remote sensing

MDPI

Article

Prediction of Groundwater Level Based on the Integration of Electromagnetic Induction, Satellite Data, and Artificial Intelligent

Fei Wang [1], Lili Han [1], Lulu Liu [1], Yang Wei [1,*] and Xian Guo [2]

[1] School of Architecture and Civil Engineering, Chengdu University, Chengdu 610106, China; wangfei@cdu.edu.cn (F.W.); halili@cdu.edu.cn (L.H.); liululu@cdu.edu.cn (L.L.)
[2] School of Earth Resources, China University of Geosciences, Wuhan 430074, China; 1202220025@cug.edu.cn
* Correspondence: weiyang@cdu.edu.cn

Abstract: Groundwater level (GWL) in dry areas is an important parameter for understanding groundwater resources and environmental sustainability. Remote sensing data combined with machine learning algorithms have become one of the important tools for groundwater level modeling. However, the effectiveness of the above-based model in the plains of the arid zone remains underexplored. Fortunately, soil salinity and soil moisture may provide an optimized solution for GWL prediction based on the application of apparent conductivity (ECa, mS/m) using electromagnetic induction (EMI). This has not been attempted in previous studies in oases in arid regions. The study proposed two strategies to predict GWL, included an ECa-based GWL model and a remote sensing-based GWL model (RS_GWL), and then compared and explored their performances and cooperation possibilities. To this end, this study first constructed the ECa prediction model and the RS_GWL with ensemble machine learning algorithms using environmental variables and field observations (474 ECa reads and 436 groundwater level observations from a mountain–oasis–desert system, respectively). Subsequently, a strategy to improve the prediction accuracy of GWL was proposed by comparing the correlation between GWL observations and the two models. The results showed that the RS_GWL prediction model explains 30% and 90% of the spatial variability in the two value domain intervals, GWL < 10 m and GWL > 10 m, respectively. The R^2 of the modeling and the validation of the ECa was 79% and 73%, respectively. Careful analysis of the scatter plots between predicted ECa and GWL revealed that when ECa varies between 0–600 mS/m, 600–800 mS/m, 800–1100 mS/m, and >1100 mS/m, the fluctuation ranges of the corresponding GWL correspond to 0–31 m, 0–15 m, 0–10 m, and 0–5 m. Finally, combining the spatial variability of ECa and RS_GWL spatial distribution map, the following optimization strategies were finally established: GWL < 5 m (in natural land with ECa > 1100 mS/m), GWL < 5 m (occupied by farmland from RS_GWL) and GWL > 10 m (from RS_GWL), and 3 < GWL < 10 m (speculated). In conclusion, this study has demonstrated that the integration of EMI technology has significantly improved the precision of forecasting shallow GWL in oasis plain regions, outperforming the outcomes achieved by the use of remote sensing data alone.

Keywords: groundwater levels; remote sensing; electromagnetic induction; machine learning; Tarim Basin

check for updates

Academic Editor: Hatim Sharif

Received: 7 December 2024
Revised: 6 January 2025
Accepted: 7 January 2025
Published: 8 January 2025

Citation: Wang, F.; Han, L.; Liu, L.; Wei, Y.; Guo, X. Prediction of Groundwater Level Based on the Integration of Electromagnetic Induction, Satellite Data, and Artificial Intelligent. *Remote Sens.* **2025**, *17*, 210. https://doi.org/10.3390/rs17020210

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Groundwater scarcity is a fundamental challenge for societies in arid and semi-arid regions worldwide [1]. The availability of groundwater resources has become a significant factor influencing population growth and distribution, industrial structure, socioeconomic development, and ecological security [2]. Groundwater serves as a stable water source and could provide a continuous supply to plants and soil organisms, provided that the root systems or soil capillary fringes could access it in close proximity [3,4]. The depth to groundwater is a crucial parameter for understanding groundwater resources and maintaining the sustainable development of water resources and the environment [5]. Groundwater at various depths substantially influences vegetation density, plant growth [6], and soil properties [7]. The spatial distribution of the groundwater level (GWL) directly reflects the availability of groundwater resources, and it has been studied extensively by hydrogeological, agricultural, and social science research communities at both local and regional scales [8].

Shallow groundwater is essential for the survival of dominant vegetation species in arid regions, such as the Tarim Basin in northwestern China, and provides sustenance for oasis core areas [9–11]. In arid ecosystems, vegetation could be subjected to water stress when an aquifer is too deep, and, conversely, excessively shallow depths could lead to soil salinity that curtails plant growth [12]. The influence of groundwater depth on the distribution of dominant vegetation species in the desert regions of the Tarim Basin was investigated and specific growth thresholds were delineated for *Populus euphratica* and *Tamarix ramosissima* [10]. Medium- and large-sized individuals of these species failed to thrive at a GWL of 6.9 m with necrotic branches emerging at 7.8 m. These field observations highlight the pivotal role that GWL plays in survival strategies of vegetation species. Accordingly, we suggest that a detailed understanding of spatial variations in GWL could provide significant guidance for ecological conservation and restoration efforts in downstream oases of arid regions.

Remote sensing has been instrumental in modeling GWL, demonstrated by a substantial body of research [5,13–19]. The ability to accurately predict key surface parameters from remote sensing data is fundamental to its application in deducing GWL because of the close relationship between these parameters and subsurface hydrology. The capacity to reveal parameters, including soil moisture [20], vegetation cover [21], groundwater storage [22–24], topography, and land surface temperature [25] at a large spatial scale, leverages the inherent strengths of remote sensing data. Concurrently, machine learning has become an essential alternative to physical models for cost-effective, large-scale simulations of spatial variations in GWL [1,26]. This strategy is predicated on the ability of the system to learn from data, recognize patterns, and make decisions with minimal human intervention [27]. Sahour et al. employed methods such as Extreme Gradient Boosting (XGB), Support Vector Machine (SVM), and Logistic Regression (LR) to establish relationships between shallow groundwater and environmental variables derived from remote sensing data [1]. Subsequently, such information was applied to map shallow groundwater distribution in the Western Desert of Egypt. However, the distribution of shallow groundwater in plain oasis areas (<10 m) presents a challenge because the GWL could be impacted by various land use practices. This phenomenon complicates the precise prediction of GWL in these areas using models informed by remote sensing datasets and machine learning methods (RS_GWL) [26]. Furthermore, owing to the inaccessibility of GWL data, studies employing this strategy in oasis regions remain scarce. Therefore, enhancing the precision of GWL simulations in shallow groundwater of plain oases is the central focus of our study.

The achievements based on the application of electromagnetic induction (EMI) on soil salinity and soil moisture may provide new insights to respond to the challenge. Huth and

Poulton [28] successfully characterized variations in total soil moisture within the upper 0.9 m of soil profiles using EMI near Walla, Queensland, Australia. Misra and Padhi [29] employed a Geonics EM38 ground conductivity meter to measure the apparent electrical conductivity (ECa) and evaluate the sensitivity of the EMI method to soil moisture fluctuations in irrigated cotton fields. Their results showed a nonlinear increase in ECa values across seasons with the accumulation of soil moisture at specific depths. The fitted model exhibited a high degree of statistical confidence ($p \leq 0.001$) and a substantial coefficient of determination (R^2, 0~1). Moghadas et al. [30] found that ECa observations could accurately track the spatiotemporal distribution and alterations in soil moisture content stemming from absorption and evaporation in the root profiles of corn. Jiang et al. [31] used the deep-rooted grass *Achnatherum splendens* as an indicator species and applied the EMI detection method to assess the relationship between the spatial distribution of *Achnatherum splendens* patches and the dynamics of soil moisture and salinity in the Qinghai Lake Basin, located in the northwest of the Tibetan Plateau. Turkeltaub et al. [32] quantified ECa at six effective depths under four predominant land-cover types (shrubland, grassland, natural fallow land, and cropland) in the northern Loess Plateau of China. These outcomes underscore the potential utility of the EMI method for investigating the distribution of soil moisture in the unsaturated zone of our study area, which could be instrumental in identifying unsustainable land-cover patterns. Extensive research has revealed a close correlation between ECa measurements and soil salinity profiles [33–35]. Based on this, a portion of the research showed the distribution characteristics of soil moisture and soil salinity were indicative of variations in groundwater depth [36–38].

Earlier studies have attempted EMI-based techniques to detect GWL. Sherlock and McDonnell [39] discovered that data form Geonics EM31 induction meter accounted for more than 80% of the variability in groundwater depth on slopes, whereas EM38 data explained more than 70% of the changes in soil moisture, determined by gravimetric methods. Hedley et al. [40] integrated EM38 (Electromagnetic induction devices) surveys with a wireless soil moisture sensor network and rainfall data for dynamic prediction of soil moisture and GWL in agricultural settings (<1.5 m). Accordingly, EMI measurements offer a promising avenue for detecting spatial variations in GWL. However, this approach has not been investigated extensively in arid oasis regions owing to the paucity of supporting data.

In the current study, we used remote sensing, machine learning, EMI, and observed GWL data to predict GWL. Initially, the performance of groundwater models constructed from remote sensing data was assessed for predicting different GWL. Subsequently, the spatial distribution of ECa within the research area was delineated and the relationship between ECa and GWL scrutinized. Finally, the advantages and limitations of both ECa and remote sensing-GWL (RS_GWL) models for identifying shallow groundwater were analyzed and the feasibility of their potential integration explored. The study was carried out in the Tarim River Basin in Xinjiang, which is one of the best-preserved mountain-oasis-desert systems (MODS), providing a valuable test site for studying groundwater level prediction.

2. Study Area

The Tarim Basin is located in southern Xinjiang, China (Figure 1). The region has an arid climate with scarce rainfall and intense evaporation, exhibiting distinct continental climatic features. The annual precipitation is generally below 50 mm, and evaporation reaches 2300 to 3000 mm [41]. Sunlight is abundant, with an annual sunshine duration range of 2800–3100 h. The average annual temperature is 10–11 °C, with accumulated temperature above 10 °C reaching 4000–4350 °C [42]. The primary soil types in the area include solonchak, poplar, saline, oasis meadow, eolian sand, and marsh soil.

The topography of the region slopes from high in the west to low in the east, with an altitude range of 779–1031 m [43], and the lowest point being at Taitema Lake. The Tarim River is the largest inland river in China and is a collective term for 114 rivers from nine major water systems, including the Aksu, Hotan Yarkand, Keriya, Kuya, Di, Kashgar, Peacock, Korda, and Weigan Rivers [44]. Water resources primarily depend on glacial meltwater. In the study area, the flow directions of groundwater and surface water are consistently west to east. The hydraulic gradient is generally 0.0002, indicating an exceptionally slow flow [45]. The vegetation type is arboreal, shrub, and grassland. The arboreal species primarily comprise *Populus euphratica* as the dominant group [26], with shrubs including *Tamarix chinensis*, *Haloxylon ammodendron*, and *Lycium ruthenicum*, including *Phragmites communis*, *Karelinia caspica*, *Glossogyne tenuifolia*, and *Alhagi sparsifolia*.

