research directions, and (2) move towards automated, synoptic water quantity and quality monitoring to inform more robust water resource management.

This systematic review is relevant to multiple research domains, including, but not limited to RS, geographic information science, computer science, data science, information science, geoscience, hydrology, and water resource management. This paper does not attempt to review the application of RS to water resources and hydrology more generally; for recent reviews of these topics, see [13,21–24]. A survey of DL applications in hydrology and water resources can be found in [25]; a survey of AI in the water domain can be found in [26]; and a survey of water quality applications using satellite data solely focused on ML can be found in [27]. This review focuses on investigating recent AI methods, including its branches ML, DL, and CV, for water information extraction (specifically water body detection and/or water quality monitoring) from RS imagery. Our review has a narrowed scope in water resources and hydrological research, but a wider and deeper scope in terms of AI methods and metrics used to assess models in both water body detection and water quality research. *By integrating both domains, we hope to develop a basis for effective computational frameworks for intelligent water monitoring systems using RS and AI.*

3. The State of the Art: Advances in Intelligent Waterbody Information Extraction

3.1. General Characteristics of the Reviewed Studies

Note that we only included and reviewed the papers that use both RS and AI/ML/DL/CV for water body and/or water quality detection (that is, the number of papers cited in our reference section is much larger than the number of papers we review in this Section 3). A word cloud visualization of the titles, abstracts, and keywords of the reviewed 56 papers are provided in Figure 2, where the top figure indicates the word cloud for all reviewed papers. The bottom left word cloud is for reviewed water body papers, and the bottom right for reviewed water quality papers.

As we can see from the word cloud for both water body extraction and water quality (see the top word cloud in Figure 2), “remote sensing”, “deep learning”, “prediction”, “classification”, “extraction”, “machine learning”, “water body”, “water quality”, and “convolutional neural network” are prominent concepts and words captured by the word cloud. Our focus is on studies that use RS for water body extraction and water quality monitoring, so many of the keywords are to be expected. However, it is perhaps surprising to see DL featured so prominently given that the shift from ML to DL models is a relatively recent phenomenon.

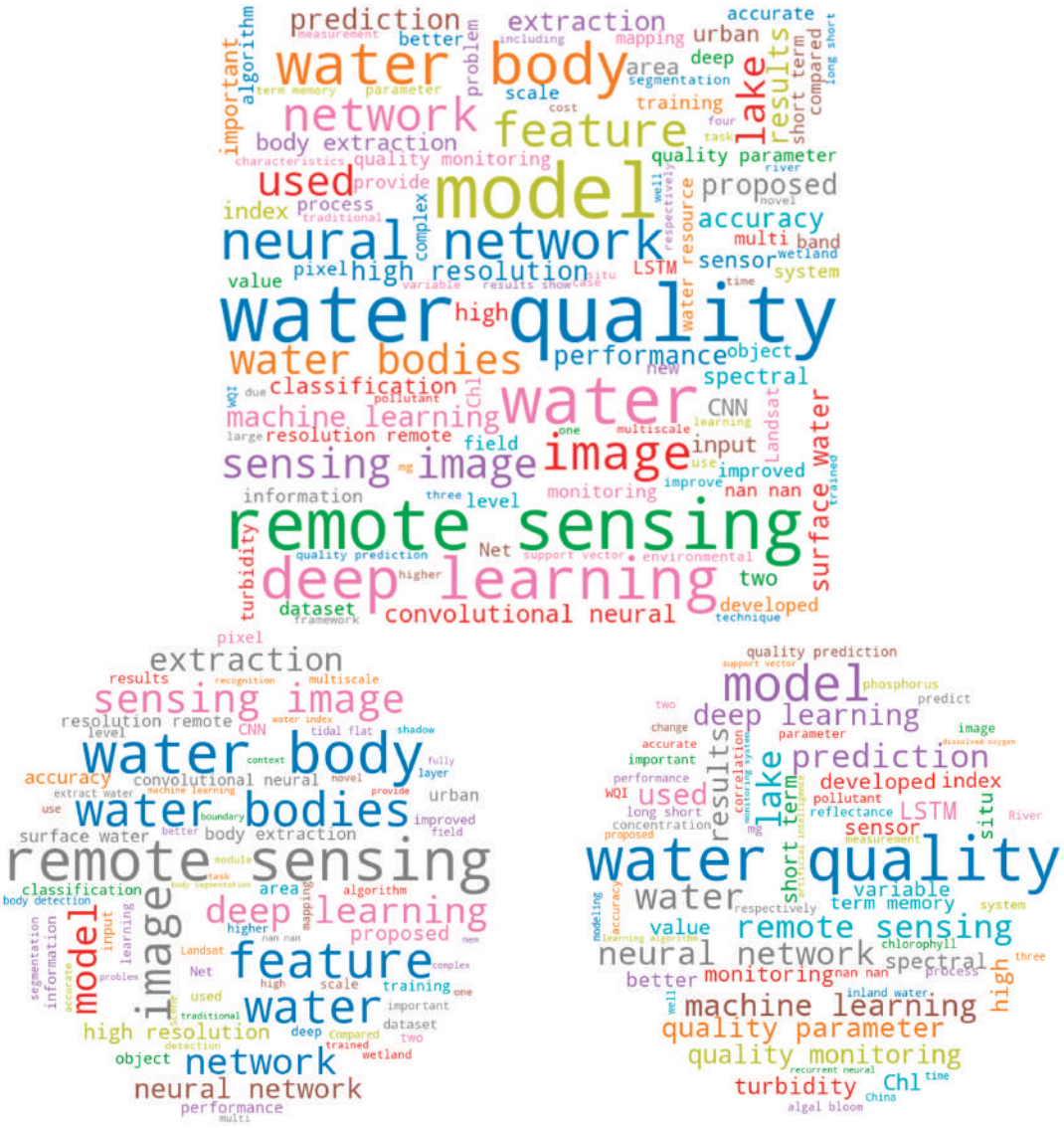


Figure 2. Word cloud visualization of all the reviewed papers (**top**), water body papers (**bottom left**), and water quality papers (**bottom right**). Note that the word clouds are generated from paper titles, abstracts, and keywords. The word clouds provide an informative (general and specific) focus of each set of the papers. For example, both water body and water quality papers share the focus on RS, DL, and neural networks (NN). We can also see that water body extraction tasks tend to focus on the use of convolutional neural networks (CNN), whereas for water quality modeling the use of long short-term memory (LSTM) networks is more prevalent. We can also see that there are specific, unique keywords for water quality, such as "turbidity", "chl", and "algal bloom".

When we separate the keyword word clouds (see the bottom two word clouds in Figure 2), this trend becomes clearer. Deep learning is much more common in water body extraction, whereas in the word cloud for water quality monitoring, "neural network" and "machine learning" are about the same size. Additionally, in the water body extraction

word cloud, "remote sensing" is featured much more heavily than it is in the water quality extraction literature. In our review, the water quality papers often involved other types of data, including in situ sensors or smaller RS devices (not satellites), whereas the water body extraction literature is dominated by RS imagery. This is related to the scale of projects in the two domains: water body extraction is usually undertaken across large spatial scales, whereas the water quality monitoring literature is still focused on smaller, often individual, bodies of water. This points to a future research direction in the water quality literature that we touch on in our review paper: we need to scale up water quality estimation using RS imagery by matching it with ground-truth water quality measurements.

Tables 2 and 3 provide a brief summary of the methods used for water body detection and water quality monitoring, elaborated in Sections 3.2 and 3.3, respectively. The general characteristics summarized by machines (i.e., the word clouds in Figure 2) align with the literature; convolutional neural network (CNN) models are indeed applied much more frequently for water body detection, and long short-term memory (LSTM) models are often used for water quality monitoring. The evaluation metrics used in the reviewed articles were also summarized and are provided in Tables 2 and 3 (a brief explanation of each metric is in Appendix C).

Table 2. Studies targeting water body detection from RS imagery using AI (note that it is ordered chronologically to show trends in data type and model usage; see the Abbreviations for a list of the acronyms).

Reference	Method	Model Comparison	RS Data Type	Evaluation Metrics
Li et al. (2011) [28]	DNN	NDWI	Landsat TM	coherence
Yang et al. (2015) [29]	AE	DNN, SVM	Landsat ETM+	accuracy
Huang et al. (2015) [30]	ELM	DT, LORSAL, RF, SVM, TB	GeoEye-1, WorldView-2	Kappa, F-score
Isikdogan et al. (2017) [31]	CNN	MDWI, MLP	Landsat	F1-score, CE, OE, precision, recall
Yu et al. (2017) [32]	CNN–LR hybrid	ANN, CNN, SVM	Landsat ETM+	accuracy
Jiang et al. (2018) [10]	MLP	MLC, NDWI	Landsat-8 OLI	Kappa, OA
Chen et al. (2018) [33]	CNN	CNN, NDWI, SVM	GaoFen-2, Zi Yuan-3	ECE, EOA, EOE, OA, PA, UA
Miao et al. (2018) [34]	CNN	DNN	Google Earth imagery	OA
Acharya et al. (2019) [35]	SVM	ANN, DT, GMB, NB, NDWI, RF, RPART	Landsat-8 OLI	Kappa, OA
Feng et al. (2019) [36]	CNN–CRF hybrid	CNN, CV-method	GaoFen-2, WorldView-2	Kappa, PCC, precision
Li et al. (2019) [37]	CNN	CNN, NDWI, SVM	GaoFen-2	F1-score
Li et al. (2019) [38]	CNN–CRF hybrid	CNN, NDWI	GaoFen-1	IoU, pixel accuracy, recall
Meng et al. (2019) [39]	CNN–SVM hybrid	CNN, SVM	GaoFen-2	accuracy, MA, UA
Isikdogan et al. (2020) [40]	CNN	CNN, MLP, MNDWI	Landsat-8	F1-score, precision, recall
Song et al. (2020) [41]	CNN	CART, KNN, RF, SVM	GaoFen-2, WorldView-3	IoU, precision, recall
Yang et al. (2020) [42]	CNN	CNN	GaoFen-2	IoU
Wang et al. (2020) [43]	CNN	CNN, NDWI	GaoFen-1	F1-score, mIoU, precision, recall
O'Neil et al. (2020) [44]	CNN	DEM, NDVI, RF	LiDAR DEMs, NAIP	precision, recall
Chen et al. (2020) [45]	CNN	NDWI, SVM	GaoFen-1, GaoFen-2, Zi Yuan-3	BOA, Kappa, OA
Dang and Li (2021) [46]	CNN	CNN	GaoFen-2, GID	mIoU, FWIoU, OA
Yuan et al. (2021) [47]	CNN	CNN, MNDWI, NDMI, NDWI	Sentinel-2	accuracy, mIoU
Tambe et al. (2021) [48]	CNN	CNN	Landsat-8 OLI	CA, F1-score, GA, IoU, precision, recall
Yu et al. (2021) [49]	CNN	CNN	GaoFen-2, Landsat-7	F1-score, OA, precision, recall
Li et al. (2021) [50]	CNN	CNN, CV-method, SVM	UAV	Kappa, F-score, OA, precision

Table 2. *Cont.*

Reference	Method	Model Comparison	RS Data Type	Evaluation Metrics
Zhang et al. (2021) [51]	CNN	CNN, MLC, NDWI, SVM	GaoFen-2	IoU, Kappa, pixel accuracy
Li et al. (2021) [52]	CNN	CNN, NDWI, SVM	GaoFen-2, GaoFen-6, Sentinel-2, Zi Yuan-3	F1-score, IoU, OA
Su et al. (2021) [53]	CNN	CNN	Landsat-8, Sentinel-2A	IoU, pixel accuracy, recall
Ovakoglou et al. (2021) [54]	KMeans	fuzzy-rules classification, Haralick's textural features of dissimilarity, Otsu valley-emphasis	Sentinel-1	Kappa, OA, precision, recall

Table 3. Studies targeting water quality monitoring from RS imagery using AI (where "/" means none. Note that it is ordered chronologically to show trends in data type and model usage) (See the Abbreviations for a full list of the acronyms).

Reference	Method	Model Comparison	RS Data Type	Evaluation Metrics
Chebud et al. (2012) [55]	DNN	/	Landsat TM	RMSE, R^2
Wang et al. (2017) [56]	SVR	index methods	spectroradiometer, water samples	RMSE, RPD, R^2
Lee and Lee (2018) [57]	LSTM	DNN, RNN	water quality time series	RMSE
Wang et al. (2019) [58]	LSTM	/	water quality time series	accuracy, cross-correlation
Pu et al. (2019) [59]	CNN	RF, SVM	Landsat-8	accuracy
Liu et al. (2019) [60]	LSTM	ARIMA, SVM	IoT data	MSE
Chowdury et al. (2019) [61]	MLP	/	IoT data	threshold value
Hafeez et al. (2019) [62]	DNN	CB, RF, SVR	Landsat	accuracy, relative variable importance
Li et al. (2019) [63]	RNN–DS hybrid	GRU, LSTM, SRN, SVR	water quality time series	MAE, MAPE, RMSE
Randrianiaina et al. (2019) [64]	DNN	/	Landsat-8	RMSE, R^2
Yu et al. (2020) [65]	LSTM	/	water quality time series	MAE, RMSE
Zou et al. (2020) [66]	LSTM	DNN, GRU, LSTM	meteorological time series, water quality time series	MAE
Peterson et al. (2020) [67]	ELR	MLR, SVR	Landsat-8, Sentinel-2	MAPE, RMSE, R^2
Hanson et al. (2020) [68]	LSTM	/	water quality time series	auto-correlation, MK statistics, RMSE
Barzegar et al. (2020) [69]	CNN–LSTM hybrid	CNN, LSTM	water quality data from multiprobe sensor	MAE, NSEC, Percentage of Bias, RMSE, Wilmott's index
Aldhyani et al. (2020) [70]	LSTM	ANN, DNN, KNN, NB, SVM	water quality time series	accuracy, F-score, MSE, precision, R, sensitivity, specificity
Li et al. (2021) [71]	RF	SVM	Sentinel-2 MSI	RMSE, RPD, R^2, Z-score
Sharma et al. (2021) [72]	CNN	CNN	UAV camera	precision, recall
Cui et al. (2021) [73]	CNN	KNN, index method, RF, SVM	Landsat-8, Sentinel-2	RPD, RMSE, R^2
Zhao et al. (2021) [74]	DNN	RBFNN	Landsat-8, water quality time series	MAE, MSE, R^2
Arias-Rodriguez et al. (2021) [75]	ELM	LR, SVR	Landsat-8, Sentinel-2 MSI, Sentinel-3 OLI	MAE, MSE, RMSE, R^2
Kravitz et al. (2021) [76]	DNN	KNN, RF, XGBoost	Landsat 8 OLI, Sentinel-2 MSI	MAPE, RMSE, RMSLE [1]
Sun et al. (2021) [77]	DNN	GPR, RF	proximal hyperspectral imager, water samples	accuracy, MRE, RMSE, R^2

[1] The authors use the abbreviation RMSELE for RMSLE in their paper (this might be a typographical error).

3.2. Recent Advances in Water Body Detection Using AI

From our systematic review (including Table 2), we provide a brief summary here about the recent advances in water body detection based on AI. (1) The most common satellite platforms were Landsat, GaoFen, Zi Yuan, WorldView, and Sentinel, although there were some manually annotated datasets. The use of UAVs and DEMs were noted but were not as common. (2) Precision, recall, overall accuracy (OA), F1-score, kappa, and intersection over union (IoU) are the most popular evaluation metrics for water body detection since it is mainly a classification task. (3) Convolutional neural networks (CNNs) are normally compared to normalized difference water index (NDWI) or another index-based method, some form of "shallow" ML model (e.g., random forest (RF), support vector machine (SVM)), or other CNN architectures). Below, we provide a more detailed review of the methods used for water body detection. As Table 2 and word clouds (see Figure 2) indicate, the dominant methods used in water body detection with AI are CNNs (Section 3.2.1). Beyond CNN-based methods, there are other methods including CNN hybrids (Section 3.2.2), artificial neural networks (ANN), multilayer perceptrons (MLP), dense neural networks (DNN), other DL methods (Section 3.2.3), and "shallow" ML based methods (Section 3.2.4).

3.2.1. CNN-Based Water Body Detection

CNN-based models are the dominant methods for water body detection, but each of them have addressed different challenges posed in water body detection from RS imagery. Based on our review, we identify the following five groups of use cases: (1) Addressing limitations of index-based methods; (2) sharpening blurred boundaries caused by CNNs; (3) Addressing spatial and spectral resolution challenges, which covers those methods that are able to recognize water body across scales, at multiple resolutions, from very high-resolution imagery, and/or integrating bands beyond RGB channels to use for CNN model training; (4) Robust detection of small/slender/irregular-shaped water bodies; (5) Others.

1. Addressing limitations of index-based methods:

Index methods (e.g., NDWI) are rule-based and fail to take advantage of context information. CNNs can overcome this, although they often blur boundaries in segmentation tasks because of the convolution operation [34]. A DenseNet was used in [43] for water feature extraction and the authors compared its performance with NDWI and several popular CNN architectures. While NDWI methods are quick, they are not as accurate as CNNs. The authors showed that DenseNet performed the best at distinguishing water from shadows and clouds. However, the authors argue that clouds often occlude optical imagery, so one way to improve their method is to combine it with microwave RS imagery.

The authors in [31] pointed out that index methods require careful calibration and that indices differ from place to place. They also suffer from false positives (from snow, ice, rock, shadows, etc.) and vary in different weather conditions (e.g., clouds). To overcome those limitations of index-based methods, the authors of [31] developed DeepWaterMap, which can classify water with high accuracy, even distinguishing it from snow, ice, shadow, and clouds. DeepWaterMap is able to classify land classes that are often misclassified as water (or vice versa); thus, it minimizes false positives during the classification process. Most importantly, the DeepWaterMap model also works across different terrains and in different weather conditions, although it is still affected by clouds. The same authors released a second version of the model, DeepWaterMap v2, in [40]. The major improvement from v1 is that the new version allows users to input large RS scenes without the need for tiling, and the authors made their network run efficiently with constant memory at inference time. This model should theoretically work across different sensor platforms as long as they have the visible, near-infrared, and shortwave infrared 1 and 2 bands, but will still sometimes classify clouds as water.

2. Sharpening blurred boundaries caused by CNNs:

CNN-based methods can overcome the limitations of index-based methods, as reported above in group (1) [34], but they often blur boundaries in segmentation tasks because of the convolution operation. To sharpen water body detection boundaries, in [34], a restricted receptive field deconvolution network (RRF DeconvNet) and a new loss function called edges weighting loss were proposed. However, the authors needed to retrain the entire network (which is very computationally expensive) instead of using transfer learning (TL).

