MDPI

Article

Improving the Resolution and Accuracy of Groundwater Level Anomalies Using the Machine Learning-Based Fusion Model in the North China Plain

Gangqiang Zhang [1,2,†] , Wei Zheng [1,2,3,4,5,*,†], Wenjie Yin [2] and Weiwei Lei [1]

1 School of Surveying and Landing Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China; 211804020040@home.hpu.edu.cn (G.Z.); wwlei@hpu.edu.cn (W.L.)
2 Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology, Beijing 100094, China; yinwenjie@qxslab.cn
3 School of Geomatics, Liaoning Technical University, Fuxin 123000, China
4 School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China
5 School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China
* Correspondence: zhengwei1@qxslab.cn
† Those authors contributed equally to this paper.

Abstract: The launch of GRACE satellites has provided a new avenue for studying the terrestrial water storage anomalies (TWSA) with unprecedented accuracy. However, the coarse spatial resolution greatly limits its application in hydrology researches on local scales. To overcome this limitation, this study develops a machine learning-based fusion model to obtain high-resolution (0.25°) groundwater level anomalies (GWLA) by integrating GRACE observations in the North China Plain. Specifically, the fusion model consists of three modules, namely the downscaling module, the data fusion module, and the prediction module, respectively. In terms of the downscaling module, the GRACE-Noah model outperforms traditional data-driven models (multiple linear regression and gradient boosting decision tree (GBDT)) with the correlation coefficient (CC) values from 0.24 to 0.78. With respect to the data fusion module, the groundwater level from 12 monitoring wells is incorporated with climate variables (precipitation, runoff, and evapotranspiration) using the GBDT algorithm, achieving satisfactory performance (mean values: CC: 0.97, RMSE: 1.10 m, and MAE: 0.87 m). By merging the downscaled TWSA and fused groundwater level based on the GBDT algorithm, the prediction module can predict the water level in specified pixels. The predicted groundwater level is validated against 6 in-situ groundwater level data sets in the study area. Compare to the downscaling module, there is a significant improvement in terms of CC metrics, on average, from 0.43 to 0.71. This study provides a feasible and accurate fusion model for downscaling GRACE observations and predicting groundwater level with improved accuracy.

Keywords: machine learning-based fusion model; GRACE; gradient boosting decision tree; groundwater level anomalies; statistical downscaling; North China Plain

1. Introduction

As a significant supply source of freshwater resources, groundwater plays a crucial role in social production and human life [1,2]. Globally, it provides drinking water for approximately two billion people [3] and irrigation for roughly 40% of areas equipped for irrigation [4]. Due to extreme climate episodes and anthropogenic actives (e.g., drought and overuse of irrigation water), groundwater resources are seriously over-exploited in some typical regions [5,6], leading to a series of environmental issues, such as land subsidence and seawater intrusion [7,8]. Therefore, understanding the dynamics changes in groundwater is necessary for the effective utilization and sustainable management of water resources [2,9].

The traditional method to monitor groundwater levels is based on ground-based measurements [10,11]. However, it is not applicable to large-scale and remote regions restricted by national policies, limited stations, and instrument accuracy. The Gravity Recovery and Climate Experiment (GRACE) satellites, successfully launched in March 2002 [12], provides a kind of new method for monitoring the global time-variable gravity field with unprecedented accuracy [13]. Additionally, it can provide continuous terrestrial water storage anomalies (TWSA) and cover most parts of the world, which is especially beneficial for areas lacking ground-based measurements. By integrating auxiliary information from hydrological models, groundwater storage anomalies (GWSA) can be further isolated from GRACE observations. Previous studies have demonstrated that GRACE missions show great potential in various fields, e.g., detection of terrestrial water storage [14–16] and groundwater storage [17,18]. Swenson et al. [14] derived region-scale water storage by applying averaging kernels to a realistic synthetic GRACE gravity signal within North American river basins. Results indicated that the accuracy might be better than 1 cm for regions with 400,000 km^2 or larger areas. Rodell et al. [17] simulated groundwater storage based on GRACE and hydrological modeling system, and the results showed that it was being depleted at a mean rate of 4.0 ± 1.0 cm/yr equivalent height of water over the Indian states of Rajasthan, Punjab and Haryana (including Delhi).