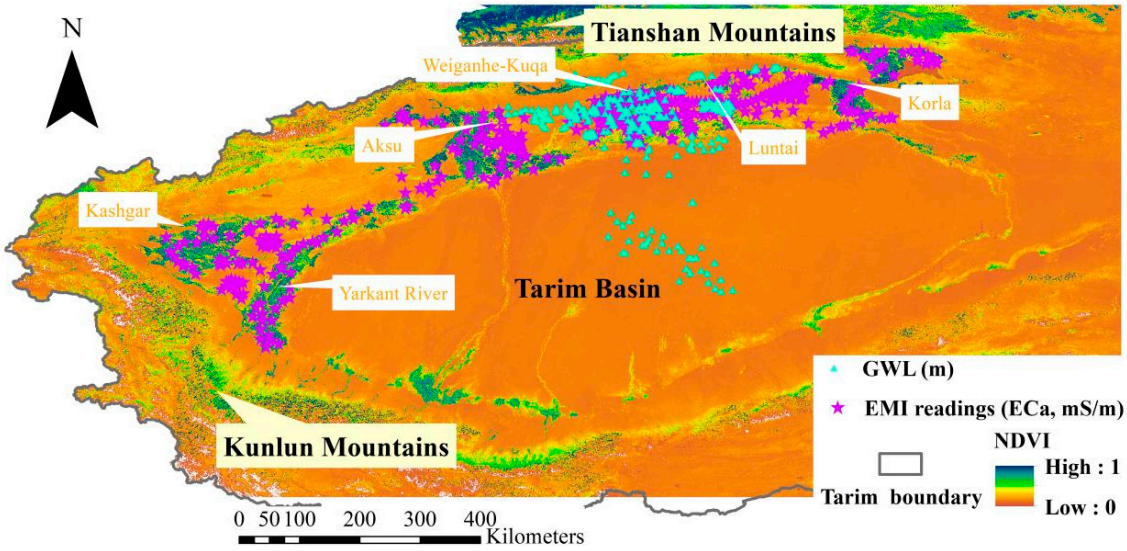

Figure 1. Distribution map of electromagnetic induction (EMI) measurement sites and groundwater level collection points overlaid on Normalized Difference Vegetation Index (NDVI) in the Tarim Basin. The boundary of the Tarim River Basin (2019) is sourced from the National Tibetan Plateau Data Centre.

3. Data and Method

3.1. Field-Measured Samples

The groundwater level observational data were obtained through environmental monitoring and assessment projects conducted prior to the development of an oil field. The samples were collected from areas such as piedmont alluvial fans, agricultural irrigation zones, ecotone (between oases and deserts), and desert regions. Groundwater was collected through mechanical drilling and from existing observation wells (Figure 2). The data collection period was from April to June 2018, and a total of 436 groundwater level values were collected during this period. The layout of these sites was designed taking into account land use and geological conditions.

The raw data of electrical conductivity (ECa) were collected utilizing EM38-MK2 form Geonics Limited, Mississauga, ON, Canada, during the summers of 2018 and 2019, and then a total of 474 samples were obtained after data processing. The standard configuration of the EM38-MK2 includes two receiver coils positioned 1 m and 0.5 m away from the transmitter, respectively. When aligned with the vertical dipole orientation, they provide effective depth range data for 1.5 m and 0.75 m, respectively, and, when aligned with the horizontal dipole orientation, they offer effective depth ranges of 0.75 m and 0.375 m, respectively [46].



Figure 2. Photos showing the field activities during the groundwater sampling and ECa observations.

The vertical mode of EM38-MK2, with a detection depth of 1.50 m, was used for sampling. Data collectors carrying EM 38-MK2 could not carry metal objects on their bodies. During the process, it was ensured that the device was maintained at a height of 5 cm above the ground. The data acquisition task was completed using an S-shaped path in the sample area (90 × 90 m).

Prior to EMI measurements, the surface temperature (T) had to be measured to correct its influence on the true ECa values at 25 °C [46] (Figure 3). After temperature correction, the raw data were screened to remove zero and negative values, followed by averaging of the data collected along the nonlinear paths to represent the apparent electrical conductivity (ECa, mS/m) of the 90 × 90 m grid cell.

$$\mathrm{ECa25\ °C} = \mathrm{ECa} \times (0.4779 + 1.3801e^{(-T/25.654)}) \quad (1)$$

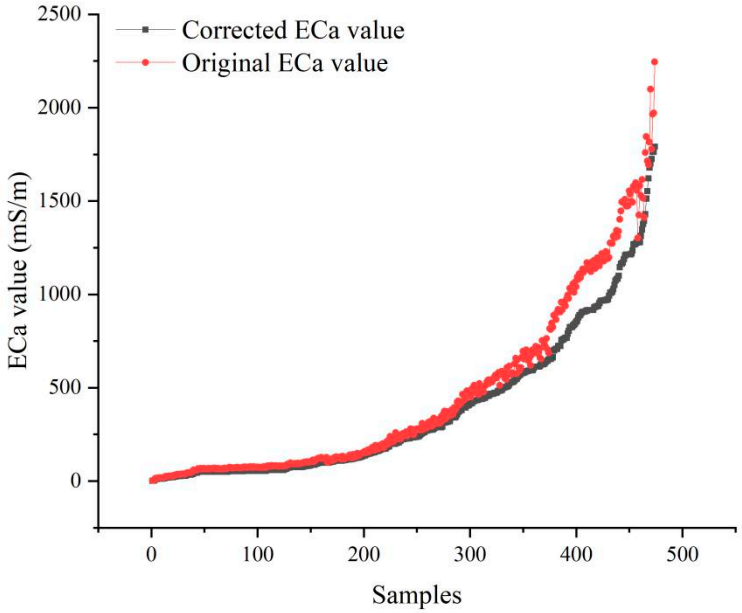

Figure 3. ECa observations before and after correction based on surface temperature.

3.2. Environmental Variables

The optical data used in this study were derived from Sentinel-2, a satellite mission that provides high-resolution multispectral imagery. The data encompassed 13 spectral bands that underwent atmospheric correction and cloud removal to ensure data quality and usability.

Vegetation indices serve as crucial proxies for soil moisture levels and the replenishment of groundwater resources [1,10]. Therefore, NDVI (normalized difference vegetation

index) annual median values were employed to avoid missing data and unusual fluctuations. The data were collected for the years 2016–2020. The sensor data were obtained from the Sentinel-2 series satellites.

The temperature could indirectly reflect the soil moisture conditions [47]. Our study leveraged multiple temperature products from the OpenLandMap series available on the Google Earth Engine (GEE) platform. The products include the Land Surface Temperature daytime monthly mean value (2000–2017), long-term MODIS LST daytime and nighttime temperature standard deviation, and long-term MODIS LST daytime and nighttime differences at 1 km, based on the 2000–2017 time series. Each product comprises 12 bands, representing 12 months of the year. In total, 36 environmental variables from the three datasets were used.

Soil bulk density (BD) is an important index for evaluating soil and water conservation ability [48]. We considered this indicator useful to aid in the determination of GWL and, therefore, include it in the environmental dataset. We obtained BD from the OpenLandMap series of Google Earth Engine (GEE) Platform. BD for this dataset has a spatial resolution of 250 m and contains 6 depths (0, 10, 30, 60, 100, and 200 cm). The OpenLandMap series also includes products related to soil moisture, which are 33 kPa (Field Capacity) and 1500 kPa (Permanent Wilting Point). They were also included in the environmental datasets.

The climatic data were sourced from the WorldClim database, which include 19 bioclimatic variables (BIO01...BIO19) [49]. Solar radiation, saturated water pressure, and wind speed from the platform were also considered. Each of these three datasets comprised 12 bands corresponding to the 12 months of the year.

Topography is a significant variable that affects groundwater, and it has been used extensively in various surface property predictions [50]. A digital elevation model was employed to calculate the Slope, Topographic Position Index (TPI), and Multi-Resolution Index of Valley Bottom Flatness (MRVBF). These calculations were performed using the System for Automated Geoscientific Analyses software (SAGA, version 7.9.0, Hamburg, Germany).

Soil type and land use practices were also considered in this study. The soil type data were obtained from the Harmonized World Soil Database (HWSD v2.0). This database is a comprehensive global soil map that offers detailed insights into soil properties, including morphology, chemistry, and physical characteristics, with a resolution of 1 km. The database is a valuable tool for research on the effects of agriculture, food security, and climate change. Differences in hydrothermal conditions under various land use patterns have significant effects on water and salt transport [51]. Therefore, we selected global 30 m surface land use/cover products [52], which are available at http://data.ess.tsinghua.edu.cn/. This product comprises seven land use types, namely farm- and grassland, shrubbery, desert, unused and urban land, and water bodies.

3.3. Machine Learning

This study combines the entire individual algorithm to build models. Linear regression is simple, efficient, and interpretive. The decision tree is easy to understand, it can handle multiple data and automatically select features. The extreme random tree is strong in randomness, fast in training, and robust. XGBoost (v2.1) has high performance, prevents overfitting, and supports parallel and missing value processing. LightGBM (v4.0) is fast and memory-saving, can handle large-scale data, and is accurate.; Random forest is resistant to overfitting, can handle high-dimensional data, and is robust. Their integrated algorithms can improve performance, enhance stability, and optimize generalization ability, making up for the shortcomings of single algorithms and adapting to a wider range of scenarios.

Linear regression is a straightforward algorithm prevalent in the machine learning domain [53]. This statistical approach is employed for predictive analytics. The linear

regression algorithm, as the name implies, delineates a linear relationship between a dependent variable (y) and one or more independent variables (x).

Decision Trees constitute a category of supervised machine learning algorithms [54] that were utilized in this study to classify or regress data through a series of true or false responses to specific questions. When rendered visually, the outcomes take the shape of a tree comprising various types of nodes, namely root, internal, and leaf. The root node marks the inception of the Decision Tree, which, subsequently, diverges into internal and leaf nodes. The leaf nodes represent the ultimate classification bins or actual values. Decision trees are characterized by their simplicity and interpretability.

The extra-tree algorithm [55], similar to the random forest algorithm, generates a multitude of Decision Trees. However, sampling for each tree is executed randomly and without replacement, yielding a distinct dataset for each tree with unique samples. A random subset of the features from the entire feature set is selected for each tree. A distinctive and pivotal feature of the extra-tree algorithm is the stochastic selection of the split value for a feature. Rather than computing a locally optimal value for data bifurcation using metrics, such as Gini impurities or entropy, the algorithm chooses a random split value. This approach fosters diversity and decorrelation among the trees.