Apart from blurring pixel boundaries, CNNs generally require many training parameters and very large training datasets to be successful. A novel convolution–inception block in a network, called W-Net, was proposed in [48], to extract water bodies from RS imagery. W-Net is able to train on fewer images compared with other CNN models and still extract water bodies accurately, and the authors pointed out that less computations are necessary due to use of inception layers. W-Net outperformed other CNN architectures, although the authors still needed to go through the time- and labor-intensive process of creating a dataset of manually annotating images.

3. Addressing resolution and band related challenges

High-resolution optical RS imagery allows for much finer detail in surface water body extraction. However, clouds and their shadows are often present in optical RS images [78]. The shadows (e.g., cloud shadows and building shadows) and water bodies share a very similar appearance in optical RS images. Therefore, water body extraction is not an easy task in the optical high-resolution RS images due to the limited spectral ranges (including blue, green, red, and near-infrared bands) and the complexity of low-albedo objects (cloud shadows, vegetation, and building shadows). Higher spatial resolution imagery often comes at the cost of less spectral channels and thus makes it difficult to extract features from complex scenes. To address this problem, a dense local feature compression (DLFC) was proposed [52] to extract bodies of water from RS imagery, and their DLFC outperformed other state-of-the-art (SOTA) CNNs, as well as an SVM and NDWI thresholding. Their results demonstrated that the DLFC is good at extracting slender water bodies and distinguishing water bodies from building shadows using multisensor data from multiple RS platforms.

TL and data augmentation (see Appendix B) are used in [37] to extract water bodies from satellite imagery. The authors showed that a CNN can outperform NDWI and an SVM in water body detection when the input data is very high resolution. There are tradeoffs, however, and the authors reported that the difficulty of hyperparameter tuning is one downside to using a CNN. A water body extraction NN, named WBE-NN, was proposed in [45] to extract water bodies from multispectral imagery at multiple resolutions while distinguishing water from shadows, and performed much better than NDWI, an SVM, and several CNN architectures. A self-attention capsule feature pyramid network (SA-CapsFPN) was proposed in [49] to extract water bodies from satellite imagery of different resolutions. SA-CapsFPN is able to recognize bodies of water across scales and different shapes and colors, as well as in varying surface and environmental conditions, although it is still entirely dependent on optical imagery as input to the CNN.

The novel MSResNet proposed in [46] learned from a large dataset of unlabeled RS imagery. MSResNet, in addition to being able to extract water bodies in an unsupervised manner, is able to recognize water bodies at multiple resolutions and of varying shapes. However, their network cannot distinguish water bodies from farms and barren areas. In addition, the CNN-based model name FYOLOv3, proposed in [51], is able to detect tidal flats at different resolutions. However, it does depend on a manually selected similarity threshold that introduces some subjectivity.

RGB band imagery is the primary focus in substantial research for water body extraction, but many more bands are available in RS imagery. A multichannel water body detection network (MC-WBDN) was created in [47], which fused the infrared and RGB channels and used them as input data for their CNN architecture. They demonstrated that when multispectral data is used, model performance for water body detection is increased and the model is more robust to lighting conditions. The proposed model MC-WBDN is

much more accurate than index-based methods such as NDWI, modified NDWI (MNDWI), and normalized difference moisture index (NDMI). MC-WBDN also outperforms other SOTA architectures such as U-Net and DeepLabV3+ for water body detection tasks. However, this method still relies on preprocessing data to make sure each input image is the same shape and free of clouds.

4. Robust detection of small/slender/irregular-shaped water bodies

Small water bodies are hard to extract from RS imagery. In [33], the authors designed a CNN (named SAPCNN), which is able to extract high-level features of water bodies from input data in a complex urban background. NDWI and SVMs cannot distinguish between water and shadows and their architecture's performance partly relies on visual inspection. Ref. [53] utilized a modified DeepLabv3+ architecture to extract bodies of water at different scales. Their focus is on extracting water bodies in urban RS images. Their network performed well on small bodies of water, but the model has problems identifying many of them because they were not properly annotated.

Mask-region-based CNNs (R-CNNs) have demonstrated success in detecting small and irregular shape water bodies. Song et al. (2020) [41] employed an R-CNN for water body detection from RS imagery, and their model outperforms many traditional ML models in identifying small water bodies and bodies of water with differing shapes. However, it is still difficult to deploy a trained NN model into a usable, production-ready form for water mapping applications. The authors reported that using NN output to create and update a vector map of water resources for stakeholders is challenging.

Yang et al. (2020) [42] also used a mask R-CNN to automate water body extraction. The authors argued that this allows them to avoid manual feature extraction in complex RS imagery. They segmented small water bodies and bodies of water with irregular shapes, although their methods suffer from poor IoU accuracy. This is primarily due to a small training set, for which DL models are ill-suited, and resulted in their models having problems identifying multiple bodies of water in RS images.

A self-attention capsule feature pyramid network (SA-CapsFPN) was proposed in [49] to extract water bodies from satellite imagery. SA-CapsFPN is able to recognize bodies of water across scales and different shapes and colors, as well as utilizing different information channels. The novel MSResNet proposed in [46], learnt from unlabeled large RS imagery, is also able to recognize water bodies at multiple resolutions and of varying shapes; however, their network cannot distinguish water bodies from farms and barren areas.

A dense local feature compression (DLFC) was proposed in [52] to extract bodies of water from RS imagery, and their DLFC outperformed other SOTA CNNs, as well as an SVM and an NDWI. Their results demonstrated that the DLFC is good at extracting slender water bodies and distinguishing water bodies from building shadows using multisensor data from multiple RS platforms.

5. Others

Extracting water bodies from RS imagery quickly and reliably is still a difficult task. Based on U-Net, [50] developed a new model called SU-Net to distinguish between water bodies, shadows, and mixed scenes. However, the authors only focused on water body extraction in urban areas and only used RGB information during the extraction process. While SU-Net performed better than an SVM and classic U-Net, it suffered when extracting water bodies from RS imagery with high reflectivity or that contained aquatic plants.

Wetlands are important ecosystems because they can keep flooding at bay and store carbon; however, they are threatened by development, climate change, and pollution. For the task of identifying wetlands, [44] combined RS imagery with hydrological properties derived from digital elevation models (DEMs) to identify wetlands. They showed that an RF performs as well as a CNN, although both models had issues distinguishing roads and trees from wetlands. This is perhaps due to their small training set. To improve performance, the authors argued that larger datasets with finer labels should be created for wetland detection.

Substantial water body detection work has focused on water bodies in urban and inland settings. Very few focus on tidal flat extraction, where sediment levels are high and the boundary of the water body itself is blurry. A CNN model called FYOLOv3 was proposed in [51], where the authors compared their model to NDWI, an SVM, a maximum likelihood classifier, U-Net, and YOLOv3. FYOLOv3 performed the best and is able to detect tidal flats at different resolutions; however, it depends on a manually-selected similarity threshold during the training process, which is a source of subjectivity.

Large sets of unlabeled water body data are available and easy to acquire, and semantic segmentation networks cannot recognize different water body shapes. A recent, very novel encoder–decoder CNN architecture named MSResNet, proposed in [46], is able to overcome those limitations. MSResNet is able to learn from unlabeled data and can also recognize water bodies of varying shapes and at multiple resolutions. However, even though their network outperforms other SOTA architectures without supervised training, their network has some issues categorizing water bodies, farms, and barren areas.

3.2.2. CNN Hybrid-Based Water Body Detection

CNNs are the SOTA models in water body extraction tasks (detailed in Section 3.2.1 above); however, their output and decisions for why they make the predictions that they do are largely a black box. Recent studies have integrated CNNs with some ML models. Interpretability was improved by using a CNN and SVM in parallel to classify wetland water bodies [39]. Wetlands are difficult/complex to identify in high-resolution satellite imagery with any single ML model. Hybrid models have shown promise in a process called decision fusion. Here, the authors pick a decision fusion threshold value by performing cross-validation on the CNN to see when it is sure or not. They then use this threshold value for the decision fusion predictions (e.g., when the CNN is not that sure, they defer to the SVM). However, the authors did not explain why they used an SVM and not some other ML model. The classifier used in [32] combines a CNN with a logistic regression (LR) model to extract water bodies. The authors emphasized that traditional ML methods for water body extraction need multispectral data and rely on lots of prior knowledge. Thus, those ML-based methods would not generalize well to different tasks. The authors also argue that single-band threshold methods are subjective. Their results demonstrated that the hybrid CNN-LR model works better than an SVM, an ANN, and other CNNs. However, their method requires segmented RS images as input.

How to accurately extract water bodies from RS images, while continuously updating the surface water maps, is an active research question. Index methods and active contour models are popular methods for water body detection tasks but are sensitive to subjective threshold values and starting conditions. Deep U-Net model was proposed to be used with a conditional random field (CRF) and regional restriction to categorize water versus non-water in satellite images [36], while reducing the blurring of edges that often occurs from CNNs for image segmentation. Although this network is highly accurate, it takes a lot of data and computation power to train. Training ML models at a single scale in single channels can cause errors when generalizing to other scales or types of RS data. Multiscale RS imagery was used with DeepLabV3+ and a CRF for water body segmentation [38]. This approach works well for training models on data from different scales, and they concluded that CNNs and CRFs together extract more accurate water boundaries at both large and small scales than CNNs alone.

3.2.3. ANN, MLP, DNN, and Other DL-Based Methods for Water Body Detection

An NN architecture called a local excitatory globally inhibitory oscillator network (LEGION) is used in [28], where the authors compared the results of LEGION trained on NDWI and spectral information, respectively. In addition, they employed object-wise classification, instead of pixel-based classification used in most other work. The authors reported that the network is very computationally expensive.

Different methods of water body extraction work (or do not work) in different areas/terrain types. Each needs subjective thresholds and/or hand-crafted features. In addition, generating large sets of labeled data is difficult and expensive, as high-dimension RS data is difficult to analyze. Objects such as shadows, clouds, and buildings are hard to distinguish from water bodies. In [29], the authors used an autoencoder for unsupervised training and concluded that their results are more accurate than for an SVM and traditional NN.

Huang et al., 2015 [30] pointed out that not many people have focused on water body detection in urban settings. This is a problem because water bodies often look similar to shadows due to buildings at certain times of the day in optical imagery. The authors employed an extreme learning machine (ELM), an SVM, a tree bagger (TB), and an RF to detect water bodies. The authors reported that the RF and TB performed much better than the SVM and ELM. However, their method depends on optical imagery with subjective thresholds set through trial and error. Specifically, their method depends on subjective threshold values in NDWI, normalized difference vegetation index (NDVI), and morphological shadow index (MSI).

Ref. [10] compared MLP, NDWI, and a maximum likelihood model for water body classification and showed that MLP performed the best. However, the maximum likelihood model could not recognize small bodies of water and thin rivers, whereas NDWI was not able to distinguish seawater from land. The MLP could identify small bodies of water better, but the analysis depended on visual assessment.

3.2.4. "Shallow" ML-Based Water Body Detection

Although most of the recent methods for water body detection used DL and/or deeper neural networks (Sections 3.2.1–3.2.3), a few studies used only "shallow" ML methods (e.g., RF and SVM). In [35], the authors used band methods (where slope, NDVI, and NDWI were added as three secondary bands to integrate extra information into ML training), and then applied an SVM, a decision tree (DT), and an RF to analyze multiband RS data for water body extraction in the Himalayas. However, while their models worked well for flat and hilly terrain, they had to parse out high elevations and snow in this method (which involves extra preprocessing and limits when/where their method can work with optical data). The authors ran different experiments to analyze which input bands (NDWI vs. individual input bands from Landsat data) worked the best but could only compare results visually. The authors concluded that adding single secondary bands is better than adding multiple in most ML algorithms except for NNs.

Sentinel-1 data and four different ML models (K-nearest neighbors classifier (KNN), fuzzy-rules classification, Haralick's textural features of dissimilarity, Otsu valley-emphasis) were employed to classify water bodies in [54]. It involved many different ML methods in tandem (i.e., the output of one ML model was fed into other processing steps), which complicates interpretability. This method did not have very high accuracy and did not work well in flooded regions, near buildings, and in the presence of aquatic vegetation. However, it was an important attempt to use synthetic aperture radar (SAR) data, which is rare in water body detection literature.

3.3. Recent Advances in Water Quality Monitoring Using AI

From Table 3, we identify the following trends in the use of AI for water quality monitoring research: (1) Water quality monitoring differs from water body detection in that it is formulated as both a classification and a regression task. Because of this, recurrent neural networks (RNNs), long short-term memory (LSTMs), and gated recurrent units (GRUs) are much more prevalent in the water quality literature. (2) Accuracy, precision, and recall are common metrics, as are some variations of mean squared error (MSE) and R^2. (3) It is important to note that while water body detection papers describe integrating multiple data sources into one analysis, this practice is much more common in water quality monitoring research. This primarily takes the form of trying to match up water quality parameters from time series data or water samples to optical satellite RS imagery. In water

quality monitoring, it is much more common to utilize Internet of Things (IoT) sensors, smaller probes such as unmanned aerial vehicle (UAVs) and stationary hyperspectral imagers, as well as government and private water quality time series data. (4) Some studies do not compare their model to any other models (detailed in Table 3), making it difficult to fully assess their methodologies.

Below, we provide a more detailed review of the methods used for water quality detection and monitoring. As our manual investigation (see Table 3) and machine summary (word cloud, see Figure 1) indicate, the dominant methods used in water quality detection with AI are LSTMs (Section 3.3.1) and ANNs, MLPs, DNNs, and other DL methods (Section 3.3.5). Beyond LSTM and ANN-based methods, there are other methods including LSTM hybrids (Section 3.3.2), CNN-based methods (Section 3.3.3), and "shallow" ML-based methods (Section 3.3.4).

3.3.1. LSTM-Based Water Quality Detection and Monitoring

Algal blooms cause serious harm to human and animal health and can damage both environments and economies. Various factors lead to algal blooms and gathering the data necessary to predict them is time- and cost-intensive. ML models can provide advanced warning for these events by taking into account time series data of basic water quality parameters. A linear regression model was compared with an MLP, an RNN, and an LSTM to predict harmful algae blooms in dammed pools from several rivers [57]. While the LSTM model was the most accurate overall, for several of the dammed pools that the authors tested, a least-squares regression model outperformed the LSTM. This casts doubt as to how the LSTM model generalizes and if it is worth the added complexity.

Water pollution is becoming an increasing problem because of rapid rates of development and urbanization. Large amounts of water quality parameters can be taken via IoT sensors, and DL techniques are well suited to finding patterns in the large quantity of data. An LSTM was used to predict future values of different water quality parameters [60]. Most importantly, the authors only used single-dimensional inputs and outputs (i.e., a 1D time series of dissolved oxygen as an input to predict dissolved oxygen at some time in the future). While the results were good, the authors noted that the architecture would benefit from training on multiple time series at the same time. The authors reported that long-term predictions on the order of 6 months into the future did not work well. Beyond monitoring water for different levels of pollutants, it is also important to find the sources of pollutants when they are identified. Cross-correlation was used to map pollutants to different water quality parameters [58]. They then used an LSTM to match pollutants to nearby polluting industries using the highly correlated water quality parameters.

Similar to LSTMs, RNNs have been demonstrated to be accurate for times series prediction but are also often criticized for being difficult to interpret. Meanwhile, process-based ecological models, although deterministic, fail to capture patterns at longer time scales. A process-based model was integrated with an RNN to better align predictions of phosphorus levels in lakes to eliminate outlier predictions. Constraining NN output with physics-based models better aligns their predictions with ecological principles [68].

Rapid development has led to decreased water quality. In [70], water quality parameters can be used to both classify the current water quality index and predict future water quality index states. However, the authors separately compared DL models for water quality prediction and ML models for water quality classification, making the methods not directly comparable. A nonlinear autoregressive neural network (NARNET), a type of ANN, performed better than an LSTM at predicting the water quality index, while an SVM performed better than other traditional ML models for classification.

3.3.2. LSTM Hybrids Water Quality Detection and Monitoring

To further improve model performance, a few recent studies have integrated other models with LSTMs. Water scarcity and drought are increasingly significant environmental challenges. Increased development is leading to worsening water pollution. Predicting the

water quality from time series data is essential, but traditional ML models fail to capture long-term temporal patterns. This causes them to make false predictions in water quality monitoring applications. An RNN–Dempster–Shafer (RNN–DS) evidence theory hybrid model was used to make sense of multiple input time series of different time scales [63]. While evidence theory did make the predictions more stable, longer-term predictions did not work very well, even with the improvements to the model. The authors pointed out one possible reason might have been not taking spatial correlations between water quality parameters into account.

Economic development and urban growth have posed water quality issues. Wavelet domain threshold denoising (WDTD) and wavelet mean fusion (WMF) were used to analyze the output of LSTM predictions for multiple water quality parameters [65]. While multiple wavelet basis functions were used to analyze predictions, the LSTM was not compared to any other models in this analysis. The authors noted that not having enough observations was a limitation while training their LSTM model.

Mangrove wetlands provide habitats for many different types of animal species in addition to preventing coastal erosion. More recent research has focused on monitoring the water quality in these environments to assess the health of coastal ecosystems. Using water quality and meteorological time series data, three different submodels were used for each water quality parameter at different time intervals and fused their output predictions [66]. The authors tested this setup with a DNN, a gated recurrent unit (GRU), and an LSTM model. While the LSTM performed the best, the authors reported that the model is not very reusable or user-friendly.

Collecting and analyzing water samples is expensive, time-consuming, and labor-intensive. Thus, many researchers choose to use sensors to remotely monitor water quality parameters, but the number of parameters they can record are often limited. Ref. [69] used a submerged multiprobe sensor to monitor several important water quality parameters over the course of 1 year. They found that a CNN–LSTM model performs better than standalone DL models and traditional ML methods for predicting water quality parameter values; however, the authors did not use a validation set during NN training and the hybrid model was able to quickly learn the training and testing set data distributions.

3.3.3. CNN-Based Water Quality Detection and Monitoring

CNNs are the dominant architecture for water body detection (Sections 3.3.1 and 3.3.2) but are not used as widely for water quality. Here, we review two very interesting but effective CNN-based methods. In situ water quality measurements work really well but are very expensive. In addition, things such as total nitrogen and phosphorus, biological oxygen demand, and dissolved oxygen are hard to measure from satellites because they have weak optical properties. A CNN was used in [59] and showed that TL beats out traditional ML models when classifying water quality from RS imagery. However, their dataset was very small, and their focus was narrow (specifically, only two lakes in China, no rivers or coastal waters covered). Water bodies are often polluted, or their quality is affected from far away and thus it is difficult to identify and report on water quality. Methods for estimating water quality at scale are essential. Turbidity can be a proxy for total suspended solids (TSS) and suspended sediment concentration (SSC), so [72] used image detection and then applied edge detectors to UAV images of water. They employed CNNs to detect changes in water color and utilized this to approximate quality. They showed that image-based turbidity detection is as accurate as actual turbidity meters, but more importantly represents a very promising method for monitoring water quality at greater spatial scales.