Although remarkable achievements have been made in large-scale areas, the application of GRACE observations in local areas is very limited due to the coarse spatial resolution (~200,000 km^2) [19]. Consequently, some downscaling methods have been proposed for providing high-resolution GRACE products, which are mainly divided into two categories: dynamic downscaling and statistical downscaling [20,21], respectively. Normally, dynamic downscaling tends to achieve regional downscaling by using the initial boundary conditions of global climate models [22,23] directly. For example, Eicker et al. [24] assimilated GRACE-derived TWSA into the WaterGAP Global Hydrology Model by introducing a new Kalman filter method, which can provide reasonable results in the Mississippi river basin. Although data assimilation methods remain consistent in the physical process, some shortcomings still require to be considered [25]. The implementation of data assimilation is relatively complicated [26], and its accuracy is subject to the full error covariance matrix of GRACE observations and hydrological models [27,28].

Compared to dynamic downscaling, statistical downscaling usually establishes the linear or non-linear relationships between input and target variables, aiming to produce local-scale information [29,30]. Initially, linear regression models are employed to downscale GRACE products [31–33]. For example, Ning et al. [31] achieved the downscaling of GRACE data in parts of Yunnan by constructing an empirical regression model based on the water balance equation, and the results indicated the feasibility for downscaling GRACE data. Yin et al. [33] proposed a new statistical downscaling algorithm by building the relationship between multi-source evapotranspiration (ET) products and GWSA in the North China Plain, which obtained desirable downscaled results but limited by the strong correlation between TWS and ET. Practically, the relationships between predictor and predictand tend to be non-linear rather than linear. The development of machine learning algorithms provides effective measures to quantify the complicated relationship by constructing non-linear models. Artificial neural networks (ANNs) have the capabilities of simulating complex hydrological characteristics to an arbitrary degree of accuracy [34,35]. This makes ANN becomes an attractive measure in the downscaling researches, which have been applied to some typical regions, e.g., the Northern High Plains [36], California's Central Valley [34], and the Lower Peninsula of Michigan [26]. Similarly, some tree-based machine learning algorithms (e.g., random forest (RF) and gradient boosting decision tree (GBDT)) become popular in regression tasks with the advantages of simplicity and effectiveness. The RF algorithm has been utilized to downscale GRACE observations and obtained satisfactory results in some areas [37,38]. As a kind of ensemble machine learning algorithm, GBDT performs well in constructing non-linear regression models, which is often employed to forecast ET [39] and urban flood [40,41], but rarely in GWLA.

Furthermore, the multi-stage machine learning algorithm may have more powerful expressive performance than a single algorithm in downscaling GRACE products. For example, Seyoum et al. [42] designed a two-layer boosted regression trees (BRT) model by utilizing GRACE data and hydrological variables in a glacial aquifer system of the United States, which can predict groundwater level anomalies (GWLA) with a high spatial resolution.

The North China Plain (NCP), which is the political, economic, and cultural center of China [43], has been suffering from water shortage and over-exploitation of groundwater for a long period [44–46]. Some downscaling researches have been conducted in the area, aiming to provide high-resolution water storage estimates [33,37,47]. Water resources managers are often more concerned with information about water levels at specified locations, while few studies are conducted with respect to this aspect. To overcome this limitation, this study proposes a machine learning-based fusion model, aiming to downscale GRACE-derived TWSA to higher spatial resolution products and predict higher-accuracy groundwater level anomalies (GWLA). The structure of this study is organized as follows. The overview of the NCP and the data sources are introduced in Section 2. Section 3 describes the structure and construction of the machine learning-based fusion model. Section 4 provides the results of the downscaling and prediction module in the NCP. The discussions and conclusions are presented in Sections 5 and 6, respectively.