The XGBoost algorithm, introduced by Chen and Guestrin [56], is essentially a gradient-boosting Decision Tree framework using both greedy and approximate algorithms to determine the tree model. Specifically, in the tree-construction phase of the XGBoost algorithm, the difference between the actual and predicted values (residuals) was initially calculated. During this process, the actual values are known, whereas the predicted values are optimized using the XGBoost algorithm through iterations over the samples. The similarity (Similarity Scores) between each root node and its leaf nodes is measured based on the residuals at different splitting points, and the gain is defined as the difference between the sum of the Similarity Scores of the leaf nodes and the similarity score of the root node. Finally, the XGBoost algorithm establishes leaf nodes at the splitting point with the highest gain, and this process is iterated to grow the model tree. Because of its high computational efficiency, strong flexibility, precise prediction, and ease of parallel computation, the XGBoost algorithm is being applied widely for geospatial information mining and image pattern recognition.

LightGBM, introduced by a Microsoft team [57], is an efficient gradient boosting Decision Tree (GBDT) algorithm that represents an improvement over traditional GBDTs and is also an ensemble learning algorithm based on the boosting framework. Conventional boosting algorithms often have limitations, particularly in terms of their scalability and operational speed. Compared with other boosting algorithms, such as XGBoost, LightGBM effectively addresses limitations, offering a significant reduction in training time and memory usage without compromising predictive accuracy. LightGBM primarily enhances the training speed through two key methodologies. The first is Gradient-based One-Side Sampling (GOSS), which selects samples with larger gradients, thereby increasing their contribution to the computational information gain. The second is Exclusive Feature Bundling (EFB), which merges certain data features to reduce dimensionality. These innovations facilitate LightGBM to effortlessly manage large-scale data and complex models, rendering it the preferred choice for high-performance machine learning tasks in various fields.

Random Forest (RF), introduced by Breiman [58], is an ensemble learning method based on bagging that uses Decision Trees as base learners within the ensemble. The algorithm functions by random sampling with replacements from the training set, constructs a Decision Tree for each set of samples, and then aggregates the predictions from all trees, typically by averaging, to generate the final outcome. The CART (Classification and Regression Trees) algorithm is employed to generate regression Decision Trees. CART

is a binary tree-based method that partitions a dataset based on features and their corresponding subranges, calculating the mean of the y-values for each subset. Subsequently, it computes the mean squared error (MSE) for both subsets and identifies the minimum MSE as the optimal split point, with the corresponding feature as the optimal splitting feature. This process is repeated, traversing through all data until the MSE is reduced below a certain threshold. The Random Forest algorithm has been widely and effectively applied in the field of remote sensing information extraction, offering robust performance and high accuracy because of its ensemble nature and the diversity of its constituent Decision Trees.

An ensemble represents a collective of models whose predictive outcomes are consolidated using techniques such as weighted averaging or voting. Prevalent strategies for ensemble construction include bagging, represented by the Random Forest algorithm, and boosting, represented by XGBoost. These methods amalgamate the outputs from multiple instances of the same algorithm, each trained on different subsets of the data or in a sequential manner to correct the errors of prior iterations. In more recent advancements, particularly those discussed in "Ensemble Selection from Libraries of Models" by Caruana et al. [59], a shift occurs toward integrating diverse models to capture a wider array of patterns in the data. The ensemble creation process typically involves the following methodological steps.

(1) Incorporate the model from the library into the ensemble that most significantly enhances its performance, as measured by a predefined error metric on a validation dataset.
(2) Iterate this process for a predetermined number of cycles, or continue until every model in the library has been considered for inclusion.
(3) From the formed nested set of ensembles, select and return the one that demonstrates optimal performance on the validation set.

In our study, the candidate subroutines for constructing the ensemble algorithm comprised a range of models, including linear regression, Decision Trees, Extra trees, XGBoost, Random Forest, and LightGBM.

3.4. Flowchart

For details of the specific process, please refer to Figure 4.

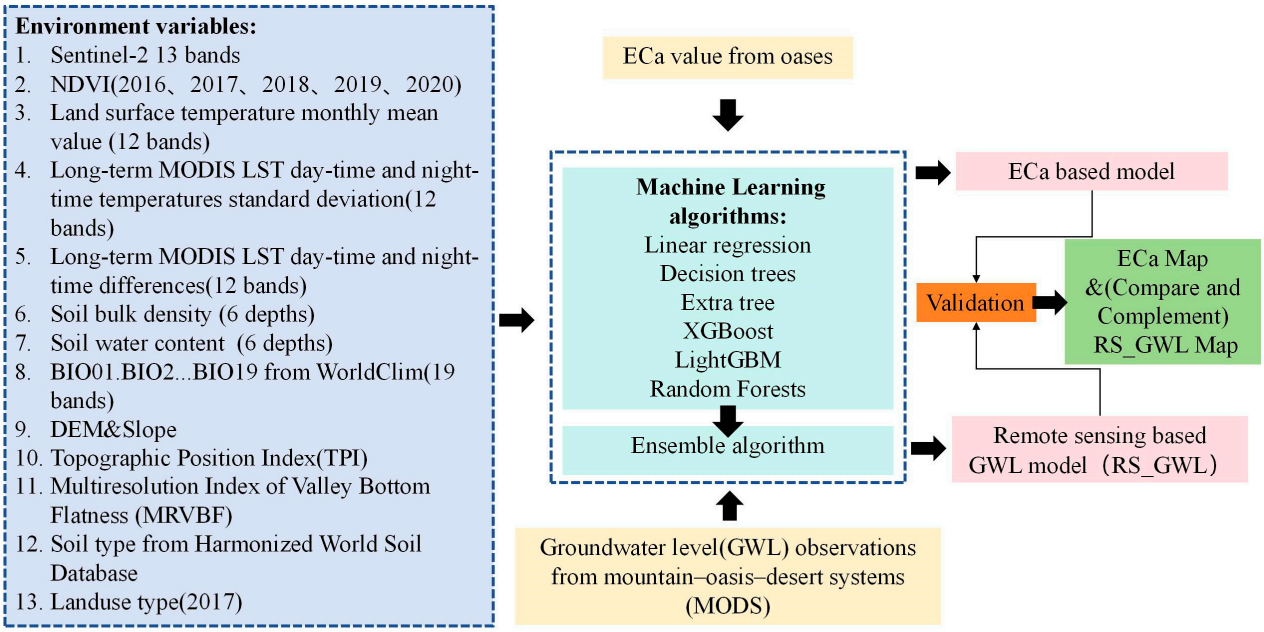

Figure 4. Flowchart for the development of an integrated strategy in GWL prediction for the study area.

(1) The spatial resolution of all environmental variables is resampled to 90 m.
(2) Optimize the algorithm by adjusting the parameters and select the modeling algorithm by comparing predicted errors of five datasets (<5 m, 5~10 m, >10 m, oasis, desert).
(3) Build ECa and RS_GWL models based on the optimal algorithms and subsequently draw ECa map and RS_GWL spatial distribution maps.
(4) Attempt to establish the linkage between the GWL observations and the predicted ECa and RS_GWL and analyze the possible relationships between them.
(5) Develop a relatively optimal strategy for GWL prediction.

3.5. Modeling and Validation

This study employed a 7:3 sampling ratio to validate the predictive outcomes of ECa and RS_GWL. The ECa observations were concentrated predominantly in the oasis areas within the MODS, whereas the samples for RS_GWL encompassed the entire MODS. To validate the ECa model, we allocated 334 samples (70% of the total) for model construction, and the remaining 140 samples (30%) for validation. The predictive accuracy was assessed using *p* value, R-squared (R^2), and Root Mean Square Error (RMSE) as metrics of evaluation. The model calculation was completed in RStudio (v7.2), and the p-value calculation was performed in Origin (v8.5).

4. Result

4.1. Comparison of Machine Learning Algorithms in GWL Prediction

The optimal modeling algorithm was selected based on the RMSE criterion (Figure 5). A comprehensive review indicated that the ensemble algorithm consistently exhibited the lowest RMSE values across all the datasets. In the two datasets in which the GWL was less than 10 m, the RMSE of all algorithms was approximately 1 m. In the dataset in which the GWL exceeded 10 m, the RMSE error of the ensemble algorithm was 10.15 m, whereas the highest RMSE was ascribed to the Random Forest algorithm at 25.10 m. In oasis areas where groundwater is used most extensively, the RMSE error of the ensemble algorithm was less than 1 m, that of the Random Forest was 1.45 m, and the highest RMSE was observed in the linear model. Boxplot results showed that the RMSE of the ensemble algorithm had the narrowest variation range, followed closely by that of XGBoost. In contrast, the RMSE variability for Random Forest, LightGBM, and Decision Tree was broader than those for the ensemble algorithm and XGBoost.

The study evaluated the explanatory power of relatively excellent machine learning methods. As shown in Figure 6, the highest explanatory power is the ensemble algorithm (R^2 = 0.93), followed by XGboost (R^2 = 0.897) and LightGBM (R^2 = 0.891). This precision implies that the ensemble model can account for 93% of the spatial variability of GWL. Based on this preliminary result, we further dissected the prediction performance of the ensemble algorithm in the GWL < 10 m range. The results show that the predicted values explain approximately 30% of the spatial variability in GWL. And in the interval of GWL > 10 m, the validation accuracy of the constructed model based on this ensemble algorithm is R^2 = 0.905.

An ensemble algorithm was employed using all available samples (436) to map the spatial distribution of GWL within the MODS, as shown in Figure 7. Mountainous regions and piedmont alluvial fan areas exhibited the deepest GWL, ranging from 80 to 139 m. In contrast, agricultural irrigation zones within the oases had the shallowest GWL of less than 4 m. The transitional zones between oases and deserts have GWL between 3 and 15 m, whereas desert areas contain GWL ranging from 7 to 20 m. These spatial distribution characteristics are consistent with the topography and land use type.