3.3.4. "Shallow" ML-Based Water Quality Detection and Monitoring

Remote water bodies are hard to monitor for water quality. A simple NN architecture was designed to estimate several water quality parameters (i.e., chlorophyll-a, turbidity, phosphorus) both before and after an ecosystem restoration project during both the dry

and wet seasons [55]. Importantly, their predictions, using seven different input bands for training the NN, were very close to the actual values.

Finding what data to input into an ML model for water quality monitoring is neither easy nor straightforward. Different indices are sensitive to different areas and varying weather and lighting conditions. To address this problem, [71] first correlated water quality parameters to different RS bands. These correlations were then used to test four ML models and their ability to predict a water quality index. Their R^2 statistics were not high, though.

3.3.5. ANN, MLP, DNN, and Other DL-Based Methods for Water Quality Detection and Monitoring

Climate change is making droughts and water shortages increasingly worse in arid regions. It is thus important to develop methods and systems for intelligent and efficient monitoring of the water resources in those regions. A water quality index for arid regions was proposed in [56] and attempted to find which bands and spectral indices are related to that water quality index. In situ water quality sampling is labor- and cost-intensive and often suffers from low temporal resolution. As bodies of water around the world are changing rapidly due to global warming, it is more important than ever to model their spatial variation through time. A point-centered regression CNN (PSRCNN) was used in [73] to analyze lake reflectance data to model water transparency. The authors concluded that their model outperformed different band ratios and traditional ML models (KNN, RF, SVM), although at the cost of generalization. The PSRCNN did not make stable predictions due to too little data.

There is currently not enough paired RS imagery and in situ water measurement to meaningfully create robust water quality monitoring applications. The generation of a synthetic dataset of atmospheric reflectances and its suitability for water quality monitoring were investigated in [76]. The synthetic dataset is physics-based and attempts to capture the natural variability in inland water reflectances and chlorophyll-a concentrations. An ANN outperforms several traditional ML models (KNN, RF, XGBoost) in predicting actual water quality parameter values when trained on the synthetic dataset, although only the ANN is validated against unseen data. Still, synthetic data generation is a promising research direction for water body and water quality detection. Without RS imagery, many water quality monitoring programs will suffer from lack of spatial coverage due to labor, time, and cost constraints. Yet while RS is a useful tool for monitoring water quality parameters, it has not been meaningfully integrated into operational water quality monitoring programs. Existing water quality time series data were used in [75] and assessed the effectiveness of multiple RS data platforms and ML models in estimating various water quality parameters. The authors showed that some sensors are poorly correlated with water quality parameters, while others are more suitable for water quality monitoring tasks. They concluded that more research needs to be carried out for assessing the suitability of paired RS imagery and in situ field data.

Current water quality monitoring systems are labor-, time-, and cost-intensive to operate. IoT sensors can monitor water quality parameters in near real time, allowing for much more data to be recorded with much higher temporal resolution. A wireless sensor network made up in part of IoT sensors was used in [61], and used an MLP to classify water quality as either good or bad. The authors utilized the MLP predictions to notify water quality managers via SMS if the water quality drops below a certain threshold value. However, because of the cost to deploy and run the network, the authors were not able to include additional water quality parameters from more types of bodies of water other than rivers. Water quality monitoring data collection is expensive and time consuming, and there are usually tradeoffs between spatial and temporal resolution when implementing data collection programs. In addition, several key water quality parameters (pH, turbidity, temperature) can be estimated directly from optical and infrared RS imagery. Randrianianina et al. [64] used RS imagery and DNNs to model water quality parameters directly, after which they extend their analysis to map the distributions of water quality

parameters to an entire lake, but they only focused on one lake and did not test their methods on other bodies of water.

As bodies of water are exposed to increased nutrient loads, harmful algal blooms can occur, leading to eutrophication. This process can create dead zones that would kill wildlife and lead to negative economic impacts. Thus, it is important to monitor chlorophyll-a levels in water bodies and predict algal blooms before they happen. Zhao et al. [74] attempted to address this need by comparing DL models to traditional ML and curve-fitting methods to predict chlorophyll-a levels using time series measurements paired with RS imagery. The authors did not have much data as they limited the data collection process to one lake. Thus, the DL models did not perform well. Additionally, the ML models used in this paper needed more data and computing than simpler models in order to perform well.

It is often difficult to monitor inland water bodies for quality because of low signal-to-noise ratios and limitations in resolution. A proximal hyperspectral imager was used in [77] with high spectral and temporal time series data for continuous water quality observations. The authors found that index-based methods of water quality detection were difficult to calibrate as thresholding values are subjective, while ML and DL models performed much better. However, the authors show that their models do not generalize well to other water bodies with different water quality parameter distributions.

Anthropogenic activities have currently threatened largely coastal ecosystems. Coastal ecosystems are complex bodies of water but monitoring them is very important. The performance of an ANN was compared to traditional ML models in [62] for predicting various water quality parameters. In some cases, traditional ML methods outperform the ANN. More importantly, the authors conducted an analysis of relative variable importance to show which sets of input data helped the ML models to learn the most. While the relative variable importance analysis is critically important, the authors only test their method in cloud-free RS imagery, limiting its utility. Additionally, while biophysical and chemical water quality parameters were analyzed, little work was carried out with bio-optical data due to issues with data availability.

While recent advances in RS capabilities for water quality detection are substantial in the literature, few papers have collected and synthesized the resources available to researchers. In a paper reviewing recent trends in RS imagery, cloud computing, and ML methods, [67] used time series data from hundreds of water quality parameters and water samples and combined them with proximal imagery, hyperspectral imagery, and two sets of data from different satellite data platforms. They showed that DNNs outperform many other traditional processing and ML techniques for assessing water quality. The authors conclude that anomaly detection using multisensor data is the most promising method for algal bloom detection. As is sometimes the case in the water body detection and water quality monitoring literature, the authors did not have a third holdout set (necessary for DL projects so that the data is not memorized).

4. Challenges and Opportunities

In this section, we first provide a brief summary and discussion of the key themes and overall insights (Section 4.1) derived from reviewing the range of research discussed above. In Section 4.2, we provide and discuss some of the major challenges we identified through our systematic survey. Specifically, those challenges shared in both domains are detailed in Section 4.2.1, those specific only to water body extraction in Section 4.2.2, and those specific to water quality monitoring in Section 4.2.3. Finally, we discuss possible research directions and related opportunities for water body detection and water quality monitoring using RS and AI in Section 4.3.

4.1. Summary and Discussion

After introducing the essential terms in AI and RS (Appendix B) and commonly used evaluation metrics in ML and DL for classification, regression, and segmentation tasks

(Appendix C), we reviewed recent and influential research for water body detection and water quality monitoring using RS and AI (Section 3).

While the research investigated in Section 3 has demonstrated the power of using RS and AI to detect water bodies and monitor water quality, very few studies thus far performed integrative research of water body and water quality using the power of RS and AI. In addition, most existing RS and AI-based work on water bodies and water quality repeat the same (or very similar) methods in a different research location or on a different (usually small) dataset. However, real intelligent water resource management applications will require serious development that goes beyond this type of research. Before operational applications can be deployed, AI models (especially DL models) need to be trained on large and representative benchmark datasets with a focus on making models generalizable and interpretable.

We noticed that most work does not include hardware specifications (e.g., what CPU/GPU the authors used to run their models) and/or processing time. To make models comparable and for the sake of replicability and reproducibility, it is essential to report such information. This is even true for index-based methods and more traditional ML models so that researchers can fully evaluate the trade-offs between runtime, accuracy, and ease of implementation. We hope our review will provide a useful guide to make future research more replicable and reproducible. From our interactive web app (the web app tool URL and its brief demo video link are provided in Appendix A), we also noticed that while most papers have an open access PDF/HTML version of their manuscripts, a sizable portion of manuscripts (16 out of 56 of reviewed articles) do not. We suggest authors provide an open access version (e.g., posting the proofreading version after acceptance to ResearchGate/arXiv) in order to increase the visibility of their research and thus to accelerate the advancement of scientific knowledge.

4.2. Identified Major Challenges

Below, we provide the most commonly posed challenges for water body and water quality research in the literature we reviewed. Those challenges shared in both domains are outlined in Section 4.2.1 and those specific to each domain are detailed in Sections 4.2.2 and 4.2.3, respectively. Here are some specific issues to water body detection and water quality monitoring.

4.2.1. Shared Common Challenges in Both Domains

A summary of the shared common challenges and identified problems in water body extraction and water quality monitoring using RS and AI are provided below.

- Methods for water body detection and water quality monitoring need to be able to work quickly and reliably on large spatial and temporal scales, and yet high-resolution RS imagery is very complex. Index methods rely on subjective threshold values that can change over time and space depending on weather conditions. Shallow ML models are more accurate, but do not work at scale. DL models are complex, require very large datasets to train on, and are very computationally expensive; also, the hyperparameter tuning process is very tedious and difficult.
- It is difficult to know exactly what data to feed to ML and DL models, and it is difficult to know what to make of the output predictions. This often requires integrative expertise and/or interdisciplinary collaboration of RS, hydrology, biology, and CV/ML expertise.
- NNs generally perform the best in water quality and water body detection tasks but are often the least stable models (i.e., they do not generalize well). This is not surprising, as the datasets used in RS problem settings are often not large enough to allow NN models (too many parameters compared with shallow ML models) to overcome overfitting (see Appendix B). Table 4 summarizes the relatively few existing datasets we identified through our systematic review.

- Both domains over-rely on optical RS imagery, and thus clouds and shadows are a persistent problem and heavily skew the results towards working only in cloud-free conditions.

Table 4. Existing datasets for waterbody extraction and water quality monitoring.

Datasets	Source	Category	Link to the Dataset	Notes
DeepWaterMap v2	[40]	Water body	https://github.com/isikdogan/deepwatermap, accessed on 15 December 2021	>1 TB of Landsat-7 imagery paired with Global Inland Water dataset labels.
2020 GaoFen Challenge	/	Water body	https://github.com/AICyberTeam/2020Gaofen, accessed on 15 December 2021	Dataset containing both 2500 optical and 1200 SAR satellite images with pixel level labels for water body segmentation.
GID-15	[79]	Water body	https://captain-whu.github.io/GID15/, accessed on 15 December 2021	150 pixel-level annotated GaoFen-2 images for semantic segmentation tasks.
LandCover.ai	[80]	Water body	https://landcover.ai/, accessed on 15 December 2021	A dataset from 2015–2018 of 10,674 annotated tiles of RGB imagery with labeled water bodies.
SEN12MS	[81]	Water body	https://mediatum.ub.tum.de/1474000, accessed on 15 December 2021	A curated dataset of 180, 662 georeferenced multispectral Sentinel-1 and -2 imagery with MODIS land cover labels.
AquaSat	[82]	Water quality	https://github.com/GlobalHydrologyLab/AquaSat, accessed on 15 December 2021	600,000 data matchups between satellite imagery and water quality measurements from 1984–2019.
Forel–Ule Index	[83]	Water quality	https://doi.org/10.6084/m9.figshare.13014299, accessed on 15 December 2021	151 data matchups between satellite imagery and water quality measurements from 2000–2018.

4.2.2. Additional Challenges in Water Body Extraction

The specific challenges and problems identified for water body extraction are summarized below.

- The majority of reviewed research focused on inland bodies of water, where only a few papers discussed applications for coastal waters (not including oceans). Moreover, many papers focus solely on only one type of water body, for example, only on lakes or rivers in a specific area. As a result, water bodies from different landscapes (e.g., inland, coastal tidal flats, urban, wetlands) are difficult to recognize with one unified method (i.e., methods do not generalize). The same applies to water bodies of different colors, especially when distinguishing them from rock, ice, snow, clouds, and shadows.
- There are very few benchmark datasets. In contrast, there are huge volumes of unlabeled data not being fully leveraged.
- CNNs blur output boundaries during the segmentation process.

4.2.3. Additional Challenges in Water Quality Monitoring

The specific challenges and problems identified for water quality monitoring are summarized below.

- Collecting in situ water quality data is very time- and labor-intensive and financially expensive; also, it often does not have adequate temporal or spatial resolution.
- RS imagery and existing corresponding field samples are often not stored together. Allowing water quality researchers to easily retrieve and locate two or more sources

of data at the same location is critical, as computational methods require such data to verify their model performance in order to generalize to new water bodies.
- Remote water bodies are difficult to monitor.
- Urbanization, pollution, and drought are having serious effects on the economy, wildlife, and human health as they deteriorate water quality.
- Ecosystems are complex and their nutrient and pollution budgets are not well understood.
- Some studies do not use a training, validation, and testing set for DL projects (all three are necessary) or do not use nearly enough data to achieve good results with DL models.

4.3. Research Directions and Opportunities

Here, we provide five research directions, each along with its promising opportunities, from our investigation and based on the posed challenges discussed in Section 4.2 above.

4.3.1. Urgent Need of Large and Comprehensive Benchmark Datasets

Large representative, balanced, and open-access benchmark datasets are critical for any domain to let AI meaningfully shine [84–86]. In computer science, especially for its branches CV and DL, there are very comprehensive, large, and open-source databases (e.g., ImageNet [87] for image classification tasks, and Microsoft COCO [88] for object detection and segmentation tasks). The availability of big and open-source image repositories has dramatically boosted recent advances in novel and robust algorithms in DL and CV, as computer science researchers do not need to worry about collecting datasets. Instead, they can focus on developing new algorithms and/or methods.

In our systematic review, we identified an urgent need for more curated, labeled datasets for intelligent water body extraction and water quality monitoring. We found some of the few available open-source datasets with water body boundary labels through our literature review, but also sought out additional datasets. We identified datasets that were not used in our literature review but contain water body labels, or datasets that were used for water body detection or water quality monitoring that did not use ML/DL/CV but would be useful for benchmarking tasks. Our search results are summarized in Table 4 above. Below, we list a few opportunities in this direction.

(1) *More public data and code: currently, most authors do not share their code and/or datasets.* See the two quoted pieces below from [25]: (a) "Lack of deep learning-ready datasets within the water field [. . .] The main problem caused by this absence of many datasets is that the research community does not build upon previous work in terms of constructing better neural network architectures and moving the state of art to the next iteration [. . .]"; (b) "[. . .] many papers are published that achieve the same task with almost identical methods but different data.". Part of this issue is a replication crisis in the water body detection and water quality monitoring literature, but it stems more broadly from the lack of public codebases and datasets.

(2) Some promising ways to generate large datasets of good quality

- AI/ML/DL models need large datasets with good quality to guarantee meaningful (unbiased and generalize well) good to great performance, thus work on obtaining large but better subsets of data. Quality > quantity is critical and in urgent demand. See one piece of such evidence reported in [44], "[. . .] site-specific models improved as more training data was sampled from the area to be mapped, with the best models created from the maximum training datasets studied: [. . .] However, performance did not improve consistently for sites at the intermediate training data thresholds. This outcome exemplifies that model improvement is an issue of not only increasing the quantity of training data, but also the quality".
- Generating synthetic data as in [76] (detailed in the second paragraph in Section 3.3.5).
- Downloading RS images from Google Earth Engine (GEE) and annotating accordingly, or, even better, developing user-friendly interactive interfaces with GEE as a backend to directly allow researchers (or even citizen science volunteers) to contribute to the

annotation of RS imagery available on GEE. To our knowledge, no RS datasets for water body detection and water quality monitoring are downloaded from GEE and then annotated, let alone interfaces for directly annotating RS imagery on GEE.

- Obtaining RS imagery from Google Earth (GE) manually or with the help of code scripts, then annotating accordingly (see [34,42,49] for examples). For instance, the following two datasets generated and used in [34,49] are both from GE, but are not shared publicly.
    - "The first dataset was collected from the Google Earth service using the BIGEMAP software (http://www.bigemap.com, accessed on 15 December 2021). We named it as the GE-Water dataset. The GE-Water dataset contains 9000 images covering water bodies of different types, varying shapes and sizes, and diverse surface and environmental conditions all around the world. These images were mainly captured by the QuickBird and Land remote-sensing satellite (Landsat) 7 systems." [49].
    - "We constructed a new water-body data set of visible spectrum Google Earth images, which consists of RGB pan-sharpened images of a 0.5 m resolution, no infrared bands, or digital elevation models are provided. All images are taken from Suzhou and Wuhan, China, with rural areas as primary. The positive annotations include lakes, reservoirs, rivers, ponds, paddies, and ditches, while all other pixels are treated as negative. These images were then divided into patches with no overlap, which provided us with 9000 images [. . .]" [34].

4.3.2. Generalization

It is important to be able to obtain a good accuracy score when training an ML/DL model, but perhaps more important is that model's ability to generalize to unseen data. The ultimate goal of ML/DL is to develop predictive models through finding statistical patterns in a training set which then generalize well to new, previously unseen data outside the training set [89]. Ideally, this is achieved by training on large and representative datasets that capture nearly all variations in the data actual distribution of values [86,89]. A model's ability to generalize is critical to the success of a model. An ML/DL model with good generalization capability will have the best trade-off between underfitting and overfitting so that a trained model obtains the best performance (See "Generalization, overfitting, underfitting and regularization" entry in Appendix B for details). Below, we outline a few ways to make AI systems more generalizable for water body detection and water quality monitoring tasks.

(1) Create robust AI methods for tiny water body detection. Depending on resolution, tiny water bodies such as ponds or small lakes in desert cities are difficult to identify yet may play a more critical role than we think.

(2) Develop NN architectures and comprehensive datasets (see Section 4.3.1) that are able to recognize water bodies not just from

- One type of body of water (e.g., ponds, lakes, rivers);
- One color (e.g., different levels of sediment, aquatic vegetation and algae, nutrients, pollutants);
- One size: Water bodies present in RS imagery come with different sizes (large and small water bodies) and various shapes. Many studies reported that it is not an easy task to correctly classify small water bodies and/or water bodies with different shapes.
- One environment setting (e.g., desert, urban, inland, coastal).

(3) Utilize data from multiple sources to train ML/DL models. From our comprehensive investigation, most of the current AI methods are only able to deal with water quality and/or water body detection data from one specific type of RS imagery. This should be improved and indicates a promising new research direction. Specifically, it will be important to focus on using data from multiple data platforms or resolutions, from varying weather conditions, and regions which have different ecosystem and terrain types. We humans can recognize water bodies in different RS imagery with different weather conditions. We

expect that machines should be able to mimic humans to perform this task well if we have robust AI algorithms and comprehensive datasets. See some example research below:

- Extraction of water bodies at multiple resolutions and scales using CNNs [49,53];
- Evaluation of CNN performance on multisensor data from multiple RS platforms [52];
- Integration of data from multiple sources (e.g., SAR, UAV, smaller sensors, water quality time series);
- Data fusion of Landsat-8 and Sentinel-2 RS imagery for water quality estimation [67]. "Virtual constellation" learning introduced in [67] could be a future direction for both water body detection and water quality estimation. A virtual constellation is constructed by using multiple RS platforms to "shorten" the revisit time and improve the spatial coverage of individual satellites. This entails fusing data sources from separate RS platforms with potentially different resolutions.