2. Study Area and Data

2.1. Study Area

The North China Plain, located in the eastern coastal region of China, lies between latitude 35° N–41° N and longitude 113° E–120° E (Figure 1). It is one of the three great plains in China, covering an area of approximately 140,000 km^2 [48]. The NCP is a central agricultural area in China, which produces about one-fourth of the country's total grain yield [10]. The main crops include winter wheat and summer maize, and the NCP supplies more than 50% of the wheat and approximately 33% of the maize production in China [49].
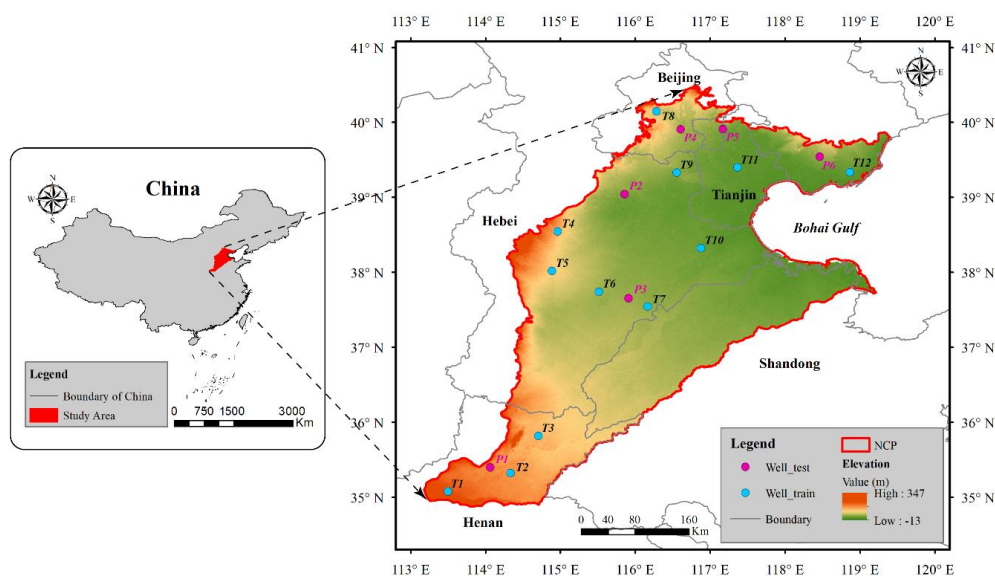
Figure 1. Location and digital elevation model map of the NCP. The blue and magenta dots represent the training and testing wells, respectively.

The NCP belongs to a continental monsoon climate with an annual average temperature between 8 and 15 °C [33,50]. The annual precipitation, most of which occurs during the growth period of summer maize, ranges from 500 to 600 mm, and annual evaporation is 900–1400 mm [51,52]. The NCP contains a shallow unconfined aquifer (40–60 m) and three confined aquifers of different depths (120–170 m, 250–360 m, and 400–600 m) [53].

2.2. Data

The fusion model, proposed in this study, is designed based on the water balance principle and machine learning methods (i.e., multiple linear regression and GBDT). Several datasets (terrestrial water storage anomalies, precipitation, runoff, evapotranspiration, soil moisture, snow water equivalent, and groundwater level) are chosen to owe to their close relationship with groundwater storage changes, as shown in Figure 2.

Variables	Resolutions	Units	Processing	Sources
GRACE (TWSA)	1°×1°	mm	Averaging	GFZ(http://gravis.gfz-potsdam.de/) CSR(http://www2.csr.utexas.edu/) JPL (https://grace.jpl.nasa.gov/)
Precipitation	1°×1° 0.25°×0.25°	mm	Remove the average and resampling	TRMM (https://disc.gsfc.nasa.gov/ datasets?keywords=TRMM)
Runoff	1°×1° 0.25°×0.25°	mm	Remove the average	GLDAS (https://disc.gsfc.nasa.gov/ datasets?keywords=GLDAS)
Evapotranspiration	1°×1° 0.25°×0.25°	mm	Remove the average	GLEAM (https://www.gleam.eu/#4)
Soil Moisture	1°×1° 0.25°×0.25°	mm	Remove the average	GLDAS (https://disc.gsfc.nasa.gov/ datasets?keywords=GLDAS)
Snow Water Equivalent	1°×1° 0.25°×0.25°	mm	Remove the average	GLDAS (https://disc.gsfc.nasa.gov/ datasets?keywords=GLDAS)
Groundwater Level	points	mm	Remove the average	Yearbook (http://www.hwcc.gov.cn/ hwcc/wwgj/xxgb/szygb/)

Figure 2. Summary of variable information (resolutions, units, processing, and sources) employed in the fusion model.