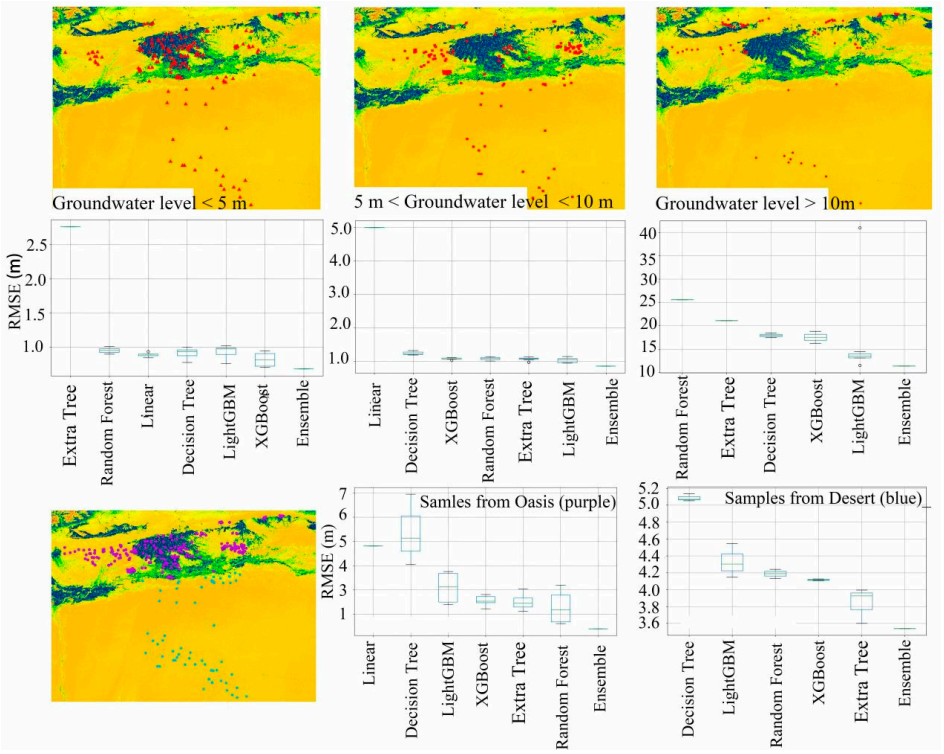

Figure 5. Comparative performance of seven algorithms in predicting GWL across five datasets (based on RMSE as the criterion). Sites in red and purple are sampling points from different zones. In the figure, yellow and orange regions indicate deserts and Gobi areas, whereas dark blue and light green signify agricultural lands and desert vegetation.

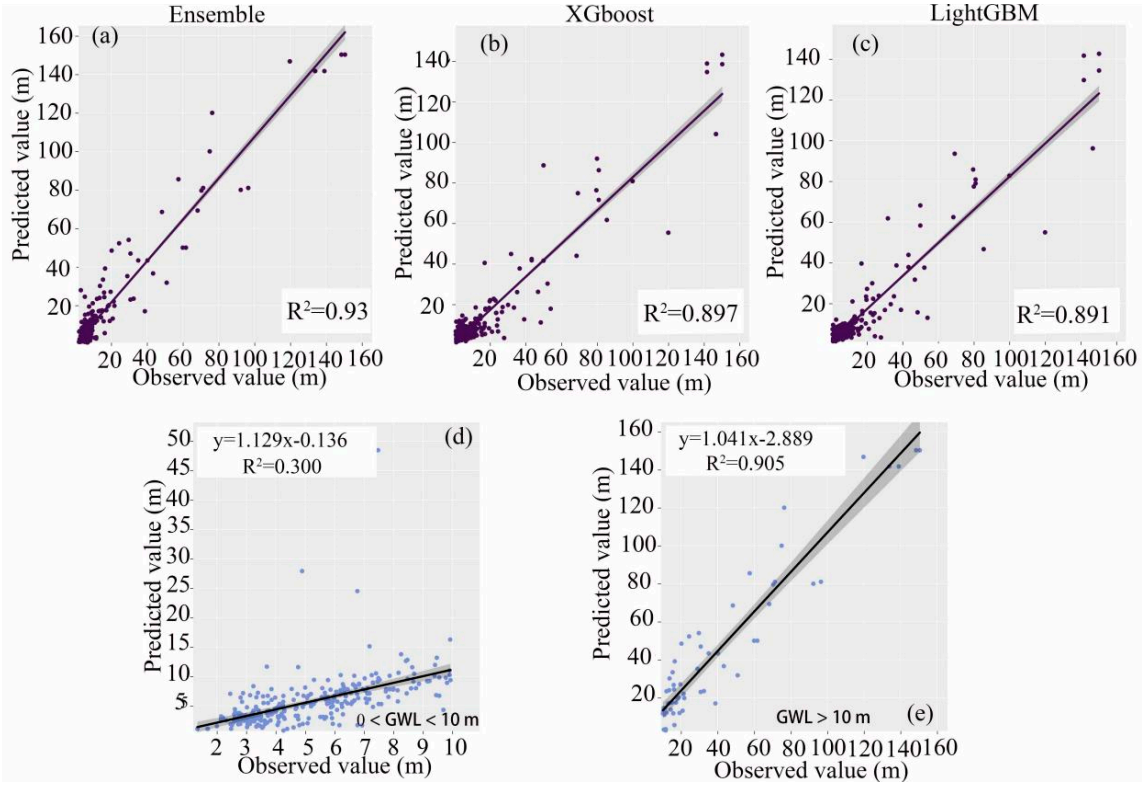

Figure 6. Validation of the GWL prediction accuracy for models constructed by the ensemble algorithm (**a**), XGboost (**b**), and LightGBM (**c**); (**d**,**e**) show the performance of the model built based on ensemble algorithm at 0 < GWL < 10 m and GWL > 10 m, respectively. The solid line and shading around the solid line represent the fitted trend and 95% confidence interval, respectively.

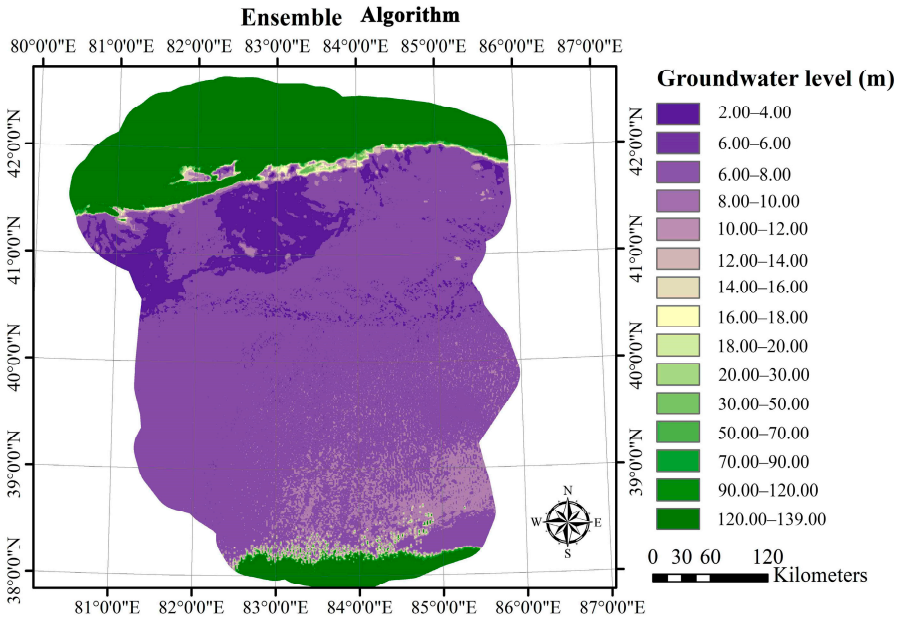

Figure 7. Spatial distribution of predicted groundwater levels in the study area (sample coverage area for GWL) based on ensemble algorithm. The figure showed that GWL in mountainous areas was mainly concentrated above 90 m (green), while GWL in oasis–desert transition zone (purple) was mainly distributed below 10 m. The GWL of the irrigated area inside the oasis was below 4 m, which was consistent with the records from the observation wells.

4.2. Electromagnetic Induction Modeling and Validation

Based on the performance of the ensemble algorithm in predicting GWL, we adopted this method to establish an ECa model. The samples were allocated to the modeling and validation datasets at a ratio of 7:3. The results indicated (Figure 8) an R^2 value of 0.78 for the ECa model, with a validation R^2 value of 0.73 and *p*-values (significance testing) less than 0.001. Subsequently, the predicted ECa values and their spatial distributions were used to analyze the correlation between the GWL and ECa.

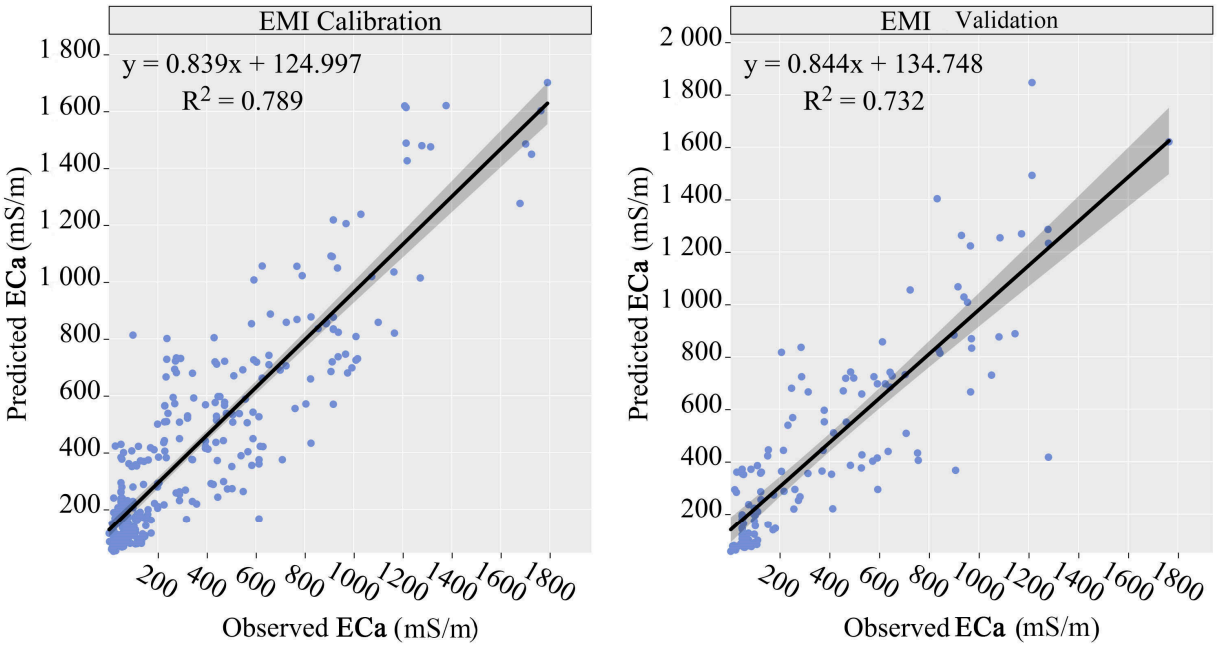

Figure 8. Assessment of predictive model accuracy for ECa calibration samples ((**left**), 70% of samples) and validation samples (**right**), 30% of samples) using the ensemble algorithm.