(4) Propose new frameworks for improving generalizability. Generalization is one of the fundamental unsolved problems in DL. The goal of a generalization theory in supervised learning is to understand when and why trained ML/DL models have small test errors [90]. The recently proposed deep bootstrap framework [90] provides a new lens for understanding generalization in DL. This new framework has the potential to advance our understanding of water domain research empowered by RS and AI by highlighting important design choices when processing RS imagery with DL.

4.3.3. Addressing Interpretability

DL has achieved significant advances with great performance in many tasks in a variety of domains, including some water domain tasks (detailed in Section 3). In the literature we reviewed for this paper, DL models have produced results comparable to, and in some scenarios even superior to, human experts. Improving predictive accuracy is important; however, improving the interpretability of ML/DL models is more important, especially through visualization techniques of ML/DL model output for later analysis by humans [18]. Interpretability is one of the primary weaknesses of DL techniques and raises wide concerns and attention in DL [91]. Due to the overparameterized and black-box nature of DL models, it is often difficult to understand the prediction results of DL models [92,93]. Understanding and explaining their black-box behaviors remains challenging due to their hierarchical, nonlinear nature. The lack of interpretability raises major concerns across several domains; for example, in high-stakes prediction applications, such as autonomous driving, healthcare, and financial services [94], the trust of DL models is critical. While many interpretation tools (e.g., image perturbation and occlusion [95], visualizing NN activation weights and class activation mapping [96,97] or attention mechanisms [98,99], feature inversion [100], local interpretable model-agnostic explanations or "LIME" [101]) have been proposed to interpret how DL models make decisions, either from a scientific perspective or a social angle, explaining the behaviors of DL models is still in progress [92]. For water domains, we list some specific potential opportunities in terms of interpretability we identified below.

- More ablation studies are needed (see Appendix B for an introduction) to investigate the role of each DL component in terms of model performance contribution and ultimately which component(s) control the model performance.
- Exploring the output of hidden layers to obtain some information to help investigate whether the model works as expected.
- Hybrid models for analyzing NN output and improving an NN's decision-making process through post-processing, for example, CNN–LR hybrids [32], CNN–CRF hybrids [36,38], CNN–SVM hybrids [39], RNN–DS hybrids [63], and CNN-LSTM hybrids [69].
- More research needs to be carried out on analyzing the importance of input data to output predictions. See examples in [62,75], each detailed below.
    - The authors in [62] systematically analyzed relative variable importance to show which sets of input data contributed to the ML models' performance. See

the quoted text below: "Relative variable importance was also conducted to investigate the consistency between in situ reflectance data and satellite data, and results show that both datasets are similar. The red band (wavelength $\approx$ 0.665 μm) and the product of red and green band (wavelength $\approx$ 0.560 μm) were influential inputs in both reflectance data sets for estimating SS and turbidity, and the ratio between red and blue band (wavelength $\approx$ 0.490 μm) as well as the ratio between infrared (wavelength $\approx$ 0.865 μm) and blue band and green band proved to be more useful for the estimation of Chl-a concentration, due to their sensitivity to high turbidity in the coastal waters".

- The authors in [75] utilized existing water quality time series data and assessed the effectiveness of multiple RS data platforms and ML models in estimating various water quality parameters. One of their interesting findings is that some sensors are poorly correlated with water quality parameters, while others are more suitable for water quality monitoring tasks. They suggested that more research needs to be carried out for assessing the suitability of paired RS imagery and in situ field data. See the quoted text below: "[. . .] assess the efficacy of available sensors to complement the often limited field measurements from such programs and build models that support monitoring tasks [. . .] We observed that OLCI Level-2 Products are poorly correlated with the RNMCA data and it is not feasible to rely only on them to support monitoring operations. However, OLCI atmospherically corrected data is useful to develop accurate models using an ELM, particularly for Turbidity (R^2 = 0.7)." (RNMCA is the acronym for the Mexican national water quality monitoring system).

- Water quality monitoring will benefit from more research exploring how well a certain ML/DL model contributes to which water quality parameter(s). See an example in [67], where the authors investigated how well DNNs could predict certain water quality parameters.
- Physics-constrained or process-based ML/DL predictions as demonstrated in [68,69].
- The need for automatic and visually-based model evaluation metrics that are better than current visual assessment as an evaluation metric. For example, automatic assessment of how DL methods are performing in large and complex RS imagery (e.g., specifically, Bayesian DL, and Gaussian DL/ML for uncertainty measurement and visualization).

4.3.4. Ease of Use

As emphasized in [13,14], one of the major current challenges for water resource management is the integration of water quality data and indices from multiple sources into usable and meaningful insights for actionable management decisions. Geovisualization, also known as geographic visualization, uses the visual representations of geospatial data and the use of cartographic techniques to facilitate thinking, understanding, knowledge construction, and decision support about human and physical environments at geographic scales of measurement [102,103]. Geovisualization is widely utilized in different domains (e.g., public health [104], crisis management [105,106], environmental analysis [107–109], and climate change strategies [110]) for the exploration and analysis of spatiotemporal data. To the best of our knowledge, very little research has leveraged geovisualization in this way for water resources management. The only piece of work similar to this we noticed is in [111], where a web interface powered by GEE allows their expert system, combined with visual analytics, to be run on any Landsat 5, 7, or 8 imagery to draw boundaries for water bodies. Geovisualization through interactive web applications provides a promising solution to the posed challenge of integrating water quality data and indices from multiple sources [112–115]. We provide a few suggested research opportunities in this direction below.

- Simply applying (or with minor modifications) existing AI/ML/CV/DL algorithms/methods to RS big data imagery-based problems is still very far away from producing

real-world applications that meet water management professionals' and policymakers' needs. As echoed in [13], "[. . .] realizing the full application potential of emerging technologies requires solutions for merging various measurement techniques and platforms into useful information for actionable management decisions, requiring effective communication between data providers and water resource managers" [116]. Much more multidisciplinary and integrative collaboration in terms of depth and breadth are in high demand. Those scholars and practitioners who have an interdisciplinary background will play a major role in this in-depth and in-breadth integration. For example, researchers who have expertise in RS but also know how to utilize AI, through collaboration with domain expertise such as water resources management officers, will significantly advance this research direction. Intuitive interactive web apps that are powered by both geovisualization and AI/ML/DL/CV will definitely make interdisciplinary collaboration much more seamless and thus easier.

    - Interactive web portal empowered by geovisualization for integration of various water quality data sources. As noted in [117], it is natural and intuitive in many studies to use "space" as the organizing paradigm.
    - More smart and responsive water management systems through the development of interactive web apps/libraries that integrate ML/DL backends and intuitive, user-friendly front ends are needed. Such systems would allow collaboration between technical experts and domain experts, including stakeholders, and even community volunteers, from anywhere at any time.
    - This requires very close collaboration and thus very integrative research from researchers in many domains (e.g., computer science, cognitive science, informatics, RS, and water-related sub-domains). We reinforce that geovisualization will be the ideal tool to make the collaboration smooth, productive, and insightful.
    - There is one recent work [118] that takes a small step in this direction, but much more work and efforts are in demand.
- Resource hubs for standardized AI/ML/DL/CV models and easy-to-follow and understandable tutorials for how to use them are needed.
- More data "matchups" as demonstrated in [82,83]. When more in situ measurements come in, they should be matched up and stored with satellite data for easy calibration studies.

4.3.5. Shifting Focus

From our investigation, it is clear that with enough annotated data and allocated computing, DL models are more accurate than traditional ML models, which are in turn more accurate than index-based methods for water body detection and water quality monitoring tasks. Increasing the accuracy of models by fractions of a percent should be given much less focus and attention moving forward. Water body detection methods are unlikely to improve upon the high rates of accuracy already reported in the literature without very high-resolution, very large, labeled datasets or the use of UAVs to detect small water bodies. Instead, we suggest that future research should focus more on reducing model parameters and making model training less computationally expensive in terms of time (e.g., designing neural networks to use constant memory at inference time [40], or by using TL [37,59]). Below, we outline some additional potential research directions we identified through our systematic review.

- As noted in Section 4.3.1, the lack of large benchmark datasets is a bottleneck in water body detection and water quality monitoring research utilizing RS imagery and AI. The dominant methods in both water domains are supervised learning, which often requires very large, labeled datasets to train on, thus, there is a clear, urgent need for semi-supervised and unsupervised learning methods [15].
    - Unsupervised learning methods are able to learn from big sets of *unlabeled* data, as demonstrated in [29,46].

- Semi-supervised learning methods are able to learn from limited good-quality labeled samples. DL models do not require feature engineering, and they are also much better at discovering intricate patterns hidden in big data. However, pure supervised DL is impractical in some situations, such as those for which the labeling tasks require domain knowledge from experts. Very few domain experts have the time and are willing to label very large sets of RS images [84]. An active learning-enabled DL approach that uses a visualization interface and methods to iteratively collect modest amounts of input from domain experts and uses that input to refine the DL classifiers [84] provides a promising direction to produce well-performing DL models with limited good-quality datasets.

- From our systematic review, we can easily see that current work on water body extraction and water quality monitoring using AI and RS are, in general, carried out separately. We call for a closer integration of water body detection and water quality monitoring research and more attention focusing on handling massive datasets that may include information in a variety of formats, of varying quality, and from diverse sources. This integration is critical as it will provide the essential foundation for developing real, intelligent water monitoring systems using RS and AI capable of producing insights used for actionable decision making.
- GEE + AI: as noted in [18], GEE is a good solution to address computational costs and overcome technical challenges of processing RS big data. However, online DL functionality is still not supported on GEE. To the best of our knowledge, the only piece of research integration of the Google AI platform with GEE is performed in [119]; however, as the authors reported, "data migration and computational demands are among the main present constraints in deploying these technologies in an operational setting". Thus, the ideal solution is to develop DL models directly on the GEE platform.
- Most current ML/DL-based RS research focuses on borrowing or slightly improving ML/DL/CV models from computer science [79,120]. Compared with natural scene images, RS data are multiresolution, multitemporal, multispectral, multiview, and multitarget [15]. Slight modifications of ML/DL/CV models simply cannot cope with the special challenges posed in RS big data. New ML/DL models specialized for RS big data are thus urgently needed [15,18]. We hope our review will draw the attention of researchers who have a multidisciplinary background to this issue. Looking deep into the mechanisms of RS and land surface processes, studying the characteristics of RS imagery would guide the design of specialized ML/DL models for RS big data and thus further improve RS applications using AI in breadth and depth [15].

5. Conclusions

Building intelligent and synoptic water monitoring systems requires automation of water body extent detection using RS imagery, from which volume can be computed, and also automation of their corresponding water quality, eventually linking the two to allow synoptic water quality monitoring. Yet, to date, water body detection and water quality monitoring research has been historically separate. Our systematic investigation indicates the following trends: deep learning is much more commonly used in water body detection, the dominant data source of which is RS imagery, whereas water quality literature often involves other types of data sources (e.g., in situ sensors, smaller RS devices that are not satellites). The trends relate to the scale of projects in the two domains: water body extraction is usually undertaken across large spatial scales, whereas the water quality monitoring literature is still only focused on smaller, often individual, bodies of water. This points to one of the future research directions in the water quality literature that we touch on above in Section 4.3; that is, we need to scale up water quality estimation using RS imagery through matching it with ground-truth water quality measurements.

Overall, based on the systematic review above, we contend that RS integrated with AI/ML/DL/CV methods, along with geovisualization, have great potential to provide

smart and intelligent support for water resources monitoring and management. Thus, this integration has considerable potential to address major scientific and societal challenges, such as climate change and natural hazards risk management.

Author Contributions: All authors have contributed to this review paper. L.Y. initiated the review, contributed to writing and overall organization, identified selected research to include in the review, supervised the web app design and development, and coordinated input from other authors. J.D. took the lead on identifying relevant literature, contributed to writing and editing the text, and provided the data for the accompanying interactive web app. S.S. contributed to the web app design and development, word clouds visualization, and editing. Q.W., C.D.L. and M.M. have contributed to editing and M.M. also contributed to writing part of the introduction section along with identifying some relevant literature. All authors have revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This material is partly based upon work supported by the funding support from the College of Arts and Sciences at University of New Mexico. The authors are also grateful to the three reviewers for their useful suggestions.

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

Abbreviations

The following abbreviations (in alphabetical order) are used in this manuscript:

AE	Autoencoder
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BOA	Boundary Overall Accuracy
CA	Class Accuracy
CART	Classification and Regression Trees
CB	Cubist Regression
CE	Commission Error
CNN	Convolutional Neural Network
COCO	Common Objects in Context
CPU	Central Processing Unit
CRF	Conditional Random Field
CV	Computer Vision
DL	Deep Learning
DNN	Dense Neural Network
DS	Dempster–Shafer Evidence Theory
DT	Decision Tree
DEM	Digital Elevation Model
ECE	Edge Commission Error
ELM	Extreme Learning Machine
ELR	Extreme Learning Regression
ESA	European Space Agency
EOE	Edge Omission Error
EOA	Edge Overall Accuracy
FN	False Negative
FP	False Positive
FWIoU	Frequency Weighted Intersection over Union
GA	Global Accuracy

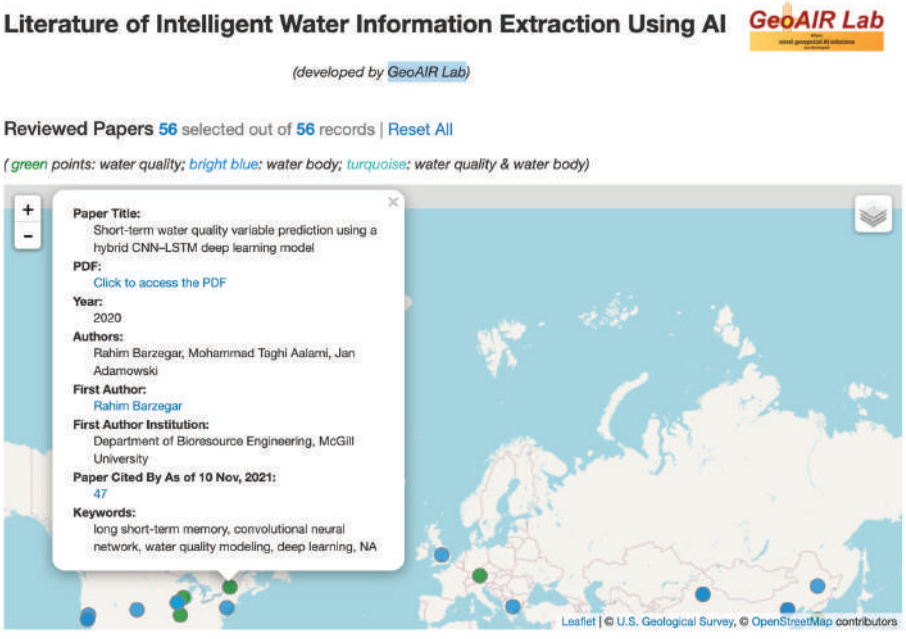
GAN	Generative Adversarial Network
GBM	Gradient Boosted Machine
GE	Google Earth
GEE	Google Earth Engine
GPR	Gaussian Process Regression
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
IoT	Internet of Things
IoU	Intersection over Union
Kappa	Kappa Coefficient
KNN	K-Nearest Neighbors Classifier
LORSAL	Logistic Regression via Variable Splitting and Augmented Lagrangian
LSTM	Long Short-Term Memory
MA	Mapping Accuracy
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
mIoU	Mean Intersection over Union
MK	Mann–Kendall
ML	Machine Learning
MLC	Maximum-Likelihood Classifier
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MNDWI	Modified Normalized Difference Water Index
MPC	Microsoft Planetary Computer
MRE	Mean Relative Error
MSE	Mean Squared Error
MSI	Morphological Shadow Index
NB	Naive Bayes Classifier
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-Infrared
NN	Neural Network
NSEC	Nash–Sutcliffe Efficiency Coefficient
OA	Overall Accuracy
OE	Omission Error
PA	Producer's Accuracy
PCC	Percent Classified Correctly
RBFNN	Radial Basis Function Neural Network
R-CNN	Region Based Convolutional Neural Network
RF	Random Forests
RMSE	Root Mean Squared Error
RMSLE	Root Mean Squared Log Error (referred to in Table 3 as RMSELE by the authors)
RNN	Recurrent Neural Network
RPART	Recursive Partitioning And Regression Trees
RPD	Relative Percent Difference
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SRN	Simple Recurrent Network (same abbreviation given for Elman Neural Network)
SOTA	State-of-the-Art
SVM	Support Vector Machine
SVR	Support Vector Regression
SWIR	Short Wave Infrared

TB	Tree Bagger
TL	Transfer Learning
TN	True Negative
TP	True Positive
VHR	Very High Resolution
UA	User's Accuracy
UAV	Unmanned Aerial Vehicle

Appendix A. The Accompanying Interactive Web App Tool for the Literature of Intelligent Water Information Extraction Using AI

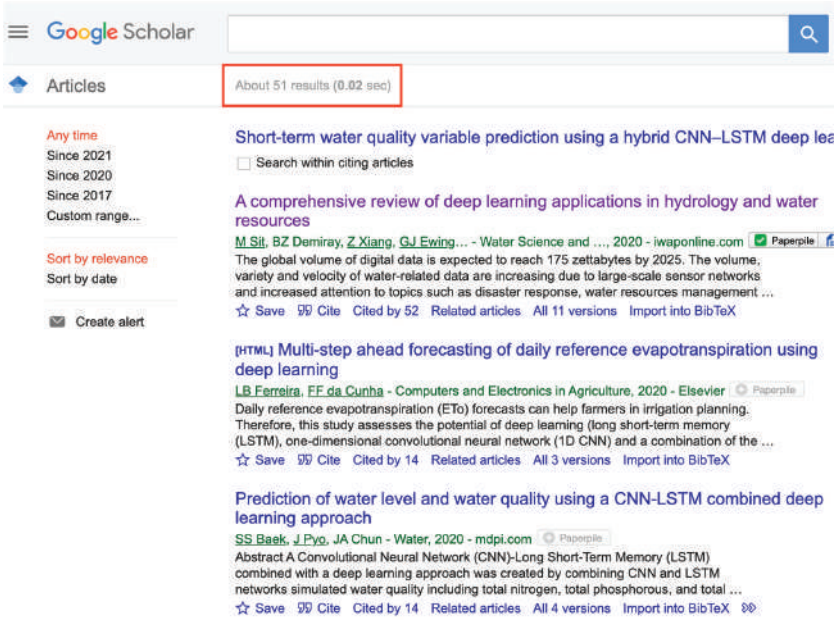
In Section 1.1, we provided a brief map and graphic summary of the papers covered in this review. To allow readers to obtain more useful and dynamic information and insights from the papers reviewed, we have developed an interactive web app. Through the web app, readers can keep track of the major researchers and access an up-to-date list of publications in the reviewed topics. Updated publications are accessible through (1) a researcher's public academic profile on Google Scholar or ResearchGate (see Figure A1a for an example), and (2) a continuously updated citations count of the papers that we reviewed in this paper (see Figure A1b for an example: the cited by as of 10 November 2021 is 47, which is when we first entered the data in our data file when we reviewed the paper, and then before this paper submission, when we clicked on the cited by URL, the page shows that the up-to-date citation number is 49). The web app can be accessed publicly, *free of charge* at

- Web app tool: https://geoair-lab.github.io/WaterFeatureAI-WebApp/index.html, accessed on 28 February 2022.
- Brief web app demo video (about 6 min duration): the video link is accessible at the web app page.