Specifically, some variables should be resampled from 0.25° to 1° for matching the spatial availability of GRACE products, and the study period covers from January 2005 to December 2014, with a total of 120 months. The schematic diagram of the water balance principle is shown in Figure 3.

GRACE (TWSA)
TRMM (P)
GLDAS (SM、R、ET)
GLEAM (ET)
MODIS (ET)

Precipitation (P)
Runoff (R)
Evapotranspiration (ET)

$$P - R - ET = \frac{ds}{dt} \quad (s: \text{TWSA})$$

$$\frac{ds}{dt} \approx \frac{dTWSA}{dt} \approx \frac{TWSA_t - TWSA_{t-1}}{\Delta t}$$

$\frac{ds}{dt}$: Terrestrial water storage change

Soil moisture (SM)
Snow water equivalent (SWE)
Groundwater storage (GWS)

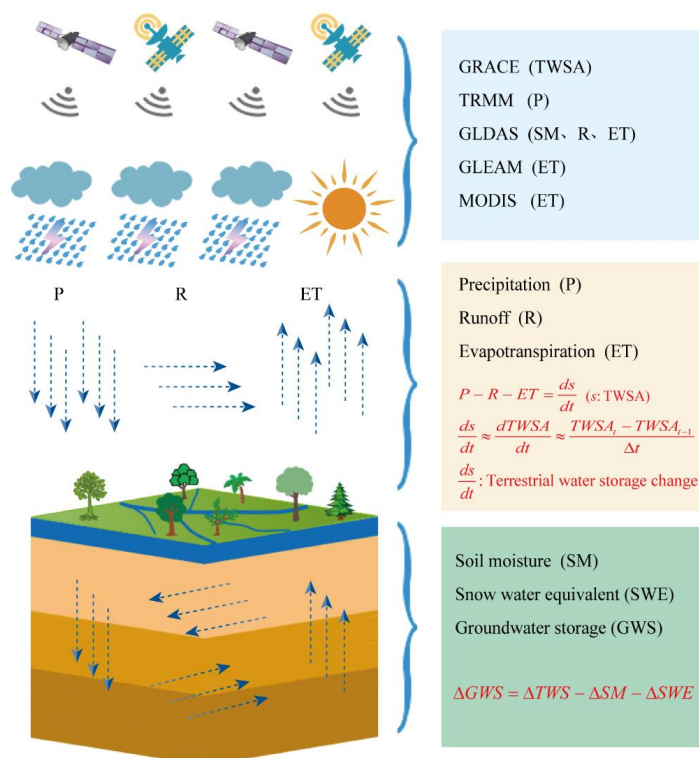$$\Delta GWS = \Delta TWS - \Delta SM - \Delta SWE$$

Figure 3. Sketch map of the water cycle.

2.2.1. GRACE TWSA

The GRACE gravity satellites, jointly developed by NASA (National Aeronautics and Space Administration) and DLR (German Aerospace Center), were launched in 2002 and successfully completed its missions in 2017 [13]. They were designed to track global mass changes or gravity variations using the K-band ranging system and low-low satellite tracking satellite mode [13,54]. In this study, GRACE observations are provided by the Jet Propulsion Laboratory (JPL), Center for Space Research (CSR), and GeoForschungsZentrum Potsdam (GFZ), respectively. The gridded-gain factors are utilized to reduce the leakage error [55], which are available at [56]. Some discrepancies exist among these three solutions due to different processing strategies and tuning parameters [57]. Therefore, we utilize the ensemble average of different solutions as the representative TWS estimates in the following discussion.