4.3. Spatial Distribution of ECa in the Tarim Basin

This study presents the first spatial distribution map of ECa values in the Tarim Basin (Figure 9). Generally, the ECa values were relatively low in over 90% of the region, which is largely ascribed to deserts with low soil moisture and irrigated agricultural lands with low soil salinity [33,60]. In contrast, regions with high ECa values were observed along the periphery of several oases adjacent to the Tarim River, specifically the Kashgar, Aksu, Kuqa–Luntai, and Ruoqiang oases. A notably large area with high ECa values is situated east of the Aksu Oasis, which encompasses the first national wetland park in Xinjiang. Additionally, contiguous zones with high ECa values exist in certain parts of the Kuqa–Luntai Oasis. Interestingly, the Ruoqiang Oasis also exhibited higher ECa-predicted values in its vicinity, which was in contrast to the dryness of the lake nearly 20 years ago [61]. This increase can be correlated with the ongoing water diversion from the BosiTeng Lake to the lower reaches of the Tarim River in recent years [62]. The expansion of Tatma Lake (around Ruoqiang Oasis) within this region has led to an increase in the GWL around the lake. This observational outcome further substantiates the authenticity of ECa prediction.

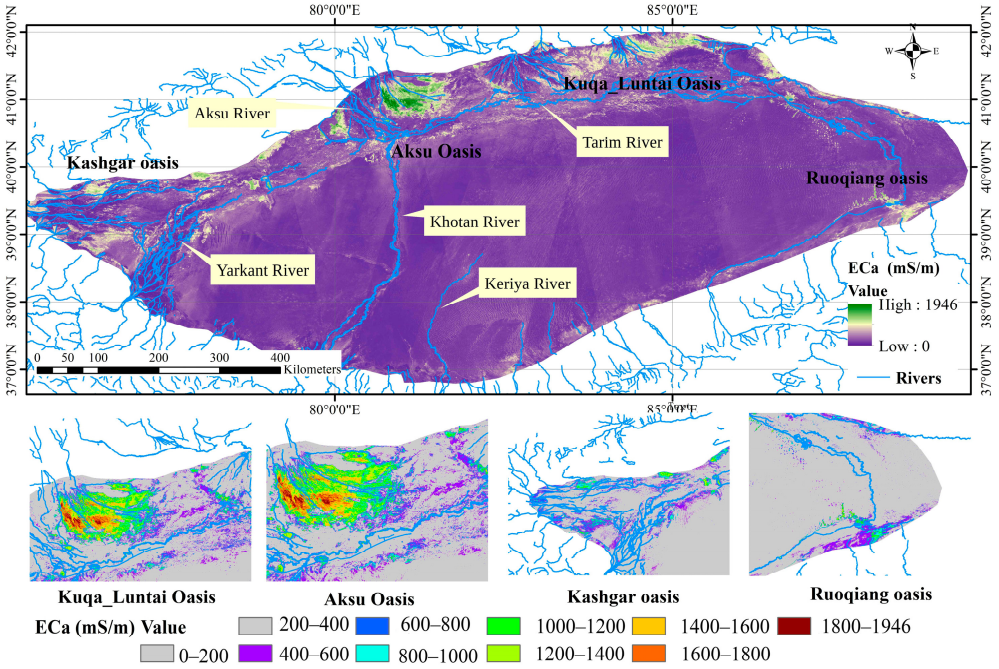

Figure 9. Spatial Distribution Map of Electromagnetic Induction (ECa) in the Tarim Basin. The predicted ECa showed high values in Kuqa–Luntai Oasis, Aksu Oasis, Kashgar Oasis, and Ruoqiang Oasis, indicating that the water and salt contents in these regions were high.

4.4. Sensitivity of Predicted Electromagnetic Induction to Groundwater Levels

The analysis of the relationship between GWL and ECa across all samples revealed no significant correlation between the two (Figure 10). However, the GWL and ECa exhibited different trends for various land use types. The fitted trend lines in the scatter plot indicated the following GWL, from shallow to deep, across different land uses: cropland, grass and shrubland, desert, saline land, and Gobi. Specifically, GWL and ECa showed a week negative relationship in areas such as saline land, desert, and Gobi, whereas no clear trend was observed in grass-, shrub-, or cropland. Considering the interference of anthropogenic irrigation in agricultural fields, this study focused on the saline soil areas, as shown in Figure 10b. The correlation between GWL and ECa was 0.62, indicating that ECa accounted for 38.60% of the spatial variability of GWL in saline land.

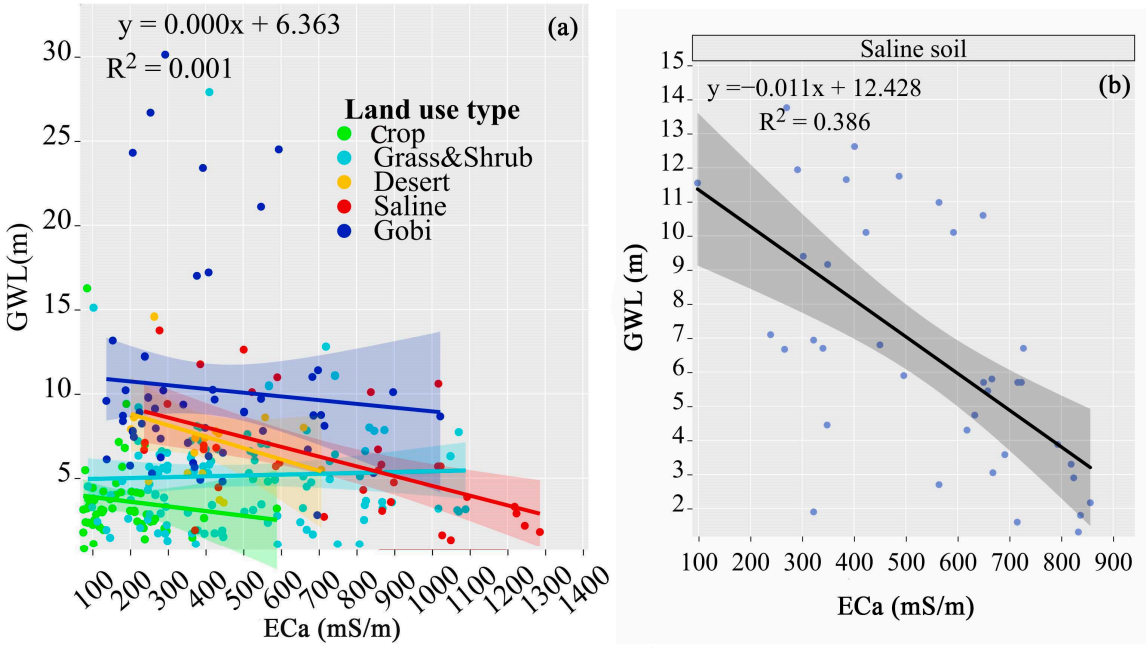

Figure 10. The relationship between ECa and GWL under different land use types. (**a**) The relationship between GWL and ECa under different land uses (**b**) The linear relationship between GWL and ECa in saline soil.

The distribution characteristics of a scatter plot constructed between ECa and GWL were analyzed further (Figure 11). The results show that as the ECa values increased, the range of GWL variations continuously converged. When the ECa values were between 0 and 600 mS/m, the GWL variation range was approximately 1.31–31 m (stage A); 600–800 mS/m corresponded to 1.31–15 m (stage B); and 800–1100 mS/m corresponded to approximately 1.60–10.60 m (stage C). The ECa observations greater than 1100 mS/m corresponded to a range of 0–5 m (stage D, the range of observed values is 1.31–3.03 m).

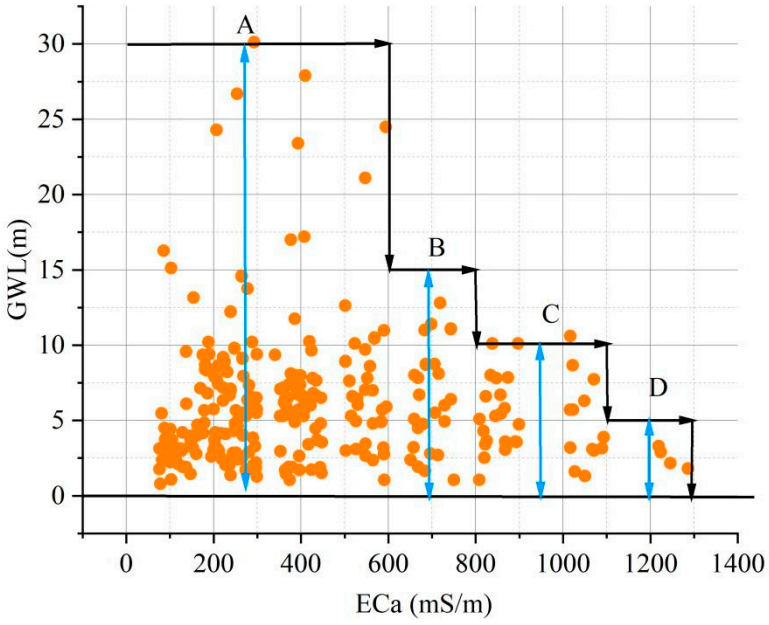

Figure 11. Stage relationship between groundwater level and ECa. Stages A, B, C, and D correspond to four intervals of ECa/GWL: 0~600 mS/m: 0~31 m; 600~800 mS/m: 0~15 m; 800~1100 mS/m: 0~10 m; and >1100 mS/m: <5 m.

5. Discussion

5.1. Performance of Machine Learning

We evaluated various machine learning algorithms to determine the optimal method for modeling GWL and ECa in the study area. Ensemble learning outperformed other approaches, particularly for deeper aquifers (>10 m) and complex environments such as oases, where human influences and subsurface conditions are intricate. Compared with popular algorithms, such as Random Forest and XGBoost, the ensemble method exhibited superior accuracy and stability. This finding highlights the potential of the ensemble method for practical applications in challenging hydrological settings. Accordingly, ensemble machine learning is a powerful tool for groundwater modeling, particularly in complex settings. Given its effectiveness, we used the predictions of the ensemble model in ECa modeling. Our study contributes to the growing body of evidence supporting the use of ensemble methods for hydrological problems [63].

5.2. Advantages and Disadvantages of RS_GWL

In the Tarim Basin, the RS_GWL model had advantages for the prediction of deep groundwater levels, but there was still challenges in the prediction of shallow groundwater. From the current perspective, the RS_GWL model, which leverages remote sensing technology, offers three distinct benefits. Firstly, the RS_GWL model, constructed using the three most effective machine learning algorithms—particularly the ensemble method—demonstrates an interpretation accuracy of approximately 90%. This precision allows the model to provide a reasonably accurate of the overall groundwater distribution within the study area. Secondly, the model's reliability is further substantiated through the application of the optimal algorithm to map the groundwater distribution. As depicted in Figure 8, the model accurately reflects the distribution of groundwater across mountainous regions, oases, and deserts, aligning with both empirical observations and the local land use patterns. For instance, the groundwater in the oasis's agricultural fields is notably shallow, whereas in the mountainous areas, the water table is deepest near the lower edge of the alluvial fan. Lastly, the RS_GWL model continues to exhibit robust performance even in the GWL > 10 m range. It is hypothesized that when the groundwater level exceeds 10 m, surface soil characteristics and vegetation are less influenced by the presence of groundwater [10,26]. Instead, external factors such as topographic variations and climatic changes, which are influenced by the terrain, become the dominant factors [26].