(**a**)

Figure A1. *Cont.*



(**b**)

Figure A1. Our highly interactive web app (accessible publicly at: https://geoair-lab.github.io/WaterFeatureAI-WebApp/index.html, accessed on 5 December 2021) provides the track of scholars and publications with just a few clicks. See an example on the pop-up. Our readers can access (1) a direct link to the PDF file of the paper (note that if there is no free, publicly available version of the paper, we link directly to the journal page of the paper so our readers can obtain the paper if their institution purchases the journal database), (2) the scholar profile (Google Scholar/ResearchGate URL) of the first author, and (3) "Cited by" Google Scholar page. (**a**) Water body and quality AI literature map pop-up. (**b**) "Cited by" Google Scholar page corresponding to the paper shown in (**a**).

Appendix B. Essential AI/ML/DL/CV Terms

In this appendix, we provide brief definitions to some essential terms (ordered alphabetically) in ML/DL/RS in our review. For readability, we group some related concepts together.

Ablation Studies: In AI, particularly in ML and DL, ablation is the removal of a component of an AI system. Ablation studies are crucial for AI, especially for DL research. An ablation study investigates the performance of an AI system by removing certain components to understand the contribution of the component to the overall system. The term is analogous to ablation in biology (removal of components of an organism). Note that ablation studies require that the systems exhibit *graceful degradation* (i.e., they continue to function even when certain components are missing or degraded). The motivation was that, while individual components are engineered, the contribution of an individual component to the overall system performance is not clear; removing components allows this analysis. Simpler is better: if we can obtain the same performance with two models, we prefer the simpler one.

Convolution, kernel (i.e., filter), and feature map [121–123]:

Convolutional layers are the major building blocks in CNNs. A convolution is the simple application of a filter (i.e., kernel) to an input that results in an activation. Repeated application of the same filter to an input results in a map of activations called a feature map, indicating the locations and strength of a detected feature in an input (e.g., an image).

Convolution: Convolution is one of the most important operations in signal and image processing. Convolution is a mathematical operation to merge two sets of information. Convolution provides a way of multiplying together two arrays of numbers, generally of different sizes, but of the same dimensionality, to produce a third array of numbers of the same dimensionality. This can be used in image processing to implement operators whose output pixel values are simple linear combinations of certain input pixel values.

A convolutional filter (i.e., kernel) is a weight matrix (vector for one-dimensional and *cube* for three-dimensional data) which operates through a sliding window on input data. The convolution is performed by determining the value of a central pixel through adding the weighted pixel values of all its neighbors together. Specifically, it is carried out by sliding the kernel over the input image, generally starting at the top left corner, so as to move the kernel through all the positions where the kernel fits entirely within the boundaries of the input image. Each kernel position corresponds to a single output pixel, the value of which is calculated by multiplying together the kernel value and the underlying image pixel value for each of the cells in the kernel, and then adding all these numbers together. The output is a new modified filtered image. Convolution is a general purpose filter effect for images. Depending on the kernel structure, the operation enhances some features of the input data (e.g., blurring, sharpening, and edge detection).

In the context of a CNN, a convolution is a linear operation that involves the multiplication of a set of weights with the input. Given that the technique was designed for two-dimensional input, the multiplication is performed between an array of input data and a two-dimensional array of weights (i.e., a filter or a kernel). Technically, note that in CNNs, although it is referred to as a "convolution" operation, it is actually a "cross-correlation". That is, in CNNs, the filter is not flipped as is required in typical image convolutions; except for this flip, both operations are identical.

Kernel (i.e., filter): A kernel is a small matrix used in image convolution, which slides over the input image from left to right and top to bottom. Differently sized kernels, which contain different patterns of numbers, produce different results through convolution operation. The size of a kernel is arbitrary, but 3×3 or 5×5 is often used. Think of a filter similar to a membrane that allows only the desired qualities of the input to pass through it.

Feature map: The feature maps of a CNN capture the application result of the filters to an input image (i.e., at each layer, the feature map is the output of that layer). Think of it as (higher level) representations of the input. The feature map(s) is/are the output image(s) of each convolutional layer(s). The resultant number of feature maps equals the number of filters.

Data augmentation (DA) [124]:

ML (especially DL) model performance often improves with an increase in the amount of data. The common case in most ML/DL applications, especially in image classification tasks, is that obtaining new training data is not easy. Thus, we need to make good use of the existing (relatively small) training set. DA is one technique to expand the training dataset from existing training data in order to improve the performance and generalizability of DL models. DA enriches (i.e., "augments") the training data by creating new examples through random transformation of existing ones. This way, we artificially boost the size of the training set, reducing overfitting. Thus, to some extent, DA can also be viewed as a regularization technique.

Image DA is perhaps the most well-known type of DA and involves creating transformed versions of images in the training dataset that belong to the same class as the original image. The ultimate goal is to expand the training dataset with new, plausible examples (i.e., variations of the training set images that are most likely to be seen by DL models). For example, a horizontal flip of a bike photo may make sense, because the photo could be taken from the left or right. A vertical flip of a bike image does not make sense and would probably not be appropriate as the model is very unlikely to see a picture of an upside down bike. Transformations for image DA include a range of operations from the field of image manipulation (e.g., rotation, shifting, resizing, flipping, zooming, exposure

adjustment, contrast change, and much more). This way, a lot of new samples can be generated from a single training example.

Note that image DA is typically only applied to the training dataset, and NOT to the validation or test dataset. This is different from data preparation such as image resizing and pixel scaling; those must be performed consistently across all datasets that interact with the model. The choice of the specific DA techniques used for a training dataset must be chosen carefully and within the context of the training dataset and knowledge of the problem domain. It can be useful to experiment with DA methods in isolation and in concert to see if they result in a measurable improvement to model performance, perhaps with a small prototype dataset, model, and training run.

DeepLabV3+ [125]: DeepLabV3 was firstly proposed to enable deep CNNs to segment features in images *at multiple scales.* ResNet-50 and ResNet-101, two variations on the popular residual network (ResNet) architecture, are the tested backbones for DeepLabV3. Through the use of residual blocks, atrous convolution, and a spatial pyramid pooling module, the authors showed that their new architecture achieved comparable performance to other SOTA models in image segmentation tasks without the need for further post-processing. The authors further improved DeepLabV3 and named the new version DeepLabV3+ [126], which combines atrous spatial pyramid pooling modules with an encoder–decoder module. This further improved the performance of DeepLabV3 while sharpening predicted feature boundaries. The DeepLabV3+ architecture is very popular in the water body extraction literature.

Generative adversarial network (GAN): GAN is a class of unsupervised DL frameworks in which two neural networks compete with each other. One network, the generator, tries to create synthetic or false images which fool the discriminator network. The discriminator, in turn, attempts to discern which images coming from the generator are actual vs. synthetic images [127]. GANs use a cooperative zero-sum game framework to learn. Among many variants of GAN, cycleGAN [128] is a technique for training unsupervised image translation models using the GAN architecture and unpaired collections of images from two different domains. CycleGAN has been demonstrated on a wide range of applications, including season translation, object transfiguration, style transfer, and generating photos from paintings.

Generalization, overfitting, underfitting and regularization (referenced [123,129,130]):

The prediction results of an ML/DL model sit somewhere between (a) low-bias, low-variance, (b) low-bias, high-variance, (c) high-bias, low-variance, and (d) high-bias, high-variance. A low-biased, high-variance model is called overfit and a high-biased, low-variance model is called underfit. A trained model achieves the best performance, through generalization, when the best trade-off between underfitting and overfitting is found. Learning with good accuracy is good, but generalization is what matters most. A good model is supposed to have both low bias and low variance. Overfitting and underfitting should both be avoided, where regularization may help.

Generalization: In ML/DL, generalization refers to the ability of a trained ML/DL model to react to new (i.e., previously unseen) data, drawn from the same distribution as the training data used to create the model. That is, after being trained on a training set, an ML/DL model can digest new data and make accurate predictions. The generalizability of an ML/DL model is central to the success of that model.

Overfitting vs. underfitting: Variance and bias are two important terms in ML. Variance refers to the variety of predicted values made by an ML model (target function). Bias means the distance of the predictions from the actual (true) target values. A high-biased model means its prediction values (average) are far from the actual values. In addition, high-variance prediction means the prediction values are highly varied.

If an ML/DL model has been trained too well on training data, it will be unable to generalize. It will make inaccurate predictions when given new data, making the model useless even though it is able to make accurate predictions for the training data. This is called overfitting. Underfitting happens when a model has not been trained enough on the

data. Underfitting models are not useful either, as they are not capable of making accurate predictions, even with the training data.

Low error rates and a high variance are good indicators of overfitting. To avoid overfitting, part of the training dataset is typically set aside as the "test set" to check whether a trained model is overfitting. If the training data has a low error rate and the test data has a high error rate, it signals overfitting. An overfit model would have very low training error on seen training data but very high error from unseen datasets (e.g., testing dataset and new datasets beyond training and testing data). This is because the model maps the training set perfectly and any deviation from the training set would result in errors. An underfit model has high training error in training data and testing error in testing data and thus in new unseen data. This is because the model cannot generalize the training data correctly. Thus, the model will have a very high training error.

Regularization (also known as shrinkage): When an ML/DL model becomes too complex, it is most likely to suffer from *overfitting*. To avoid overfitting, regularization is a collection of methods to constrain and make an ML/DL model simpler and less flexible. Specifically, regularization methods are used to avoid high variance (i.e., bias/underfitting) and overfitting and thus to increase generalization. Intuitively, it follows that the function the model represents is simpler, less unsteady. Thus, predictions are smoother, and overfitting is less likely. Certain approaches are applied to different ML algorithms, for example, pruning for DT, dropout techniques for NN, and adding a penalty parameter to the cost function in regression.

Google Earth (GE): GE is a computer software, formerly known as Keyhole Earth-Viewer, that renders a 3D representation of Earth based primarily on satellite imagery. It has a web version at https://earth.google.com/web/, accessed on 2 January 2022. Since GE version 4.3, Google fully integrated Street View into Google Earth. Street View displays 360° panoramic street-level photos of select cities and their surroundings. The photos were taken by cameras mounted on automobiles, can be viewed at different scales and from many angles, and are navigable by arrow icons imposed on them.

Google Earth Engine (GEE) and Microsoft Planetary Computer (MPC):

GEE and MPC share similar goals (e.g., cloud storage and computing support for geospatial datasets), but have their own primary focus. For example, GEE is the pioneer in the area of RS cloud computing (launched in 2010, has 495 datasets in total as of 22 December 2021), and MPC, launched in 2020 (contains 17 datasets in total as of 22 December 2021), with a primary focus on climate change and sustainable environmental studies.

GEE [131,132]: GEE is a cloud-based platform for planetary-scale geospatial analysis, launched in 2010 by Google. GEE combines a multipetabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities. Scientists, researchers, and developers use GEE to detect changes, map trends, and quantify differences on the Earth's surface. GEE brings Google's massive computational capabilities to bear a variety of high-impact societal problems (e.g., deforestation, drought, disaster, disease, food security, water management, climate monitoring, and environmental protection). GEE has been available for commercial use from 2021 and remains free for academic and research use.

MPC [133,134]: The world lacks comprehensive, global environmental data. Microsoft Chief Environmental Officer (CEO), Dr. Lucas Joppa, imagines an international database that would provide the world with "information about every tree, every species, all of our natural resources". Microsoft President Brad Smith further emphasized that "it should be as easy for anyone in the world to search the state of the planet as it is to search the internet for driving directions or dining options", and Microsoft believes technology and AI is the key to get there, in hopes that this information will allow people to "come together and solve some of the greatest environmental and sustainability challenges we face today".

To support sustainability decision-making with the power of cloud computing and AI, similar to GEE, since December 2020, Microsoft is using ML and computing power to aggregate global environmental data (contributed by individuals around the world coupled with machinery placed in water, space, land, and air environments) into a planetary

computer for a sustainable future. MPC, described as a "global portfolio of applications connecting trillions of data points", is designed to use AI to synthesize environmental data into practical information regarding the Earth's current ecosystems. For the first time, there will be a concise and comprehensive compendium of international ecosystem data. Not only will this allow for essential environmental information to be readily available to individuals across the world, but the planetary computer will predict future environmental trends through ML. In short, MPC integrates a multipetabyte catalog of global environmental data with APIs, a flexible scientific computing environment that allows people to answer global questions about that data, and applications that place those answers in the hands of conservation stakeholders.

Image classification: The concept of image classification in RS and ML/DL settings has different meanings. In RS research, the image classification is at pixel level (this is what semantic segmentation does in CV, ML, and DL settings; see the concept definition below). In contrast, in an ML and DL setting, image classification does not refer to assigning each individual pixel to a class (e.g., vegetation, water), but rather to assign the entire image to a specific class (e.g., flooded vs. not flooded) [135].

Instance segmentation: Unlike semantic segmentation, instance segmentation identifies each object instance of each pixel for every known object within an image. Thus, labels are instance-aware. Instance segmentation is essential to tasks such as counting the number of objects and reasoning about occlusion.

Normalized difference moisture index (NDMI) [136,137]: Normalized difference moisture index (NDMI) is a satellite-derived index from the near-infrared (NIR) and short wave infrared (SWIR) channels of RS imagery (note that some literature used NDMI interchangeably with NDWI; check the NDWI entry in this Appendix B for clarification).

NDMI is sensitive to the moisture levels in vegetation, and thus used to determine vegetation water content. It can be used to monitor droughts as well as monitor fuel levels in fire-prone areas. NDMI uses NIR and SWIR bands to create a ratio designed to mitigate illumination and atmospheric effects. It is calculated as a ratio between the NIR and SWIR values from RS imagery, see the formula below. For example, in Landsat 4–7, NDMI = (Band 4 − Band 5)/(Band 4 + Band 5). In Landsat 8, NDMI = (Band 5 − Band 6)/(Band 5 + Band 6). Delivered NDMI is a single band image. Similar to NDVI, NDMI values are between −1 and 1.

$$\text{NDMI} = (\text{NIR} - \text{SWIR})/(\text{NIR} + \text{SWIR})$$

Normalized difference vegetation index (NDVI) [138]: NDVI is a pixel-wise mathematical calculation rendered on an image. It is an indicator of plant health, calculated by comparing the values of absorption and reflection of red and near-infrared (NIR) light. A single NDVI value can be determined for every pixel in an image, ranging from an individual leaf to a 500-acre wheat field, depending on the RS imagery resolution.

$$\text{NDVI} = (\text{NIR} - \text{Red})/(\text{NIR} + \text{Red})$$

NDVI values always fall between −1 and 1. Values between −1 and 0 indicate dead plants, or inorganic objects (e.g., water surfaces, manmade structures such as houses, stones/rocks, roads, clouds, snow). Bare soil usually falls within 0.1–0.2 range; and plants will always have positive values between 0.2 and 1 (1 being the healthiest plants). Healthy, dense vegetation canopy should be above 0.5, and sparse vegetation will most likely fall within 0.2 to 0.5. However, it is only a rule of thumb and we should always take into account the season, type of plant, and regional peculiarities to meaningfully interpret NDVI values.

Normalized difference water index (NDWI) and modified NDWI (MNDWI) [139–141]: The NDWI is an RS-based indicator sensitive to the change in the water content of leaves or water content in water bodies (detailed below). There are two versions of NDWI.

One was defined to monitor changes in water content of leaves, using near-infrared (NIR) and short-wave infrared (SWIR) wavelengths, proposed by Gao in 1996 [139] (to avoid confusion of the two versions of NDWI, this version is also called NDMI, see NDMI entry in this Appendix B).

$$\text{NDWI} = (\text{NIR} - \text{SWIR})/(\text{NIR} + \text{SWIR})$$

The other version of NDWI, proposed by McFeeters in 1996, was defined to monitor changes related to water content in water bodies, using green and NIR wavelengths [140]. The calculation formula is given below. It is obvious that the NDWI in the papers we reviewed in this article is the version of water content in water bodies. Modification of normalized difference water index (MNDWI) was proposed [141] for improved detection of open water by replacing NIR spectral band with SWIR.

$$\text{NDWI} = (\text{Green} - \text{NIR})/(\text{Green} + \text{NIR})$$

PyTorch [142]: PyTorch is an open-source deep learning framework developed and maintained by Facebook Artificial Intelligence Research (FAIR). At its core, PyTorch is a mathematical library that performs efficient computation and automatic differentiation on graph-based models. Achieving this directly is challenging, although thankfully, the modern PyTorch API provides classes and methods that allow you to easily develop a suite of deep learning models.

Random forest (RF): It is an ML (particularly, ensemble learning) algorithm that can be used for both continuous (regression) and categorical (classification) tasks [143]. RF is widely accepted as an efficient ensemble approach for land cover classification using RS data. It handles imbalanced data, missing values, and outliers well [144].

Semantic segmentation: In contrast to instance segmentation, semantic segmentation aims to predict categorical labels for each pixel for every known object within an image, without differentiating object instances [145]. Thus, its labels are class-aware.

Support vector machine (SVM): SVM is a (supervised) machine learning algorithm that provides solutions for both classification and regression problems. The support-vector clustering [146] algorithm applies the statistics of support vectors (developed in the support vector machine algorithm) to categorize unlabeled data and is one of the most widely used clustering algorithms in many applications.

TensorFlow: TensorFlow is an open-source deep learning framework developed and maintained by Google. Although using TensorFlow directly can be challenging, the modern tf.keras API brings the simplicity and ease of use of Keras to the TensorFlow project.