2.2.2. TRMM Precipitation

The Tropical Rainfall Measurement Mission (TRMM) is a joint project of NASA and Japan Aerospace Exploration Agency (JAXA), aiming to analyze the impact of rainfall data on weather and climate [58]. The monthly precipitation products (2003–2015) used in this study are the TRMM 3b43 with the spatial resolution of $0.25° \times 0.25°$. The dataset can be obtained from the Goddard Earth Sciences Data and Information Services Center (GES DISC) [59]. Previous studies [16] have demonstrated that TRMM data match well with gauged stations compared to other remotes sensing products, thus utilized in this study.

2.2.3. GLDAS Data

The Global Land Data Assimilation System (GLDAS) is developed by the Goddard Space Flight Center (GSFC). The primary goal of the GLDAS is to ingest satellite- and ground-based observational data products, using advanced land surface modeling and data assimilation techniques, in order to generate optimal fields of land surface states and fluxes [60]. Up to now, there have been four land surface models (LSM), namely Mosaic [61], Community Land Model (CLM) [62], Noah [63], and Variable Infiltration Capacity (VIC) [64], respectively. The Noah model is selected to provide some water-budget variables, including runoff (R), soil moisture (SM), and snow water equivalent (SWE). The runoff includes surface runoff and underground runoff, and soil moisture is the sum of four soil water layers. The datasets from the Noah model include two kinds of temporal resolutions (3-h and monthly scale) and spatial resolutions ($0.25° \times 0.25°$ and $1° \times 1°$). Monthly datasets are chosen in this study, with the resolutions of $0.25°$ and $1°$, which are available at [65].

2.2.4. GLEAM Product

GLEAM (Global Land Evaporation Amsterdam Model) is a set of algorithms dedicated to estimating global evapotranspiration by combining satellite observations and the Priestley and Taylor equation [66]. It has been continuously revised and updated since 2011, and the third version of the model was released in 2017 [67]. The latest version GLEAM v3.3 contains two kinds of data sets (v3.3a and v3.3b), differing in their forcing and temporal coverage [68]. In this study, the GLEAM v3.3a is employed to provide monthly estimates in evapotranspiration, spanning the 36 years from 1980 to 2018.

2.2.5. Groundwater Level

Monthly groundwater level, collected from the Haihe River Basin Water Resources Bulletin [69], is used to validate the accuracy of the downscaled TWSA and predicted water level. Groundwater monitoring wells are unevenly distributed across the NCP, and there are abnormal jumps, data gaps, and outliers in some wells. Therefore, the collected data require pre-processing as follows: (1) ignore the wells with more missing months and obvious errors; (2) aggregate the groundwater level to pixels values on $0.25°$ cells by using the simple average of groundwater observations within the pixel; (3) remove the

mean value to obtain the groundwater level anomalies data. A total of 18 wells are selected for this study (Figure 1), and 12 wells are used to train the downscaled model, while the remaining 6 wells are used to test the performance of models.

3. Methods

3.1. Gradient Boosting Decision Tree

The gradient boosting decision tree is an algorithm that combines a series of weak learners into one strong learner [70]. Although the GBDT algorithm can be used for both classification and regression tasks, we only consider the latter in this study. Gradient boosting of regression trees can produce competitive, highly robust, interpretable procedures for all models, especially appropriate for mining less than clean data [70]. Different from other traditional regression methods, the GBDT algorithm obtains the global convergence by following the direction of the negative gradient, which will decrease the running time for getting the results [71]. The calculation core of GBDT is to learn and forecast by continually passing the residual sum of the conclusions of all the previous decision trees until the sum of the predicted values, and the input target residuals are minimized. Consequently, this study tries to develop the downscaling and prediction models based on the regression function of the ensemble algorithm.