While the RS_GWL model's performance is diminished in the GWL<10 m interval compared to the GWL > 10 m range, the significance test between the predicted and observed values, with a p-value of less than 0.0001(significance testing), indicates a strong statistical correlation. This suggests that the model can effectively differentiate between groundwater levels in agricultural lands and those in areas with natural vegetation growth within this specific interval. However, the model's validation accuracy reveals a limitation in its ability to precisely determine the spatial distribution of GWL, with only 30% explanatory power attributed to spatial variability. The study speculates that this may be related to whether environmental variables accurately characterize surface soils and the reality that there are inherently many different land use patterns corresponding to the same GWL.

5.3. Characterization of ECa in GWL Prediction

ECa values below 800 mS/m lose their capacity to indicate GWL at depth of >10 m. This limitation arises because the scatter distribution within this range exhibits a random pattern. As illustrated in Figure 11, the scatter points within this interval encompass a wide variety of local land use types. However, there is no significant correlation between ECa and GWL across most of these land use areas. The specific reasons for this lack of correlation

are as follows. (1) Mixing GWL observations of multiple crop types reduces the correlation between it and ECa, which is caused by crops being in different phenological cycles and irrigation cycles (Figure 12). (2) Gobi, situated at a higher elevation, features a deeper groundwater table and lower salinity levels. Soil moisture in this region is predominantly replenished by rainfall, rather than groundwater, which affects the ECa measurements. (3) The desert's weak capillary action results in the absence of water and salt in the surface layer [64] (Figure 12). This absence impedes the electromagnetic induction [65], thereby affecting the ECa's ability to reflect GWL. (4) This phenomenon occurred because the variability of ECa values depends on many factors such as soil moisture, clay content, salinity, and cation exchange capacity [66,67].

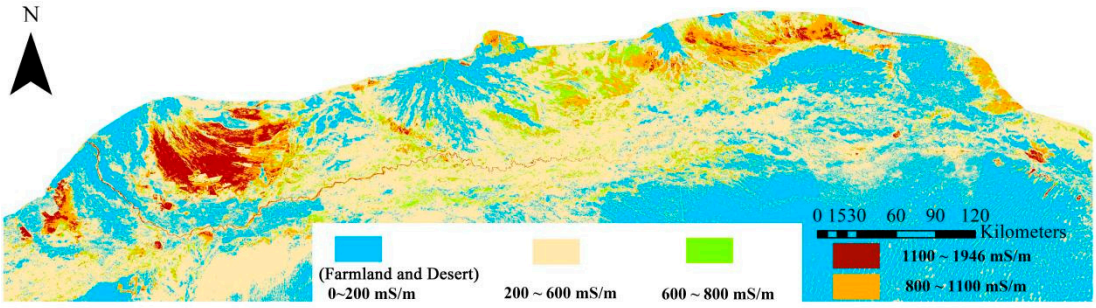

Figure 12. Spatial Distribution Characteristics of ECa within the Oasis Region.

When the ECa ranges from 800 to 1100 mS/m, the variation in GWL is constrained to depths within 10 m. This range suggests that GWL at these depths can influence the water and salt dynamics within the upper 1.5 m of the soil. A comparative analysis of land use distribution within this ECa range indicates that saline soils and shrublands are primarily found in this interval. In the research conducted by Imin et al. [10], it was observed that GWL could support the survival of various sizes of *Populus euphratica* and *Tamarix ramosissima* within the 0 to 10 m depth range. Specifically, individuals of *Tamarix ramosissima*, categorized by small-, medium-, and large sizes, were noted to thrive in areas where the GWL was situated between 2.1 and 7.8 m. The well-developed root systems of these shrubs enable the upward transfer of deep water and salts to the shallow root zone [68–70], making these elements detectable by electromagnetic induction.

When the ECa exceeds 1100 mS/m, the range of the GWL is confined to a depth of 5 m. Notably, the correlation between GWL and ECa was higher in saline land because of the sensitivity of ECa to soil salinity and soil moisture [31]. This depth range is a critical indicator of significant soil salinization occurring at the surface. As depicted in Figure 10a, the sample points within this ECa interval are predominantly found in saline soils. Figure 9 further illustrates that regions with ECa values above 1100 mS/m are primarily located in the Aksu and Kuqa Oases. This distribution aligns with the findings on salinity revealed by Peng et al. [71] and Wang et al. [72]. The distribution of GWL < 5 m, shown in Figures 7 and 12, also indicates that the GWL is predominantly found within the Aksu and Kuqa Oases, which is consistent with the patterns of soil salinity [73].

5.4. Synergistic Strategy for Predicting Groundwater Levels

Subsequent research will synthesize the strengths of both ECa and the RS_GWL to produce more robust predictive outcomes. This study proposes integrating the predictive results of RS_GWL into those of the ECa map, as shown in Figure 13. Agricultural irrigated areas in GWL < 5 m and GWL > 10 m, originating from RS-GWL predictions, are confirmed to be integrated into the ECa map. This has been verified by the ground sample points and land use maps (Figures 6e and 7). Secondly, the ECa readings above 1100 mS/m

have been demonstrated to effectively identify regions in which the GWL was less than 5 m under natural ecological conditions. Up to this point, the areas with GWL less than 5 m and greater than 10 m were delineated, an intermediate zone of 5 m < GWL < 10 m was established (Figure 13). This depth range is critical for vegetation survival in an oasis–desert belt. Living conditions for vegetation deteriorate rapidly once the GWL falls below 10 m [9,10]. After identifying the regions within this GWL range, the national government could rationally allocate resources according to the ecological needs and economic development goals of such regions.

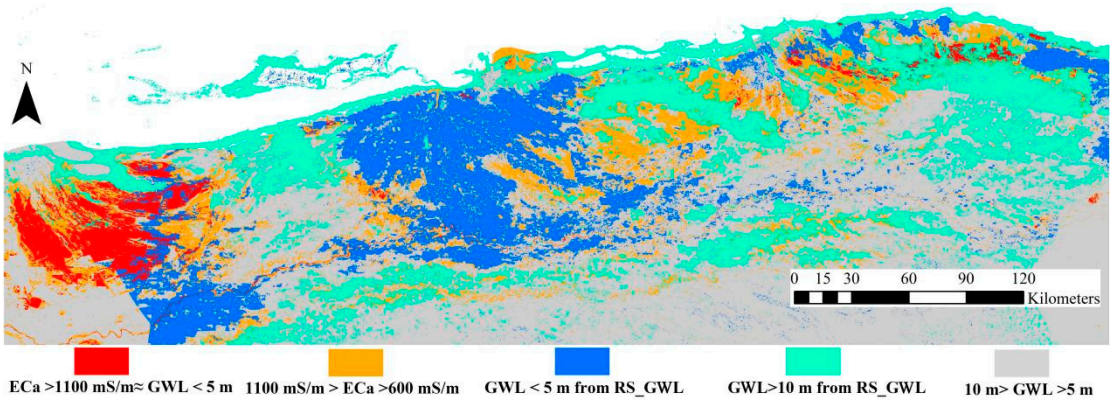

Figure 13. Spatial distribution characteristics of groundwater levels in oasis areas of the study region using an integrated ECa and RS_GWL strategy.

The integration strategy also reflects the important complement of ECa to RS_GWL predictions when detecting GWL in areas with natural vegetation cover (Figure 13). The red area within the black rectangle is the more defined area of GWL < 5 m. However, this area does not overlap with the areas with GWL < 5 m in the RS_GWL predictions (blue area). Only a small portion of the eastern side shows that the predictions of the two different models perceive this region simultaneously (GWL < 5 m). A similar phenomenon occurs in the salinized areas of the Kuqa_Luntai Oasis. This result emphasizes the advantage of ECa in predicting shallow groundwater (GWL < 5 m) in naturally vegetated areas, which may be related to the fact that water–salt variations in soils at depths of 0–1.5 m are more indicative of surface cover–subsurface ecosystems symbiosis [31,33]. In contrast, remote sensing data can only detect the surface to a depth of 5 cm [74], which may be one of the reasons for the weakness of RS_GWL in predicting shallow groundwater in natural vegetation areas.

Although the relationship between areas with ECa readings greater than 1100 mS/m and GWL has been established, the ECa value range of 600–1100 mS/m still requires attention as it likely corresponds to a GWL range of less than 10 m. This range indicates that soil moisture and salt content in the area are influenced by groundwater resources and could notably affect the signal from the EMI equipment. ECa readings in the 600–1100 mS/m range were typically found on the periphery of areas with a GWL less than 5 m (Figure 13) or along the banks of rivers.

5.5. Limitations and Future Directions

Although this research proposed strategies that could improve the accuracy of GWL forecasting, open questions remain regarding the environmental variables to be employed and the influence of temporal ECa on hydrological processes. Initially, we need to recognize that the relationship between ECa and GWL mentioned in the study is based on the prediction accuracy of ECa. Although the prediction accuracy of ECa reached more than 73% ($p < 0.0001$), there is still room for improvement. The employment of high-precision

and high-resolution physical parameters should be considered first. Then, the EM38-MK2 device was limited to a detection depth of 1.50 m beneath the surface; however, the potential concurrent use of EM31 could help address the deficiencies of EM38-MK2 in predicting deeper GWL. The EM31-MK2 was effective down to a depth of 6 m, within which fluctuations in water and salt content were connected closely to the distribution of vegetation communities. In addition, research has indicated that the type and density of vegetation in arid regions are crucial for GWL identification. However, the current high spatial resolution land use data at a scale of 1 m lack the specificity required to detail vegetation types in arid regions. Finally, the area where ECa and GWL data collection intersect within the study region was confined to local oases, limiting the comprehensive mapping of GWL in regions of dense human activity.

Capitalizing on the ability of EMI to identify GWL at depths less than 5 m, our study will continue using historical Landsat data to project ECa variations over the past 20 y. In this way, analyses of spatiotemporal trends in shallow GWL (<5 m) can be facilitated. We anticipate these findings to elucidate the mechanisms through which climate change affects local water resources.