Transfer learning (TL): TL is one powerful technique that makes learning in (deep) ML transferable. TL was initially proposed in [147] and recently received considerable attention due to recent significant advances in DL [123,148–152]. Inspired by humans' capabilities to transfer knowledge across domains (e.g., the knowledge gained while learning violin can be helpful to learn piano faster), TL aims to leverage learned knowledge from a related domain to achieve a desirable learning performance with minimized number of labeled samples in a target domain [151]. The main idea behind TL is that it is more efficient to take a DL model trained on an (unrelated) massive image dataset (e.g., ImageNet [87]) in one domain, and transfer its knowledge to a smaller dataset in another domain instead of training a DL classifier from scratch [153], as there are universal, low-level features shared between images for different problems.

U-Net: CNNs gave decent results in easier image segmentation problems but have not made any good progress on complex ones. This is where UNet comes in. UNet was first designed especially for medical image segmentation in [154]. It demonstrated such good results that it was used in many other fields afterwards. UNet is an improved architecture developed for biomedical image segmentation [154]. The UNet architecture stems from a fully convolutional network (FCN) first proposed by Long and Shelhamer in [155] and its architecture was modified and extended to work with fewer training images and to yield

more precise segmentations. The architecture of UNet resembles a "U", which justifies its name.

The UNet architecture includes three sections: the contraction, the bottleneck, and the expansion section. The bottommost layer mediates between the contraction layer and the expansion layer. The number of expansion blocks is the same as the number of contraction blocks. Most importantly, UNet uses a novel loss weighting scheme for each pixel such that there is a higher weight at the border of segmented objects. Specifically, all pixel-wise softmax applied on the resultant image is followed by a cross-entropy loss function. Each pixel is classified into one of the classes. The idea is that even in segmentation, every pixel has to lie in some predefined category. Thus, a segmentation problem was converted into a multiclass classification and it performed very well compared to the traditional loss functions.

Appendix C. Common Evaluation Metrics in AI/ML/DL/CV Classification and Regression, and Segmentation Tasks

Many evaluation criteria have been proposed and are frequently used to assess the performance of AI/ML/DL/CV models. No single evaluation metric can tell a full story of a trained model. To better select appropriate evaluation metrics for certain domain problems and tasks, in this appendix, we provide brief definitions to some commonly used evaluation metrics (ordered alphabetically; referenced [123,129,130,156,157]) in AI/ML/DL/CV for classification, regression, and segmentation tasks in our review (i.e., those listed in the field of "Evaluation metrics" in Tables 2 and 3). For readability, we group some related metrics together. In the following formulas, TP refers to true positive, FP to false positive, FN to false negative, and TN to true negative. TP samples are those that are in the positive category and are correctly predicted as positive. FPs are not annotated as the positive category but are incorrectly predicted as positive. TNs are correctly predicted as negative, while FNs are predicted as negative when they are actually labeled as positive.

Accuracy, overall accuracy (OA), commission error (CE), omission error (OE), producer's accuracy (PA), user's accuracy (UA), and pixel accuracy (PixA) [31,156,158–161]:

To better understand the metrics in this group, let us use the same confusion matrix shown below in Figure A2 to calculate the accuracy metrics in this group. Confusion matrix, also called error matrix, is a table that allows us to visualize the performance of a classification algorithm by comparing the predicted value of the target variable with its actual value [162].

		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	**27**
	Forest	5	31	1	**37**
	Urban	7	2	22	**31**
	Total	**33**	**39**	**23**	***95***

Figure A2. Example confusion matrix. The classified data indicate the ML/DL model predicted results and the reference data refer to the actual manually annotated data (image source: [161]).

(Average) Accuracy: Classification accuracy is the number of correct predictions made as a ratio of all predictions made. Accuracy with a binary classifier is measured as the following:

$$\text{Accuracy (for binary classifier)} = (TP + TN)/(TP + TN + FP + FN)$$

Note, however, that (average) accuracy for a multiclass classifier is calculated as the average of each accuracy per category (i.e., sum of accuracy for each category/number of categories) (see the definition and examples of binary classification and multiclass classification in Appendix A4 in [84]). For the example confusion matrix shown in Figure A2 (it is a multiclass classification problem), the (average) accuracy is calculated as follows:

$$\text{(average) accuracy} = (21/27 + 31/37 + 22/31)/3 = 77.5\%$$

Accuracy is perhaps the most common evaluation metric for classification problems, and it is also the most misused. It is really only meaningful and appropriate when there are an equal number of observations in each category and that all predictions and prediction errors are equally important, which is often not the case. Accuracy alone cannot tell a full meaningful story of the ML/DL models, especially when a dataset encounters a severe data imbalance problem (detailed in [86]); other metrics, such as F-score, need to tell whether an ML/DL is not suffering from overfitting when the trained model has very high accuracy.

OA: It essentially tells us out of all of the samples what proportion were classified correctly. OA is usually expressed as a percent, with 100% accuracy being a perfect classification where all samples were classified correctly. OA is the easiest to calculate and understand but ultimately only provides very basic accuracy information. OA is formally defined as follows, where N is the number of total samples. OA calculation from the example confusion matrix in Figure A2 is (21 + 31+ 22)/95 = 74/95 = 77.9%

$$\text{OA} = \textit{Number of correctly classified samples}/N = (TP + TN)/N$$

OE [31]: Errors of omission refer to samples that were left out (or omitted, as its name implies) from the correct category in the classified results. An example of OE is when pixels of a certain thing (such as maple trees), are not classified as maple trees.

OE is sometimes also referred to as a type II error (false negative). An OE in one category will be counted as a CE in another category. OEs are calculated by reviewing the reference sites for incorrect classifications. In the example confusion matrix shown in Figure A2, this is carried out by going down the columns for each category and adding together the incorrect classifications and dividing them by the total number of samples for each category. A separate OE is generally calculated for each category, as this will allow us to evaluate the classification accuracy and error for each category. OE is the inverse of the PA (i.e., OE = 1 − PA).

OE example based on the confusion matrix shown in Figure A2:

Water: Incorrectly classified reference sites: 5 + 7 = 12. Total # of reference sites = 33.

$$\text{OE} = 12/33 = 36\%$$

Forest: Incorrectly classified reference sites: 6 + 2 = 8. Total # of reference sites = 39.

$$\text{OE} = 8/39 = 20\%$$

Urban: Incorrectly classified reference sites: 0 + 1 = 1. Total # of reference sites = 23.

$$\text{OE} = 1/23 = 4\%$$

CE [31]: Errors of commission are in relation to the classified results. An example of an CE is when a pixel predicts the presence of a feature (such as trees) and, in reality, it is

absent (no trees are actually present). CE is sometimes also referred to as a type I error (false positive). CEs are calculated by reviewing the classified sites for incorrect classifications. This is performed by going across the rows for each class and adding together the incorrect classifications and dividing them by the total number of classified sites for each class. CE is the inverse of the UA (i.e., CE = 1 − UA). This makes sense and is easy to interpret, as when the predicted results are very reliable (with high UA score), the classification error would be low.

CE example based on the confusion matrix shown in Figure A2:

Water: Incorrectly classified sites: 6 + 0 = 6. Total # of classified sites = 27.

$$CE = 6/27 = 22\%$$

Forest: Incorrectly classified sites: 5 + 1 = 6. Total # of classified sites = 37.

$$CE = 6/37 = 16\%$$

Urban: Incorrectly classified sites: 7 + 2 = 9. Total # of classified sites = 31.

$$CE = 9/31 = 29\%$$

PA: Similar to UA, PA is category-level-based accuracy. PA is the accuracy from the point of view of the "producer". PA tells us how often real features in the ground truth are correctly shown in the classified results, or the probability that a certain ground truth category is classified as such. PA is formally defined as the following and is complement of the omission error (OE). PA = 100% − OE.

PA = Number of correctly classified reference samples for a particular category/Number of samples from reference (i.e., annotated) data for that category = 1 − omission error

PA example based on the example confusion matrix in Figure A2:

PA for water category = Correctly classified reference sites for water category/Total # of reference sites for water category = 21/33 = 64%.

PA for forest category = Correctly classified reference sites for forest category/Total # of reference sites for water category = 31/39 = 80%.

PA for urban category = Correctly classified reference sites for urban category/Total # of reference sites for uran category = 22/23 = 96%.

UA: Similar to PA, UA is category-level-based accuracy. UA is the accuracy from the point of view of a "user", not the "producer". UA essentially tells us how often the classified category will actually align with the ground truth. This is referred to as reliability (memory tip: users often care about reliability). The UA is a complement of the commission error (i.e., UA = 100% − Commission Error). UA is defined as the following:

UA = *Number of correctly classified samples for a particular category/Number of samples classified (i.e., predicted) to that category* = 1 − *commission error.*

UA example based on the example confusion matrix in Figure A2:

UA for *water* category = *Correctly classified sites for water category/Total # of classified sites for water category* = 21/27 = 78%.

UA for *forest* category = *Correctly classified sites for forest category/Total # of classified sites for water category* = 31/37 = 84%.

UA for *urban* category = *Correctly classified sites for urban category/Total # of classified sites for uran category* = 22/31 = 70%.

PixA [158]: Pixel accuracy is perhaps the easiest to understand metric conceptually. It is the percent of pixels in the image that are classified correctly. It is the simplest metric, simply computing a ratio between the amount of properly classified pixels and the total

number of pixels. See the PixA calculation formula below, where N represents the total number of pixels in the assessment image, which equals $TP + TN + FP + FN$. TP denotes the number of target-pixels that were correctly detected, FN denotes the number of water body pixels not classified, FP is the number of nontarget pixels classified, and TN is the number of nontarget pixels classified as nontarget pixels. This metric can sometimes provide misleading results when the category representation is small within the image, as the measure will be biased in mainly reporting how well the classifier identifies negative category (i.e., where the category we care about, such as the water body category, is not present).

$$\text{PixA} = (TP + TN)/N$$

Edge overall accuracy (EOA), edge commission error (ECE), and edge omission error [33]:

The authors in [33] defined a few evaluation metrics for water edge pixel extraction accuracy. See the following steps for how these metrics are computed.

1. Manually draw the boundary of water body.
2. Apply morphological expansion to the water body boundary from step (1) to create a buffer zone, which is centered on the boundary line (radius = three pixels).
3. Finally, the pixels in the buffer area are judged.

Let the total number of pixels in the buffer area be M, the number of correctly classified pixels be M_R, the number of missing pixels be M_O, and the number of false alarm pixels be M_C. EOA, EOE, and ECE are defined as below:

$$\text{EOA} = M_R/M \times 100\%$$

$$\text{EOE} = M_o/M \times 100\%$$

$$\text{ECE} = M_c/M \times 100\%$$

Intersection over union (IoU), mean intersection over union (mIoU), and frequency weighted intersection over union (FWIoU):

In the formal definitions below, TP, TN, FP, and FN are the number of true positive, true negative, false positive, and false negative samples, respectively.

IoU [163,164]: It is the most popular and simple evaluation metric for object detection and image segmentation used to measure the overlap between any two shapes such as two bounding boxes or masks (e.g., ground-truth and predicted bounding boxes). Values of IoU lie between 0 and 1, where 0 means two boxes do not intersect and 1 indicates two boxes completely overlap. If the prediction is completely correct, IoU = 1. The lower the IoU, the worse the prediction.

mIoU [43]: It is a common evaluation metric for semantic image segmentation, which first computes the IOU for each semantic class and then computes the average over classes. The formula is given below.

$$\text{mIoU} = TP/(TP + FP + FN)$$

FWIoU [46,158]: It is an improvement over mIoU. As its name implies, it weights each class importance depending on their appearance frequency. The formal definition of FWIoU is given below, where n is the number of categories.

$$\text{FWIoU} = \frac{1}{n+1} \sum_{i=0}^{n} \left(\frac{TP_i}{TP_i + TN_i + FN_i} \times \frac{TP_i + FN_i}{TP_i + FP_i + TN_i + FN_i} \right)$$

Kappa statistic [156,159,165–172]:

Kappa (aka Cohen's kappa) statistic, a statistic that is frequently used to measure inter-annotator reliability (i.e., agreement) and also intra-annotator reliability for qualitative (i.e., categorical) items, is a very useful, but underutilized, metric. The importance of rater reliability is important because it represents the extent to which the data collected in a study

are correct representations of measured variables. Note that this measure is to compare labeling by different human annotators, not a classifier versus a ground truth.

Cohen's kappa statistic is a very good measure that can handle both multiclass and imbalanced class problems very well. In ML, for a multiclass classification problem (see Appendix A4.2 in [84] for the definition and other types of classification tasks), measures such as accuracy, precision, or recall do not provide the complete picture of the performance of a classifier. In addition, for imbalanced class problems (see section II.D Imbalanced data in [86] for details about data imbalance), measures such as accuracy are misleading, so measures such as precision and recall are used. There are different ways to combine the two, such as the F-measure, but the F-measure does not have a very good intuitive explanation, other than it being the harmonic mean of precision and recall.

The kappa statistic can be calculated by the following formula, where Pr(a) represents the actual observed agreement, and Pr(e) represents expected (i.e., estimated) chance agreement). Thus, Pr(a) = OA.

$$\text{Kappa Statistic} = (\text{Pr(a)} - \text{Pr(e)})/(1 - \text{Pr(e)})$$

Note that the sample size consists of the number of observations made across which raters are compared. Cohen specifically discussed two raters in his papers. The kappa is based on the chi-square table, and the Pr(e) is obtained through the following formula [166], where: cm^1, cm^2, rm^1, rm^2 represent column 1 marginal, column 2 marginal, row 1 marginal, row 2 marginal, respectively, and n represents the number of observations (not the number of raters).

$$\text{Expected (Chance) Agreement} = \frac{\left(\frac{cm^1 \times rm^1}{n}\right) + \left(\frac{cm^2 \times rm^2}{n}\right)}{n}$$

Similar to most correlation statistics, the kappa score can range from -1 to +1. Scores above 0.8 are generally considered good agreement; zero or lower mean no agreement (practically random labels). According to the scheme of [165], a value of <0 indicates no agreement, 0–0.20 is slight, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial, and 0.81–1 is almost perfect agreement.

Kappa is one of the most commonly used statistics to test interrater reliability, but it has limitations. Judgments about what level of kappa should be acceptable for health research are questioned. Cohen's suggested interpretation may be too lenient for health-related studies because it implies that a score as low as 0.41 might be acceptable [166]. Additional measures have been proposed to make use of the kappa framework.

For example, in [159], the authors advocate against the use of kappa and proposed the alternative measures of quantity and allocation disagreement. Quantity disagreement (QD) is the disagreement between the classification and reference data resulting from a difference in proportion of categories. Allocation disagreement (AD) assesses a difference in the spatial location of categories. The two measures (i.e., QD and AD) sum to overall error (i.e., 1–OA).

Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) [123, 129,130,173]:

MAE: also called mean absolute deviation, MAE finds the average of the absolute differences between actual and predicted values. It gives an idea of how wrong the predictions were. MAE measure gives an idea of the magnitude of the error, but no idea of the direction (e.g., over- or underpredicting). MAE is defined as below [174], where y_i is the actual true value, and $\hat{y}_i$ is the predicted value. MAE value lies between 0 to ∞. Small value indicates a better model, and a value of 0 indicates no error, or perfect predictions.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

MAE is more robust to the outliers than MSE, as it is not sensitive to outliers. MAE treats larger and small errors equally. The main reason is that in MSE, through squaring

the errors, the outliers, which usually have higher errors than other samples, obtain more attention and dominance in the final error and thus impact the model parameters. In addition, there is an intuitive maximum likelihood (MLE) interpretation behind MSE and MAE metrics. If we assume a linear dependence between features and targets, then MSE and MAE correspond to the MLE on the model parameters by assuming Gaussian and Laplace priors on the model errors, respectively.

MAPE [175]: MAPE, also known as mean absolute percentage deviation (MAPD), is the mean or average of the absolute percentage errors of forecasts. Error is defined as actual value (i.e., observed value) minus the forecasted value. Percentage errors are summed without regard to sign to compute MAPE. It is the most common measure used to forecast error and works best if there are no extremes to the data (and no zeros). Because absolute percentage errors are used, it avoids the problem of canceling positive and negative errors. The formula is given below, where M is mean absolute percentage error, n is number of times the summation iteration happens, At is the actual value, and Ft is the forecast value. The smaller the MAPE, the better the forecast.

$$M = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

Precision, recall, sensitivity, specificity, and F-score [156]:

Each measure in this group is a set-based measure [176]. The values of those measures are all from 0 to 1, with the best value at 1 and the worst score at 0.

Precision: The precision is mathematically defined by the following formula. Precision attempts to answer the question What proportion of positive identifications was actually correct? Precision refers to the proportion of the samples that is correctly classified amongst the samples predicted to be positive and is equivalent to user's accuracy (UA) for the positive category, which is also equivalent to 1 − commission error.

$$\text{Precision} = TP/(TP + FP)$$

Recall (also called sensitivity or true positive rate): it refers to the proportion of the reference data for the positive category that is correctly classified and is equivalent to producer's accuracy (also equivalent to 1 − omission error) for the positive category. It is calculated by the following formula. Recall attempts to answer the following question: What proportion of actual positives was identified correctly?

$$\text{Recall} = TP/(TP + FN)$$

Specificity (also called true negative rate): it refers to the proportion of negative samples that is correctly predicted and is equivalent to the producer's accuracy (PA) for the negative category [177].

$$\text{Specificity} = TN/(TN + FP)$$

F-score (also called F1-score, F measure): Depending on the application domain, we may need to give a higher priority to recall or precision, but there are many applications where both recall and precision are important. Thus, it is natural to think of a way to combine these two metrics into a single one. One popular metric that combines precision and recall is called F1-score. The F1-score can be interpreted as a weighted harmonic mean of the precision and recall and is formally defined as below. There is always a trade-off between precision and recall of a model; if making the precision too high, we would see a drop in the recall rate, and vice versa.

$$\text{F1-score} = (2 \times Precision \times Recall)/(Precision + Recall)$$

The generalized version of F-score is defined as follows. F1-score is a special case of F_β when β = 1.

$$F_\beta = \left(1 + \beta^2\right) \times \frac{precision \times recall}{\beta^2 \times precision + recall}$$

R^2, mean squared error (MSE), root mean squared error (RMSE), and root mean squared logarithmic error (RMSLE) [123,129,130,173]:

R^2 is based on correlation between actual and predicted value; MAE is based on absolute value of error; MSE and RMSE are both based on square of error.

R^2: R-squared, also known as the coefficient of determination, is a value between 0 and 1 that measures how well a regression line fits the data (i.e., indication of the goodness of fit of a set of predictions to the actual values in a regression model). The value range of R^2 lies between 0 and 1 for no-fit and perfect fit, respectively. R^2 is not sensitive to outliers.

The R-squared formula compares our fitted regression line to a baseline model. This baseline model is considered the "worst" model. The baseline model is a flat line that predicts that every value of y will be the mean value of y. R-squared checks to see if our fitted regression line will predict y better than the mean.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

SS_{RES} refers to the residual sum of squared errors of the regression model; y_i is the actual value, and $\hat{y}_i$ is the predicted value through the regression model. For example, if the actual y value was 58 but we had predicted it would be 47 then the residual squared error would be 121 and we would add that to the rest of the residual squared errors for the model.