3.2. Downscaling Approach Based on the Noah Model

The GRACE-derived TWSA includes plenty of water storage information (e.g., groundwater storage, soil moisture, and snow water equivalent), and part of these variables can also be simulated by the GLDAS-Noah model. Consequently, we can obtain two kinds of TWSA products from GRACE missions and the Noah model, respectively. Some discrepancies will exist in these products due to the absence of groundwater storage and anthropogenic factor in the Noah model. In order to obtain finer-resolution and higher-accuracy TWSA products, a regression model (called GRACE-Noah model for short) is employed to downscale GRACE data in this study, which can incorporate GRACE products and the Noah model [32,72]. In the GRACE-Noah model, the simulated TWSA is treated as "truth", and the bias can be calculated by the following formula:

$$B = TWSA^1_{Noah,i} - TWSA^1_{GRACE,i}, \tag{1}$$

where B is the bias of two kinds of TWSA products on the 1° grid; $TWSA^1_{Noah,i}$ is the normalized TWSA simulated by the Noah model; $TWSA^1_{GRACE,i}$ is the normalized TWSA derived from GRACE. The subscript 1 represents the 1 degree. Then, the GRACE products can be downscaled from 1° to 0.25° by the formula:

$$TWSA^{0.25}_{end,i} = TWSA^{0.25}_{Noah,i} - \frac{B \times A \times TWSA^{0.25}_{Noah\_pre,i}}{\sum\left(TWSA^{0.25}_{Noah\_pre,i} \times a_i\right)}, \tag{2}$$

where $TWSA^{0.25}_{end,i}$ is the downscaled GRACE TWSA with the 0.25° spatial resolution, $TWSA^{0.25}_{Noah,i}$ is the normalized TWSA simulated form the Noah model, B is the bias obtained by Formula (1), A is the area of the 1° grid (m²), $TWSA^{0.25}_{Noah\_pre,i}$ is the pre-normalized TWSA simulated from the Noah model, a_i is the area of the 0.25° grid (m²).

3.3. Multiple Linear Regression

Multiple Linear Regression (MLR) is a regression modeling method with multiple independent and dependent variables [73,74]. The essential parameter estimation method is the least squares method, which is used mainly to find the best function by minimizing the sum of squares of errors. Compared to the single regression model, the multiple linear regression model is more practical and accurate in simulating the relationship between independent and dependent variables, which can better achieve the prediction

and estimation of TWSA products. The following formula can describe the MLR method used in this study:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \tag{3}$$

where y is the dependent variable (TWSA), $x_1, x_2, \cdots, x_n$ represent the independent variables (e.g., P, R, and ET), a_0 is the constant value, $a_1, a_2, \cdots, a_n$ are the weights of n variables.