6. Conclusions

In the Tarim Basin, characterized by scant rainfall and robust evaporation, native vegetation is highly sensitive to groundwater depth. When the GWL exceeds the ecological groundwater level threshold, the ecological role of groundwater is substantially weakened, leading to significant deterioration of natural oases. Remote sensing data have been applied extensively in monitoring GWL; however, models based on remote sensing still face challenges in terms of the precise location of shallow groundwater. The study explored two paths to retrieve GWL. First, the remote sensing and geographic information data and the observed GWL were input into an optimal machine learning algorithm, and then the predicted GWL of the study area was generated. The second path was to predict the spatial variation in the ECa of the study area based on the remote sensing and related geographic datasets with the support of the machine learning algorithm. The predicted ECa was then correlated with the GWL and analyzed for a relationship. Finally, we compared the differences between the two paths in predicting GWL and drew the following conclusions.

(1) The prediction accuracies of the groundwater level model (RS_GWL) constructed on the basis of remotely sensed data were $R^2 = 0.30$ and $R^2 = 0.91$ in the intervals of 0~10 m and GWL > 10 m, respectively;
(2) With a continuous increase in ECa values, variability in the observed GWL decreased. In particular, when the ECa readings were between 800 and 1100 mS/m, the GWL fluctuated within the 0–10 m range, and when the ECa readings exceeded 1100 mS/m, the GWL changed within an area of less than 5 m.
(3) By integrating the advantages of both ECa and RS_GWL models, GWL < 5 m (natural vegetation growth area and ECa > 1100 mS/m area), GWL < 5 m (occupied by farmland from RS_GWL), GWL > 10 m (from RS_GWL), and 5 < GWL < 10 m (be speculated) were identified.

In conclusion, this study indicates that the remote sensing method for detecting groundwater levels (GWL) is not only feasible but also offers significant advantages in terms of cost and spatial efficiency compared to traditional exploration methods. Furthermore, the incorporation of EMI enhances the accuracy of predicting shallow GWL in oasis plain areas, surpassing the results obtained from the direct application of remote sensing data alone.

Author Contributions: Conceptualization, Y.W. and X.G.; Methodology, Y.W.; Software, Y.W. and X.G.; Formal analysis, X.G.; Investigation, Y.W. and X.G.; Data curation, L.H.; Writing—original draft, F.W.; Writing—review & editing, L.L. and Y.W.; Supervision, Y.W. and X.G.; Funding acquisition, F.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (42101363). This project was also supported by the Research Initiation Fund of Chengdu University (2081923044 and 2081923045).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We are very proud to have been invited to participate in the environmental assessment project, and we also thank the truck-mounted drilling operators and samplers who worked in the difficult conditions.

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

References

1. Sahour, H.; Sultan, M.; Abdellatif, B.; Emil, M.; Abotalib, A.Z.; Abdelmohsen, K.; Vazifedan, M.; Mohammad, A.T.; Hassan, S.M.; Metwalli, M.R.; et al. Identification of shallow groundwater in arid lands using multi-sensor remote sensing data and machine learning algorithms. *J. Hydrol.* **2022**, *614*, 128509. [CrossRef]
2. Awadh, S.M.; Al-Mimar, H.; Yaseen, Z.M. Groundwater availability and water demand sustainability over the upper mega aquifers of Arabian Peninsula and west region of Iraq. *Environ. Dev. Sustain.* **2021**, *23*, 1–21. [CrossRef]
3. Mendes, M.P.; Ribeiro, L.; David, T.S.; Costa, A. How dependent are cork oak (*Quercus suber* L.) woodlands on groundwater? A case study in southwestern Portugal. *For. Ecol. Manag.* **2016**, *378*, 122–130. [CrossRef]
4. Wu, X.; Wang, X.; Wang, P.; Gu, Y.; Li, Y. Effects of groundwater depth on ecological stoichiometric characteristics of assimilated branches and soil of two desert plants. *Front. Plant Sci.* **2023**, *14*, 1225907. [CrossRef]
5. Pan, S.; Wang, Z.; Su, Q.; Sun, T.; Zhang, Y. Groundwater level monitoring model using multi-temporal images in arid region of northwest China. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 745–750.
6. Martinetti, S.; Fatichi, S.; Floriancic, M.; Burlando, P.; Molnar, P. Field evidence of riparian vegetation response to groundwater levels in a gravel-bed river. *Ecohydrology* **2021**, *14*, e2264. [CrossRef]
7. Chen, Y.; Chen, Y.; Xu, C.; Ye, Z.; Li, Z.; Zhu, C.; Ma, X. Effects of ecological water conveyance on groundwater dynamics and riparian vegetation in the lower reaches of Tarim River, China. *Hydrol. Process. Int. J.* **2010**, *24*, 170–177. [CrossRef]
8. Zhang, D.; Liu, X.; Simmons, C.T.; Zhang, L.; Zhang, Q. Changes in groundwater levels across China from 2005 to 2016. *J. Hydrol.* **2023**, *623*, 129781. [CrossRef]
9. Hao, X.; Li, W.; Huang, X.; Zhu, C.; Ma, J. Assessment of the groundwater threshold of desert riparian forest vegetation along the middle and lower reaches of the Tarim River, China. *Hydrol. Process. Int. J.* **2010**, *24*, 178–186. [CrossRef]
10. Imin, B.; Dai, Y.; Shi, Q.; Guo, Y.; Li, H.; Nijat, M. Responses of two dominant desert plant species to the changes in groundwater depth in hinterland natural oasis, Tarim Basin. *Ecol. Evol.* **2021**, *11*, 9460–9471. [CrossRef]
11. Sun, D.; Zhao, C.; Wei, H.; Peng, D. Simulation of the relationship between land use and groundwater level in Tailan River basin, Xinjiang, China. *Quat. Int.* **2011**, *244*, 254–263. [CrossRef]
12. Maihemuti, B.; Simayi, Z.; Alifujiang, Y.; Aishan, T.; Abliz, A.; Aierken, G. Development and evaluation of the soil water balance model in an inland arid delta oasis: Implications for sustainable groundwater resource management. *Glob. Ecol. Conserv.* **2021**, *25*, e01408. [CrossRef]
13. Chen, S.; Aidi, H.; Wenke, G. Remote Sensing Monitoring Method for Groundwater Level on Aeolian Desertification Area. *J. Water Chem. Technol.* **2020**, *42*, 522–529. [CrossRef]
14. Kaur, L.; Rishi, M.S. Integrated geospatial, geostatistical, and remote-sensing approach to estimate groundwater level in Northwestern India. *Environ. Earth Sci.* **2018**, *77*, 786. [CrossRef]
15. Ndou, N.N.; Palamuleni, L.G.; Ramoelo, A. Modelling depth to groundwater level using SEBAL-based dry season potential evapotranspiration in the upper Molopo River Catchment, South Africa. *Egypt. J. Remote Sens. Space Sci.* **2018**, *21*, 237–248. [CrossRef]
16. Silva, C.d.O.F.; Manzione, R.L.; Albuquerque Filho, J.L. Combining remotely sensed actual evapotranspiration and GIS analysis for groundwater level modeling. *Environ. Earth Sci.* **2019**, *78*, 462. [CrossRef]
17. Stateczny, A.; Narahari, S.C.; Vurubindi, P.; Guptha, N.S.; Srinivas, K. Underground Water Level Prediction in Remote Sensing Images Using Improved Hydro Index Value with Ensemble Classifier. *Remote Sens.* **2023**, *15*, 2015. [CrossRef]

18. Suganthi, S.; Elango, L.; Subramanian, S. Groundwater potential zonation by Remote Sensing and GIS techniques and its relation to the Groundwater level in the Coastal part of the Arani and Koratalai River Basin, Southern India. *Earth Sci. Res. J.* **2013**, *17*, 87–95.
19. Tiwari, A.K.; Singh, P.K.; Chandra, S.; Ghosh, A. Assessment of groundwater level fluctuation by using remote sensing and GIS in West Bokaro coalfield, Jharkhand, India. *ISH J. Hydraul. Eng.* **2016**, *22*, 59–67. [CrossRef]
20. Gao, L.; Sadeghi, M.; Ebtehaj, A.; Wigneron, J.-P. A temporal polarization ratio algorithm for calibration-free retrieval of soil moisture at L-band. *Remote Sens. Environ.* **2020**, *249*, 112019. [CrossRef]
21. Costa, L.; Nunes, L.; Ampatzidis, Y. A new visible band index (vNDVI) for estimating NDVI values on RGB images utilizing genetic algorithms. *Comput. Electron. Agric.* **2020**, *172*, 105334. [CrossRef]
22. Frappart, F.; Ramillien, G. Monitoring groundwater storage changes using the Gravity Recovery and Climate Experiment (GRACE) satellite mission: A review. *Remote Sens.* **2018**, *10*, 829. [CrossRef]
23. Miro, M.E.; Famiglietti, J.S. Downscaling GRACE remote sensing datasets to high-resolution groundwater storage change maps of California's Central Valley. *Remote Sens.* **2018**, *10*, 143. [CrossRef]
24. Rateb, A.; Scanlon, B.R.; Pool, D.R.; Sun, A.; Zhang, Z.; Chen, J.; Clark, B.; Faunt, C.C.; Haugh, C.J.; Hill, M. Comparison of groundwater storage changes from GRACE satellites with monitoring and modeling of major US aquifers. *Water Resour. Res.* **2020**, *56*, e2020WR027556. [CrossRef]
25. Duan, S.-B.; Li, Z.-L.; Li, H.; Göttsche, F.-M.; Wu, H.; Zhao, W.; Leng, P.; Zhang, X.; Coll, C. Validation of Collection 6 MODIS land surface temperature product using in situ measurements. *Remote Sens. Environ.* **2019**, *225*, 16–29. [CrossRef]
26. Wei, Y.; Wang, F.; Hong, B.; Yang, S. Revealing spatial variability of groundwater level in typical ecosystems of the Tarim Basin through ensemble algorithms and limited observations. *J. Hydrol.* **2023**, *620*, 129399. [CrossRef]
27. Arshad, A.; Mirchi, A.; Vilcaez, J.; Umar Akbar, M.; Madani, K. Reconstructing high-resolution groundwater level data using a hybrid random forest model to quantify distributed groundwater changes in the Indus Basin. *J. Hydrol.* **2024**, *628*, 130535. [CrossRef]
28. Huth, N.I.; Poulton, P.L. An electromagnetic induction method for monitoring variation in soil moisture in agroforestry systems. *Soil. Res.* **2007**, *45*, 63–72. [CrossRef]
29. Misra, R.K.; Padhi, J. Assessing field-scale soil water distribution with electromagnetic induction method. *J. Hydrol.* **2014**, *516*, 200–209. [CrossRef]
30. Moghadas, D.; Jadoon, K.Z.; McCabe, M.F. Spatiotemporal monitoring of soil water content profiles in an irrigated field using probabilistic inversion of time-lapse EMI data. *Adv. Water Resour.* **2017**, *110*, 238–248. [CrossRef]
31. Jiang, Z.-Y.; Li, X.-Y.; Wu, H.-W.; Zhang, S.-Y.; Zhao, G.-Q.; Wei, J.-Q. Linking spatial distributions of the patchy grass Achnatherum splendens with dynamics of soil water and salt using electromagnetic induction. *Catena* **2017**, *149*, 261–272. [CrossRef]
32. Turkeltaub, T.; Wang, J.; Cheng, Q.; Jia, X.; Zhu, Y.; Shao, M.-A.; Binley, A. Soil moisture and electrical conductivity relationships under typical Loess Plateau land covers. *Vadose Zone J.* **2022**, *21*, e20174. [CrossRef]
33. Farzamian, M.; Bouksila, F.; Paz, A.M.; Santos, F.M.; Zemni, N.; Slama, F.; Ben Slimane, A.; Selim, T.; Triantafilis, J. Landscape-scale mapping of soil salinity with multi-height electromagnetic induction and quasi-3d inversion in Saharan Oasis, Tunisia. *Agric. Water Manag.* **2023**, *284*, 108330. [CrossRef]
34. Gu, S.; Jiang, S.; Li, X.; Zheng, N.; Xia, X. Soil salinity simulation based on electromagnetic induction and deep learning. *Soil. Tillage Res.* **2023**, *230*, 105706. [CrossRef]
35. Paz, A.M.; Castanheira, N.; Farzamian, M.; Paz, M.C.; Gonçalves, M.C.; Monteiro Santos, F.A.; Triantafilis, J. Prediction of soil salinity and sodicity using electromagnetic conductivity imaging. *Geoderma* **2020**, *361*, 114086. [CrossRef]
36. Kiremit, M.S.; Arslan, H. Effects of different shallow and saline groundwater depths on soil salinity, evapotranspiration, grain yield and spike traits of winter wheat. *J. Agron. Crop Sci.* **2023**, *209*, 402–421. [CrossRef]
37. Pham, T.V.; Thi Do, T.A.; Tran, H.D.; Thi Do, A.N. Assessing groundwater potential for mitigating salinity issues in agricultural areas of southern Dong Nai province, Vietnam. *Groundw. Sustain. Dev.* **2024**, *25*, 101177. [CrossRef]
38. Zhang, H.; Li, Y.; Meng, Y.-l.; Cao, N.; Li, D.-s.; Zhou, Z.-g.; Chen, B.-l.; Dou, F.-g. The effects of soil moisture and salinity as functions of groundwater depth on wheat growth and yield in coastal saline soils. *J. Integr. Agric.* **2019**, *18*, 2472–2482. [CrossRef]
39. Sherlock, M.D.; McDonnell, J.J. A new tool for hillslope hydrologists: Spatially distributed groundwater level and soilwater content measured using electromagnetic induction. *Hydrol. Process.* **2003**, *17*, 1965–1977. [CrossRef]
40. Hedley, C.B.; Roudier, P.; Yule, I.J.; Ekanayake, J.; Bradbury, S. Soil water status and water table depth modelling using electromagnetic surveys for precision irrigation scheduling. *Geoderma* **2013**, *199*, 22–29. [CrossRef]
41. Jiapaer, G.; Chen, X.; Bao, A. A comparison of methods for estimating fractional vegetation cover in arid regions. *Agric. For. Meteorol.* **2011**, *151*, 1698–1710. [CrossRef]
42. Bao, A.; Huang, Y.; Ma, Y.; Guo, H.; Wang, Y. Assessing the effect of EWDP on vegetation restoration by remote sensing in the lower reaches of Tarim River. *Ecol. Indic.* **2017**, *74*, 261–275. [CrossRef]