SS_{TOT} is the total sum of squared errors. This compares the actual y values to the baseline model (i.e., the mean). We square the difference between all the actual y values and the mean $\overline{y}$ and add them together.

MSE: MSE is perhaps the most popular metric used for regression problems. It essentially finds the mean (i.e., average) of the square of the difference (i.e., squared error) between actual and estimated values. Similar to MAE, MSE provides a gross idea of the magnitude of error. Let us assume we have a regression model that predicts the price of houses in the Boston area and let us say for each house we also have the actual price the house was sold for. The MSE can be calculated as the following, where N is the number of samples, y_i is the actual house price, and $\hat{y}_i$ is the predicted value through the regression model. MSE value lies between 0 to ∞. Small value indicates a better model. Sensitive to outliers, it punishes larger errors more. MSE incorporates both the variance and the bias of the predicting model.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

MSE measures how far the data are from the model's predicted values, whereas R^2 measures how far the data are from the model's predicted values compared to how far the data are from the mean. The difference between how far the data are from the model's predicted values and how far the data are from the mean is the improvement in prediction from the regression model.

RMSE: very straightforward, RMSE is the square root of MSE. Sometimes people use RMSE to have a metric with scale as the target values. Taking the house pricing prediction example, RMSE essentially shows what is the average deviation in your model predicted house prices from the target values (the prices the houses are sold for). Similar to MSE, RMSE value lies between 0 to ∞, with a small value indicating a better model. Similar to MSE, RMSE is sensitive to outliers and punishes larger errors more. The value of RMSE is always greater than or equal to MAE (RMSE >= MAE). The greater difference between them indicates greater variance in individual errors in the sample.

RMSLE: both RMSE and RMSLE are the techniques to find out the difference between the actual values and the predicted values by an ML/DL model. RMSLE is the root mean squared error of the log-transformed predicted and log-transformed actual values. RMSLE is formally defined as follows, where X denotes the predicted value and Y denotes the actual value, and n is the number of samples. Note that RMSLE adds 1 to both actual and predicted values before taking the natural logarithm to avoid taking the natural log of possible 0 (zero) values.

$$\text{RMLSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(log(x_i+1) - \log(y_i+1))^2}$$

RMSLE is very robust to outliers. When we compare the formula of the RMSE and RMSLE, the only difference is the log function. Basically, what changes is the variance measured. This small difference makes RMSLE much more robust to outliers than RMSE. In RMSE, outliers can explode the error term to a very high value, but in RMLSE, the outliers are drastically scaled down, therefore nullifying their effect.

RMSLE is often used when we do not want to penalize huge differences in the predicted and the actual values when both predicted and true values are huge numbers. (1) If both predicted and actual values are small: RMSE and RMSLE is same. (2) If either predicted or the actual value is big: RMSE > RMSLE. (3) If both predicted and actual values are big: RMSE > RMSLE (RMSLE becomes almost negligible).

References

1. UN Water. Climate Change Adaptation: The Pivotal Role of Water (2010). UN Water. 2010. Available online: https://www.unwater.org/publications/climate-change-adaptation-pivotal-role-water/#:~{}:text=Higher%20temperatures%20and%20changes%20in,likely%20to%20be%20adversely%20affected (accessed on 15 December 2021).
2. U.S. Bureau of Reclamation California-Great Basin Area Office Water Facts—Worldwide Water Supply. Available online: https://www.usbr.gov/mp/arwec/water-facts-ww-water-sup.html (accessed on 3 December 2021).
3. Reidmiller, D.R.; Avery, C.W.; Easterling, D.R.; Kunkel, K.E.; Lewis, K.L.M.; Maycock, T.K.; Stewart, B.C. *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment*; U.S. Global Change Research Program: Washington, DC, USA, 2018; Volume II. [CrossRef]
4. IPCC (Intergovernmental Panel on Climate Change). *Climate Change 2014–Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report*; Cambridge University Press: Cambridge, UK, 2014; ISBN 9781107058071.
5. Steffen, W.; Richardson, K.; Rockström, J.; Cornell, S.E.; Fetzer, I.; Bennett, E.M.; Biggs, R.; Carpenter, S.R.; de Vries, W.; de Wit, C.A.; et al. Planetary Boundaries: Guiding Human Development on a Changing Planet. *Science* **2015**, *347*, 1259855. [CrossRef] [PubMed]
6. Rockström, J.; Steffen, W.; Noone, K.; Persson, A.; Chapin, F.S., 3rd; Lambin, E.F.; Lenton, T.M.; Scheffer, M.; Folke, C.; Schellnhuber, H.J.; et al. A Safe Operating Space for Humanity. *Nature* **2009**, *461*, 472–475. [CrossRef] [PubMed]
7. Walker, B.; Salt, D. *Resilience Thinking: Sustaining Ecosystems and People in a Changing World*; Island Press: Washington, DC, USA, 2006; ISBN 9781597266222.
8. Jiménez Cisneros, B.E.; Oki, N.W.; Arnell, G.; Benito, J.G.; Cogley, P.; Döll, T.; Jiang, S.S. Mwakalila Freshwater Resources. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., et al., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2014; pp. 229–269, ISBN 9781107058163.
9. Yamazaki, D.; Trigg, M.A.; Ikeshima, D. Development of a Global ~90m Water Body Map Using Multi-Temporal Landsat Images. *Remote Sens. Environ.* **2015**, *171*, 337–351. [CrossRef]
10. Jiang, W.; He, G.; Long, T.; Ni, Y. Detecting Water Bodies In Landsat8 Oli Image Using Deep Learning. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-3*, 669–672. [CrossRef]
11. Shao, Z.; Fu, H.; Li, D.; Altan, O.; Cheng, T. Remote Sensing Monitoring of Multi-Scale Watersheds Impermeability for Urban Hydrological Evaluation. *Remote Sens. Environ.* **2019**, *232*, 111338. [CrossRef]
12. Wang, X.; Xie, H. A Review on Applications of Remote Sensing and Geographic Information Systems (GIS) in Water Resources and Flood Risk Management. *Water* **2018**, *10*, 608. [CrossRef]
13. El Serafy, G.Y.H.; Schaeffer, B.A.; Neely, M.-B.; Spinosa, A.; Odermatt, D.; Weathers, K.C.; Baracchini, T.; Bouffard, D.; Carvalho, L.; Conmy, R.N.; et al. Integrating Inland and Coastal Water Quality Data for Actionable Knowledge. *Remote Sens.* **2021**, *13*, 2899. [CrossRef]

14. Brown, C.M.; Lund, J.R.; Cai, X.; Reed, P.M.; Zagona, E.A.; Ostfeld, A.; Hall, J.; Characklis, G.W.; Yu, W.; Brekke, L. The Future of Water Resources Systems Analysis: Toward a Scientific Framework for Sustainable Water Management. *Water Resour. Res.* **2015**, *51*, 6110–6124. [CrossRef]
15. Zhang, X.; Zhou, Y.; Luo, J. Deep Learning for Processing and Analysis of Remote Sensing Big Data: A Technical Review. *Big Earth Data* **2021**, 1–34. [CrossRef]
16. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
17. Hoeser, T.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sens.* **2020**, *12*, 1667. [CrossRef]
18. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Prabhat Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
19. Boyd, C.E. *Water Quality: An Introduction*; Springer Nature: Cham, Switzerland, 2019; ISBN 9783030233358.
20. Ahuja, S. *Monitoring Water Quality: Pollution Assessment, Analysis, and Remediation*; Newnes: London, UK, 2013; ISBN 9780444594044.
21. Ramadas, M.; Samantaray, A.K. Applications of Remote Sensing and GIS in Water Quality Monitoring and Remediation: A State-of-the-Art Review. In *Water Remediation*; Bhattacharya, S., Gupta, A.B., Gupta, A., Pandey, A., Eds.; Springer: Singapore, 2018; pp. 225–246, ISBN 9789811075513.
22. Bijeesh, T.V.; Narasimhamurthy, K.N. Surface Water Detection and Delineation Using Remote Sensing Images: A Review of Methods and Algorithms. *Sustain. Water Resour. Manag.* **2020**, *6*, 68. [CrossRef]
23. Gholizadeh, M.H.; Melesse, A.M.; Reddi, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors* **2016**, *16*, 1298. [CrossRef]
24. Sibanda, M.; Mutanga, O.; Chimonyo, V.G.P.; Clulow, A.D.; Shoko, C.; Mazvimavi, D.; Dube, T.; Mabhaudhi, T. Application of Drone Technologies in Surface Water Resources Monitoring and Assessment: A Systematic Review of Progress, Challenges, and Opportunities in the Global South. *Drones* **2021**, *5*, 84. [CrossRef]
25. Sit, M.; Demiray, B.Z.; Xiang, Z.; Ewing, G.J.; Sermet, Y.; Demir, I. A Comprehensive Review of Deep Learning Applications in Hydrology and Water Resources. *Water Sci. Technol.* **2020**, *82*, 2635–2670. [CrossRef]
26. Doorn, N. Artificial Intelligence in the Water Domain: Opportunities for Responsible Use. *Sci. Total Environ.* **2021**, *755*, 142561. [CrossRef]
27. Hassan, N.; Woo, C.S. Machine Learning Application in Water Quality Using Satellite Data. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *842*, 012018. [CrossRef]
28. Li, M.; Xu, L.; Tang, M. An Extraction Method for Water Body of Remote Sensing Image Based on Oscillatory Network. *J. Multimed.* **2011**, *6*, 252–260. [CrossRef]
29. Yang, L.; Tian, S.; Yu, L.; Ye, F.; Qian, J.; Qian, Y. Deep Learning for Extracting Water Body from Landsat Imagery. *Int. J. Innov. Comput. Inf. Control* **2015**, *11*, 1913–1929.
30. Huang, X.; Xie, C.; Fang, X.; Zhang, L. Combining Pixel- and Object-Based Machine Learning for Identification of Water-Body Types From Urban High-Resolution Remote-Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2097–2110. [CrossRef]
31. Isikdogan, F.; Bovik, A.C.; Passalacqua, P. Surface Water Mapping by Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4909–4918. [CrossRef]
32. Yu, L.; Wang, Z.; Tian, S.; Ye, F.; Ding, J.; Kong, J. Convolutional Neural Networks for Water Body Extraction from Landsat Imagery. *Int. J. Comput. Intell. Appl.* **2017**, *16*, 1750001. [CrossRef]
33. Chen, Y.; Fan, R.; Yang, X.; Wang, J.; Latif, A. Extraction of Urban Water Bodies from High-Resolution Remote-Sensing Imagery Using Deep Learning. *Water* **2018**, *10*, 585. [CrossRef]
34. Miao, Z.; Fu, K.; Sun, H.; Sun, X.; Yan, M. Automatic Water-Body Segmentation From High-Resolution Satellite Images via Deep Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 602–606. [CrossRef]
35. Acharya, T.D.; Subedi, A.; Lee, D.H. Evaluation of Machine Learning Algorithms for Surface Water Extraction in a Landsat 8 Scene of Nepal. *Sensors* **2019**, *19*, 2769. [CrossRef] [PubMed]
36. Feng, W.; Sui, H.; Huang, W.; Xu, C.; An, K. Water Body Extraction From Very High-Resolution Remote Sensing Imagery Using Deep U-Net and a Superpixel-Based Conditional Random Field Model. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 618–622. [CrossRef]
37. Li, L.; Yan, Z.; Shen, Q.; Cheng, G.; Gao, L.; Zhang, B. Water Body Extraction from Very High Spatial Resolution Remote Sensing Data Based on Fully Convolutional Networks. *Remote Sens.* **2019**, *11*, 1162. [CrossRef]
38. Li, Z.; Wang, R.; Zhang, W.; Hu, F.; Meng, L. Multiscale Features Supported DeepLabV3+ Optimization Scheme for Accurate Water Semantic Segmentation. *IEEE Access* **2019**, *7*, 155787–155804. [CrossRef]
39. Meng, X.; Zhang, S.; Zang, S. Lake Wetland Classification Based on an SVM-CNN Composite Classifier and High-Resolution Images Using Wudalianchi as an Example. *J. Coast. Res.* **2019**, *93*, 153–162. [CrossRef]
40. Isikdogan, L.F.; Bovik, A.; Passalacqua, P. Seeing Through the Clouds With DeepWaterMap. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1662–1666. [CrossRef]

41. Song, S.; Liu, J.; Liu, Y.; Feng, G.; Han, H.; Yao, Y.; Du, M. Intelligent Object Recognition of Urban Water Bodies Based on Deep Learning for Multi-Source and Multi-Temporal High Spatial Resolution Remote Sensing Imagery. *Sensors* **2020**, *20*, 397. [CrossRef]
42. Yang, F.; Feng, T.; Xu, G.; Chen, Y. Applied Method for Water-Body Segmentation Based on Mask R-CNN. *JARS* **2020**, *14*, 014502. [CrossRef]
43. Wang, G.; Wu, M.; Wei, X.; Song, H. Water Identification from High-Resolution Remote Sensing Images Based on Multidimensional Densely Connected Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 795. [CrossRef]
44. O'Neil, G.L.; Goodall, J.L.; Behl, M.; Saby, L. Deep Learning Using Physically-Informed Input Data for Wetland Identification. *Environ. Model. Softw.* **2020**, *126*, 104665. [CrossRef]
45. Chen, Y.; Tang, L.; Kan, Z.; Bilal, M.; Li, Q. A Novel Water Body Extraction Neural Network (WBE-NN) for Optical High-Resolution Multispectral Imagery. *J. Hydrol.* **2020**, *588*, 125092. [CrossRef]
46. Dang, B.; Li, Y. MSResNet: Multiscale Residual Network via Self-Supervised Learning for Water-Body Detection in Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 3122. [CrossRef]
47. Yuan, K.; Zhuang, X.; Schaefer, G.; Feng, J.; Guan, L.; Fang, H. Deep-Learning-Based Multispectral Satellite Image Segmentation for Water Body Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7422–7434. [CrossRef]
48. Tambe, R.G.; Talbar, S.N.; Chavan, S.S. Deep Multi-Feature Learning Architecture for Water Body Segmentation from Satellite Images. *J. Vis. Commun. Image Represent.* **2021**, *77*, 103141. [CrossRef]
49. Yu, Y.; Yao, Y.; Guan, H.; Li, D.; Liu, Z.; Wang, L.; Yu, C.; Xiao, S.; Wang, W.; Chang, L. A Self-Attention Capsule Feature Pyramid Network for Water Body Extraction from Remote Sensing Imagery. *Int. J. Remote Sens.* **2021**, *42*, 1801–1822. [CrossRef]
50. Li, W.; Li, Y.; Gong, J.; Feng, Q.; Zhou, J.; Sun, J.; Shi, C.; Hu, W. Urban Water Extraction with UAV High-Resolution Remote Sensing Data Based on an Improved U-Net Model. *Remote Sens.* **2021**, *13*, 3165. [CrossRef]
51. Zhang, L.; Fan, Y.; Yan, R.; Shao, Y.; Wang, G.; Wu, J. Fine-Grained Tidal Flat Waterbody Extraction Method (FYOLOv3) for High-Resolution Remote Sensing Images. *Remote Sens.* **2021**, *13*, 2594. [CrossRef]
52. Li, M.; Wu, P.; Wang, B.; Park, H.; Yang, H.; Wu, Y. A Deep Learning Method of Water Body Extraction From High Resolution Remote Sensing Images With Multisensors. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3120–3132. [CrossRef]
53. Su, H.; Peng, Y.; Xu, C.; Feng, A.; Liu, T. Using Improved DeepLabv3+ Network Integrated with Normalized Difference Water Index to Extract Water Bodies in Sentinel-2A Urban Remote Sensing Images. *JARS* **2021**, *15*, 018504. [CrossRef]
54. Ovakoglou, G.; Cherif, I.; Alexandridis, T.K.; Pantazi, X.-E.; Tamouridou, A.-A.; Moshou, D.; Tseni, X.; Raptis, I.; Kalaitzopoulou, S.; Mourelatos, S. Automatic Detection of Surface-Water Bodies from Sentinel-1 Images for Effective Mosquito Larvae Control. *JARS* **2021**, *15*, 014507. [CrossRef]
55. Chebud, Y.; Naja, G.M.; Rivero, R.G.; Melesse, A.M. Water Quality Monitoring Using Remote Sensing and an Artificial Neural Network. *Water Air Soil Pollut.* **2012**, *223*, 4875–4887. [CrossRef]
56. Wang, X.; Zhang, F.; Ding, J. Evaluation of Water Quality Based on a Machine Learning Algorithm and Water Quality Index for the Ebinur Lake Watershed, China. *Sci. Rep.* **2017**, *7*, 12858. [CrossRef]
57. Lee, S.; Lee, D. Improved Prediction of Harmful Algal Blooms in Four Major South Korea's Rivers Using Deep Learning Models. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1322. [CrossRef]
58. Wang, P.; Yao, J.; Wang, G.; Hao, F.; Shrestha, S.; Xue, B.; Xie, G.; Peng, Y. Exploring the Application of Artificial Intelligence Technology for Identification of Water Pollution Characteristics and Tracing the Source of Water Quality Pollutants. *Sci. Total Environ.* **2019**, *693*, 133440. [CrossRef]
59. Pu, F.; Ding, C.; Chao, Z.; Yu, Y.; Xu, X. Water-Quality Classification of Inland Lakes Using Landsat8 Images by Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 1674. [CrossRef]
60. Liu, P.; Wang, J.; Sangaiah, A.; Xie, Y.; Yin, X. Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. *Sustainability* **2019**, *11*, 2058. [CrossRef]
61. Chowdury, M.S.U.; Emran, T.B.; Ghosh, S.; Pathak, A.; Alam, M.M.; Absar, N.; Andersson, K.; Hossain, M.S. IoT Based Real-Time River Water Quality Monitoring System. *Procedia Comput. Sci.* **2019**, *155*, 161–168. [CrossRef]
62. Hafeez, S.; Wong, M.S.; Ho, H.C.; Nazeer, M.; Nichol, J.; Abbas, S.; Tang, D.; Lee, K.H.; Pun, L. Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-II Waters: A Case Study of Hong Kong. *Remote Sens.* **2019**, *11*, 617. [CrossRef]
63. Li, L.; Jiang, P.; Xu, H.; Lin, G.; Guo, D.; Wu, H. Water Quality Prediction Based on Recurrent Neural Network and Improved Evidence Theory: A Case Study of Qiantang River, China. *Environ. Sci. Pollut. Res. Int.* **2019**, *26*, 19879–19896. [CrossRef]
64. Randrianiaina, J.J.C.; Rakotonirina, R.I.; Ratiarimanana, J.R.; Fils, L.R. Modelling of Lake Water Quality Parameters by Deep Learning Using Remote Sensing Data. *Am. J. Geogr. Inf. Syst.* **2019**, *8*, 221–227.
65. Yu, Z.; Yang, K.; Luo, Y.; Shang, C. Spatial-Temporal Process Simulation and Prediction of Chlorophyll-a Concentration in Dianchi Lake Based on Wavelet Analysis and Long-Short Term Memory Network. *J. Hydrol.* **2020**, *582*, 124488. [CrossRef]
66. Zou, Q.; Xiong, Q.; Li, Q.; Yi, H.; Yu, Y.; Wu, C. A Water Quality Prediction Method Based on the Multi-Time Scale Bidirectional Long Short-Term Memory Network. *Environ. Sci. Pollut. Res. Int.* **2020**, *27*, 16853–16864. [CrossRef]
67. Peterson, K.T.; Sagan, V.; Sloan, J.J. Deep Learning-Based Water Quality Estimation and Anomaly Detection Using Landsat-8/Sentinel-2 Virtual Constellation and Cloud Computing. *GISci. Remote Sens.* **2020**, *57*, 510–525. [CrossRef]

68. Hanson, P.C.; Stillman, A.B.; Jia, X.; Karpatne, A.; Dugan, H.A.; Carey, C.C.; Stachelek, J.; Ward, N.K.; Zhang, Y.; Read, J.S.; et al. Predicting Lake Surface Water Phosphorus Dynamics Using Process-Guided Machine Learning. *Ecol. Modell.* **2020**, *430*, 109136. [CrossRef]
69. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-Term Water Quality Variable Prediction Using a Hybrid CNN–LSTM Deep Learning Model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [CrossRef]
70. Aldhyani, T.H.H.; Al-Yaari, M.; Alkahtani, H.; Maashi, M. Water Quality Prediction Using Artificial Intelligence Algorithms. *Appl. Bionics Biomech.* **2020**, *2020*, 6659314. [CrossRef]
71. Li, X.; Ding, J.; Ilyas, N. Machine Learning Method for Quick Identification of Water Quality Index (WQI) Based on Sentinel-2 MSI Data: Ebinur Lake Case Study. *Water Sci. Technol. Water Supply* **2021**, *21*, 1291–1312. [CrossRef]
72. Sharma, C.; Isha, I.; Vashisht, V. Water Quality Estimation Using Computer Vision in UAV. In Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence), Noida, India, 28–29 January 2021; pp. 448–453.
73. Cui, Y.; Yan, Z.; Wang, J.; Hao, S.; Liu, Y. Deep Learning-Based Remote Sensing Estimation of Water Transparency in Shallow Lakes by Combining Landsat 8 and Sentinel 2 Images. *Environ. Sci. Pollut. Res. Int.* **2022**, *29*, 4401–4413. [CrossRef] [PubMed]
74. Zhao, X.; Xu, H.; Ding, Z.; Wang, D.; Deng, Z.; Wang, Y.; Wu, T.; Li, W.; Lu, Z.; Wang, G. Comparing Deep Learning with Several Typical Methods in Prediction of Assessing Chlorophyll-a by Remote Sensing: A Case Study in Taihu Lake, China. *Water Supply* **2021**, *21*, 3710–3724. [CrossRef]
75. Arias-Rodriguez, L.F.; Duan, Z.; de Díaz-Torres, J.; Basilio Hazas, M.; Huang, J.; Kumar, B.U.; Tuo, Y.; Disse, M. Integration of Remote Sensing and Mexican Water Quality Monitoring System Using an Extreme Learning Machine. *Sensors* **2021**, *21*, 4118. [CrossRef]
76. Kravitz, J.; Matthews, M.; Lain, L.; Fawcett, S.; Bernard, S. Potential for High Fidelity Global Mapping of Common Inland Water Quality Products at High Spatial and Temporal Resolutions Based on a Synthetic Data and Machine Learning Approach. *Front. Environ. Sci.* **2021**, *19*. [CrossRef]
77. Sun, X.; Zhang, Y.; Shi, K.; Zhang, Y.; Li, N.; Wang, W.; Huang, X.; Qin, B. Monitoring Water Quality Using Proximal Remote Sensing Technology. *Sci. Total Environ.* **2021**, *803*, 149805. [CrossRef] [PubMed]
78. Chen, Y.; Fan, R.; Bilal, M.; Yang, X.; Wang, J.; Li, W. Multilevel Cloud Detection for High-Resolution Remote Sensing Imagery Using Multiple Convolutional Neural Networks. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 181. [CrossRef]
79. Tong, X.-Y.; Xia, G.-S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-Cover Classification with High-Resolution Remote Sensing Images Using Transferable Deep Models. *Remote Sens. Environ.* **2020**, *237*, 111322. [CrossRef]
80. Boguszewski, A.; Batorski, D.; Ziemba-Jankowska, N.; Dziedzic, T.; Zambrzycka, A. LandCover. Ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water and Roads From Aerial Imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 1102–1110.
81. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS—A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *arXiv* **2019**, arXiv:1906.07789. [CrossRef]
82. Ross, M.R.V.; Topp, S.N.; Appling, A.P.; Yang, X.; Kuhn, C.; Butman, D.; Simard, M.; Pavelsky, T.M. AquaSat: A Data Set to Enable Remote Sensing of Water Quality for Inland Waters. *Water Resour. Res.* **2019**, *55*, 10012–10025. [CrossRef]
83. Wang, S.; Li, J.; Zhang, W.; Cao, C.; Zhang, F.; Shen, Q.; Zhang, X.; Zhang, B. A Dataset of Remote-Sensed Forel-Ule Index for Global Inland Waters during 2000–2018. *Sci. Data* **2021**, *8*, 26. [CrossRef]
84. Yang, L.; MacEachren, A.M.; Mitra, P.; Onorati, T. Visually-Enabled Active Deep Learning for (Geo) Text and Image Classification: A Review. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 65. [CrossRef]
85. Yang, L.; Gong, M.; Asari, V.K. Diagram Image Retrieval and Analysis: Challenges and Opportunities. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Virtual, 14–19 June 2020; pp. 180–181.
86. Yang, L.; MacEachren, A.M.; Mitra, P. Geographical Feature Classification from Text Using (active) Convolutional Neural Networks. In Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Virtual, 14–17 December 2020; pp. 1182–1198.
87. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
88. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
89. Maaten, L.; Chen, M.; Tyree, S.; Weinberger, K. Learning with Marginalized Corrupted Features. In Proceedings of the International Conference on Machine Learning (ICML), Atlanta, GA, USA, 16–21 June 2013; pp. 410–418.
90. Nakkiran, P.; Neyshabur, B.; Sedghi, H. The Deep Bootstrap Framework: Good Online Learners Are Good Offline Generalizers. *arXiv* **2020**, arXiv:2010.08127.
91. Montavon, G.; Samek, W.; Müller, K.-R. Methods for Interpreting and Understanding Deep Neural Networks. *Digit. Signal Process.* **2018**, *73*, 1–15. [CrossRef]
92. Li, X.; Xiong, H.; Li, X.; Wu, X.; Zhang, X.; Liu, J.; Bian, J.; Dou, D. Interpretable Deep Learning: Interpretations, Interpretability, Trustworthiness, and beyond. *arXiv* **2021**, arXiv:2103.10689.
93. Doshi-Velez, F.; Kim, B. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv* **2017**, arXiv:1702.08608.

94. Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* **2019**, *8*, 832. [CrossRef]
95. Fong, R.C.; Vedaldi, A. Interpretable Explanations of Black Boxes by Meaningful Perturbation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3429–3437.
96. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 818–833.
97. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
98. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; Bach, F., Blei, D., Eds.; pp. 2048–2057.
99. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
100. Mahendran, A.; Vedaldi, A. Understanding Deep Image Representations by Inverting Them. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5188–5196.
101. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv* **2016**, arXiv:1602.04938.
102. Laurini, R. *Geographic Knowledge Infrastructure: Applications to Territorial Intelligence and Smart Cities*; Elsevier: Oxford, UK, 2017.
103. MacEachren, A.M.; Gahegan, M.; Pike, W.; Brewer, I.; Cai, G.; Lengerich, E.; Hardisty, F. Geovisualization for Knowledge Construction and Decision Support. *IEEE Comput. Graph. Appl.* **2004**, *24*, 13–17. [CrossRef]
104. Lan, Y.; Desjardins, M.R.; Hohl, A.; Delmelle, E. Geovisualization of COVID-19: State of the Art and Opportunities. *Cartographica* **2021**, *56*, 2–13. [CrossRef]
105. MacEachren, A.M.; Cai, G. Supporting Group Work in Crisis Management: Visually Mediated Human—GIS—Human Dialogue. *Environ. Plann. B Plann. Des.* **2006**, *33*, 435–456. [CrossRef]
106. Tomaszewski, B.; MacEachren, A.M. Geovisual Analytics to Support Crisis Management: Information Foraging for Geo-Historical Context. *Inf. Vis.* **2012**, *11*, 339–359. [CrossRef]
107. Harrower, M.; MacEachren, A.; Griffin, A.L. Developing a Geographic Visualization Tool to Support Earth Science Learning. *Cartogr. Geogr. Inf. Sci.* **2000**, *27*, 279–293. [CrossRef]
108. Cova, T.J.; Dennison, P.E.; Kim, T.H.; Moritz, M.A. Setting Wildfire Evacuation Trigger Points Using Fire Spread Modeling and GIS. *Trans. GIS* **2005**, *9*, 603–617. [CrossRef]
109. Cliburn, D.C.; Feddema, J.J.; Miller, J.R.; Slocum, T.A. Design and Evaluation of a Decision Support System in a Water Balance Application. *Comput. Graph.* **2002**, *26*, 931–949. [CrossRef]
110. Kiss, E.; Zichar, M.; Fazekas, I.; Karancsi, G.; Balla, D. Categorization and Geovisualization of Climate Change Strategies Using an Open-Access WebGIS Tool. *Infocommun. J.* **2020**, *12*, 32–37. [CrossRef]
111. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-Resolution Mapping of Global Surface Water and Its Long-Term Changes. *Nature* **2016**, *540*, 418–422. [CrossRef]
112. Brodlie, K.; Fairbairn, D.; Kemp, Z.; Schroeder, M. Connecting People, Data and Resources—distributed Geovisualization. In *Exploring Geovisualization*; Elsevier: Amsterdam, The Netherlands, 2005; pp. 423–443.
113. Robinson, A.C. *Design for Synthesis in Geovisualization*; The Pennsylvania State University: Ann Arbor, MI, USA, 2008.
114. Robinson, A.C. Supporting Synthesis in Geovisualization. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 211–227. [CrossRef]
115. Andrienko, G.; Andrienko, N.; Jankowski, P.; Keim, D.; Kraak, M.-J.; MacEachren, A.; Wrobel, S. Geovisual Analytics for Spatial Decision Support: Setting the Research Agenda. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 839–857. [CrossRef]
116. Schaeffer, B.A.; Schaeffer, K.G.; Keith, D.; Lunetta, R.S.; Conmy, R.; Gould, R.W. Barriers to Adopting Satellite Remote Sensing for Water Quality Management. *Int. J. Remote Sens.* **2013**, *34*, 7534–7544. [CrossRef]
117. Smith, M.J.; Hillier, J.K.; Otto, J.-C.; Geilhausen, M. Geovisualization. In *Treatise on Geomorphology*; Shroder, J.F., Ed.; Academic Press: Cambridge, MA, USA, 2013; Volume 3, pp. 299–325.
118. Sit, M.; Sermet, Y.; Demir, I. Optimized Watershed Delineation Library for Server-Side and Client-Side Web Applications. *Open Geospat. Data Softw. Stand.* **2019**, *4*, 8. [CrossRef]
119. Mayer, T.; Poortinga, A.; Bhandari, B.; Nicolau, A.P.; Markert, K.; Thwal, N.S.; Markert, A.; Haag, A.; Kilbride, J.; Chishtie, F.; et al. Deep Learning Approach for Sentinel-1 Surface Water Mapping Leveraging Google Earth Engine. *ISPRS Open J. Photogramm. Remote Sens.* **2021**, *2*, 100005. [CrossRef]
120. Cheng, G.; Zhou, P.; Han, J. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [CrossRef]
121. CS231n: Convolutional Neural Networks for Visual Recognition. Available online: http://cs231n.stanford.edu/ (accessed on 8 December 2021).
122. Mather, P.M.; Koch, M. *Computer Processing of Remotely-Sensed Images: An Introduction*; John Wiley & Sons: Hoboken, NJ, USA, 2011; ISBN 9781119956402.
123. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
124. Chollet, F. *Deep Learning with Python*, 2nd ed.; Manning Publications Co.: Shelter Island, NY, USA, 2021; ISBN 9781617296864.

125. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
126. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
127. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
128. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
129. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
130. Hastie, T.; Friedman, J.; Tibshirani, R. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2001.
131. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
132. Google Earth Engine. Available online: https://earthengine.google.com/ (accessed on 15 December 2021).
133. Toadvine, H. Microsoft Launches a "Planetary Computer" To Support Global Sustainability. Available online: https://earth.org/planetary-computer/ (accessed on 22 December 2021).
134. Microsoft Planetary Computer. Available online: https://planetarycomputer.microsoft.com/ (accessed on 15 December 2021).
135. Yang, L.; Cervone, G. Analysis of Remote Sensing Imagery for Disaster Assessment Using Deep Learning: A Case Study of Flooding Event. *Soft Comput.* **2019**, *23*, 13393–13408. [CrossRef]
136. Wilson, E.H.; Sader, S.A. Detection of Forest Harvest Type Using Multiple Dates of Landsat TM Imagery. *Remote Sens. Environ.* **2002**, *80*, 385–396. [CrossRef]
137. Skakun, R.S.; Wulder, M.A.; Franklin, S.E. Sensitivity of the Thematic Mapper Enhanced Wetness Difference Index to Detect Mountain Pine Beetle Red-Attack Damage. *Remote Sens. Environ.* **2003**, *86*, 433–443. [CrossRef]
138. Measuring Vegetation (NDVI & EVI). Available online: https://earthobservatory.nasa.gov/features/MeasuringVegetation (accessed on 20 December 2021).
139. Gao, B.-C. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
140. McFEETERS, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
141. Xu, H. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]
142. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 8026–8037.
143. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
144. Pal, M. Random Forest Classifier for Remote Sensing Classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]
145. Wu, X.; Sahoo, D.; Hoi, S.C.H. Recent Advances in Deep Learning for Object Detection. *Neurocomputing* **2020**, *396*, 39–64. [CrossRef]
146. Ben-Hur, A.; Horn, D.; Siegelmann, H.T.; Vapnik, V. Support Vector Clustering. *J. Mach. Learn. Res.* **2001**, *2*, 125–137. [CrossRef]
147. Pratt, L.Y. Discriminability-Based Transfer between Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Denver, CO, USA, 29 November–2 December 1993; pp. 204–211.
148. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
149. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A Survey of Transfer Learning. *J. Big Data* **2016**, *3*, 9. [CrossRef]
150. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; pp. 270–279.
151. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2019**, *109*, 43–76. [CrossRef]
152. Li, C.; Zhang, S.; Qin, Y.; Estupinan, E. A Systematic Review of Deep Transfer Learning for Machinery Fault Diagnosis. *Neurocomputing* **2020**, *407*, 121–135. [CrossRef]
153. Bar, Y.; Diamant, I.; Wolf, L.; Lieberman, S.; Konen, E.; Greenspan, H. Chest Pathology Detection Using Deep Learning with Non-Medical Training. In Proceedings of the 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), Brooklyn, NY, USA, 16–19 April 2015; pp. 294–297.
154. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.
155. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

156. Maxwell, A.E.; Warner, T.A. Thematic Classification Accuracy Assessment with Inherently Uncertain Boundaries: An Argument for Center-Weighted Accuracy Assessment Metrics. *Remote Sens.* **2020**, *12*, 1905. [CrossRef]
157. Sokolova, M.; Lapalme, G. A Systematic Analysis of Performance Measures for Classification Tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
158. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation. *Appl. Soft Comput.* **2018**, *70*, 41–65. [CrossRef]
159. Pontius, R.G.; Millones, M. Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]
160. Yang, X.; Zhao, S.; Qin, X.; Zhao, N.; Liang, L. Mapping of Urban Surface Water Bodies from Sentinel-2 MSI Imagery at 10 M Resolution via NDWI-Based Image Sharpening. *Remote Sens.* **2017**, *9*, 596. [CrossRef]
161. GSP216Online Accuracy Metrics. Available online: http://gsp.humboldt.edu/olm_2019/courses/GSP_216_Online/lesson6-2/metrics.html (accessed on 23 December 2021).
162. Stehman, S.V. Selecting and Interpreting Measures of Thematic Classification Accuracy. *Remote Sens. Environ.* **1997**, *62*, 77–89. [CrossRef]
163. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 658–666.
164. Padilla, R.; Netto, S.L.; da Silva, E.A.B. A Survey on Performance Metrics for Object-Detection Algorithms. In Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020; pp. 237–242.
165. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [CrossRef]
166. McHugh, M.L. Interrater Reliability: The Kappa Statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]
167. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]
168. Stehman, S.V.; Foody, G.M. Key Issues in Rigorous Accuracy Assessment of Land Cover Products. *Remote Sens. Environ.* **2019**, *231*, 111199. [CrossRef]
169. Stehman, S.V.; Wickham, J.D. Pixels, Blocks of Pixels, and Polygons: Choosing a Spatial Unit for Thematic Accuracy Assessment. *Remote Sens. Environ.* **2011**, *115*, 3044–3055. [CrossRef]
170. Foody, G.M. Status of Land Cover Classification Accuracy Assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]
171. Congalton, R.G. Accuracy Assessment and Validation of Remotely Sensed and Other Spatial Information. *Int. J. Wildland Fire* **2001**, *10*, 321–328. [CrossRef]
172. Stehman, S.V. Sampling Designs for Accuracy Assessment of Land Cover. *Int. J. Remote Sens.* **2009**, *30*, 5243–5272. [CrossRef]
173. Strutz, T. Data Fitting and Uncertainty. Available online: https://link.springer.com/book/9783658114558 (accessed on 31 December 2021).
174. Willmott, C.J.; Matsuura, K. Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
175. Rosenblad, A. The Concise Encyclopedia of Statistics. *J. Appl. Stat.* **2011**, *38*, 867–868. [CrossRef]
176. Schütze, H.; Manning, C.D.; Raghavan, P. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008; Volume 39.
177. Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In Proceedings of the Advances in Information Retrieval, Santiago de Compostela, Spain, 21–23 March 2005; pp. 345–359.

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
www.mdpi.com

Sensors Editorial Office
E-mail: sensors@mdpi.com
www.mdpi.com/journal/sensors

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-0365-9595-5