43. Li, Z.; Yu, G.-A.; Brierley, G.J.; Wang, Z.; Jia, Y. Migration and cutoff of meanders in the hyperarid environment of the middle Tarim River, northwestern China. *Geomorphology* **2017**, *276*, 116–124. [CrossRef]
44. Fan, Y.; Chen, Y.; Liu, Y.; Li, W. Variation of baseflows in the headstreams of the Tarim River Basin during 1960–2007. *J. Hydrol.* **2013**, *487*, 98–108. [CrossRef]
45. Zhang, H.; Xue, L.; Wei, G.; Dong, Z.; Meng, X. Assessing vegetation dynamics and landscape ecological risk on the mainstream of Tarim River, China. *Water* **2020**, *12*, 2156. [CrossRef]
46. Tromp-van Meerveld, H.J.; McDonnell, J.J. Assessment of multi-frequency electromagnetic induction for determining soil moisture patterns at the hillslope scale. *J. Hydrol.* **2009**, *368*, 56–67. [CrossRef]
47. Pablos, M.; Martínez-Fernández, J.; Piles, M.; Sánchez, N.; Vall-llossera, M.; Camps, A. Multi-temporal evaluation of soil moisture and land surface temperature dynamics using in situ and satellite observations. *Remote Sens.* **2016**, *8*, 587. [CrossRef]
48. Kern, J.S. Evaluation of soil water retention models based on basic soil physical properties. *Soil. Sci. Soc. Am. J.* **1995**, *59*, 1134–1141. [CrossRef]
49. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Clim.* **2017**, *37*, 4302–4315. [CrossRef]
50. Condon, L.E.; Maxwell, R.M. Evaluating the relationship between topography and groundwater using outputs from a continental-scale integrated hydrology model. *Water Resour. Res.* **2015**, *51*, 6602–6621. [CrossRef]
51. Gong, L.; Ran, Q.; He, G.; Tiyip, T. A soil quality assessment under different land use types in Keriya river basin, Southern Xinjiang, China. *Soil. Till Res.* **2015**, *146*, 223–229. [CrossRef]
52. Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* **2019**, *64*, 370–373. [CrossRef]
53. Su, X.; Yan, X.; Tsai, C.L. Linear regression. *Wiley Interdiscip. Rev. Comput. Stat.* **2012**, *4*, 275–294. [CrossRef]
54. Rokach, L.; Maimon, O. Decision trees. In *Data Mining and Knowledge Discovery Handbook*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 165–192.
55. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *MLear* **2006**, *63*, 3–42. [CrossRef]
56. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
57. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3149–3157.
58. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
59. Caruana, R.; Niculescu-Mizil, A.; Crew, G.; Ksikes, A. Ensemble selection from libraries of models. In Proceedings of the Twenty-First International Conference on Machine Learning, Banff, AB, Canada, 4–8 July 2004; p. 18.
60. Ding, J.; Yang, S.; Shi, Q.; Wei, Y.; Wang, F. Using Apparent Electrical Conductivity as Indicator for Investigating Potential Spatial Variation of Soil Salinity across Seven Oases along Tarim River in Southern Xinjiang, China. *Remote Sens.* **2020**, *12*, 2601. [CrossRef]
61. Li, W.; Huang, F.; Shi, F.; Wei, X.; Zamanian, K.; Zhao, X. Human and climatic drivers of land and water use from 1997 to 2019 in Tarim River basin, China. *Int. Soil Water Conserv. Res.* **2021**, *9*, 532–543. [CrossRef]
62. Ye, Z.; Chen, S.; Zhang, Q.; Liu, Y.; Zhou, H. Ecological water demand of Taitema Lake in the lower reaches of the Tarim River and the Cherchen River. *Remote Sens.* **2022**, *14*, 832. [CrossRef]
63. Zounemat-Kermani, M.; Batelaan, O.; Fadaee, M.; Hinkelmann, R. Ensemble machine learning paradigms in hydrology: A review. *J. Hydrol.* **2021**, *598*, 126266. [CrossRef]
64. Voigt, C.; Klipsch, S.; Herwartz, D.; Chong, G.; Staubwasser, M. The spatial distribution of soluble salts in the surface soil of the Atacama Desert and their relationship to hyperaridity. *Glob. Planet. Chang.* **2020**, *184*, 103077. [CrossRef]
65. Taghizadeh-Mehrjardi, R.; Minasny, B.; Sarmadian, F.; Malone, B.P. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* **2014**, *213*, 15–28. [CrossRef]
66. Aragüés Lafarga, R.; Urdanoz, V.; Cetin, M.; Kirda, C.; Daghari, H.; Ltifi, W.; Lahlou, M.; Douaik, A. Soil salinity related to physical soil characteristics and irrigation management in four Mediterranean irrigation districts. *Agric. Water Manag.* **2011**, *98*, 959–966. [CrossRef]
67. Coppola, A.; Smettem, K.; Ajeel, A.; Saeed, A.; Dragonetti, G.; Comegna, A.; Lamaddalena, N.; Vacca, A. Calibration of an electromagnetic induction sensor with time-domain reflectometry data to monitor rootzone electrical conductivity under saline water irrigation. *Eur. J. Soil. Sci.* **2016**, *67*, 737–748. [CrossRef]
68. Cao, Q.; Yang, B.; Li, J.; Wang, R.; Liu, T.; Xiao, H. Characteristics of soil water and salt associated with *Tamarix ramosissima* communities during normal and dry periods in a semi-arid saline environment. *Catena* **2020**, *193*, 104661. [CrossRef]

69. Fu, A.; Li, W.; Chen, Y. The threshold of soil moisture and salinity influencing the growth of *Populus euphratica* and *Tamarix ramosissima* in the extremely arid region. *Environ. Earth Sci.* **2012**, *66*, 2519–2529. [CrossRef]
70. Xia, J.; Zhang, S.; Zhao, X.; Liu, J.; Chen, Y. Effects of different groundwater depths on the distribution characteristics of soil-Tamarix water contents and salinity under saline mineralization conditions. *Catena* **2016**, *142*, 166–176. [CrossRef]
71. Peng, J.; Biswas, A.; Jiang, Q.; Zhao, R.; Hu, J.; Hu, B.; Shi, Z. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* **2019**, *337*, 1309–1319. [CrossRef]
72. Wang, N.; Chen, S.; Huang, J.; Frappart, F.; Taghizadeh, R.; Zhang, X.; Wigneron, J.-P.; Xue, J.; Xiao, Y.; Peng, J. Global Soil Salinity Estimation at 10 m Using Multi-Source Remote Sensing. *J. Remote Sens.* **2024**, *4*, 0130. [CrossRef]
73. Ma, S.; He, B.; Xie, B.; Ge, X.; Han, L. Investigation of the spatial and temporal variation of soil salinity using Google Earth Engine: A case study at Werigan–Kuqa Oasis, West China. *Sci. Rep.* **2023**, *13*, 2754. [CrossRef]
74. Richer-de-Forges, A.C.; Chen, Q.; Baghdadi, N.; Chen, S.; Gomez, C.; Jacquemoud, S.; Martelet, G.; Mulder, V.L.; Urbina-Salazar, D.; Vaudour, E. remote sensing data for digital soil mapping in French research—A review. *Remote Sens.* **2023**, *15*, 3070. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.