3.4. Fusion Model Design

To obtain high-resolution and high-accuracy GWLA, the machine learning-based fusion model is developed within the NCP, which mainly consists of three modules. Specifically, Module #1 is used to downscale GRACE-derived TWSA from 1° to 0.25° by using different algorithms (MLR, GBDT, and GRACE-Noah), which is also called the downscaling module. Module #2 is employed to incorporate climate variables with in-situ levels based on the GBDT algorithm, which is also named as the data fusion module. With respect to Module #3, it accepts the downscaled TWSA from Module #1 and the fused GWLA from Module #2. Then, these variables are integrated into a big model for obtaining GWLA in the whole study area, and as module is named the prediction module. The conceptual map of the fusion model is shown in Figure 4.
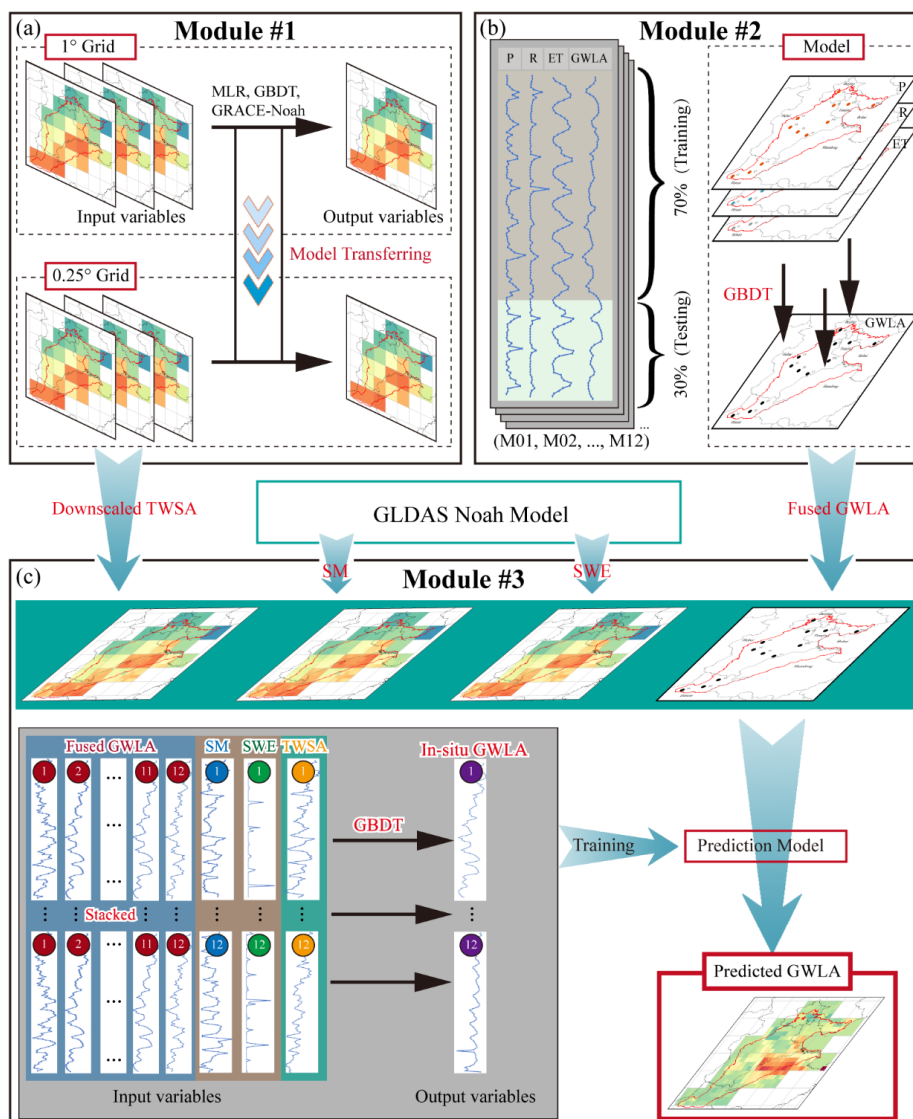
Figure 4. The conceptual map of the machine learning-based fusion model. (**a**) Downscaling module; (**b**) Data fusion module; (**c**) Prediction module.

3.4.1. Module #1 for Downscaling

This study utilizes two traditional machine learning models (MLR and GBDT) and one hydrological model (GLDAS-Noah) to downscale GRACE-derived TWSA into the higher resolution, as shown in Figure 4a. The detailed process of this module is described as follows:

Step 1: Climate variables (P, R, and ET) are chosen as the predictor based on the water balance equation, and GRACE-derived TWSA is selected as the predictand. The climate variables are resampled from $0.25°$ to $1°$ in accordance with that of GRACE products.

Step 2: Under the coarse resolution, these downscaling models are developed for each $1°$ grid based on the input variables (P, R, and ET) and the output variable (TWSA). Each model is continuously trained and tested by adjusting the core parameters until it can achieve satisfactory metrics (e.g., root mean square error (RMSE) [75], Nash-Sutcliffe efficiency coefficient (NSE) [76,77], and correlation coefficient (CC) [78]), and the mean absolute error (MAE) [79].

Step 3: The relationship is constructed between predictors and predictands at the resolution of $1°$. It is assumed that the relationship is still accurate under different spatial resolutions. The downscaled TWSA can be obtained in the study area by employing higher-resolution variables into these models.

Step 4: The bias of simulated TWSA and GRACE-derived TWSA is calculated at the resolution of $1°$ for each grid. Then, these values are resampled to 0.25 based on the Kriging interpolation [80] and assigned to the corresponding grid.

Step 5: GWSA can be isolated from the downscaled TWSA based on the auxiliary information provided by the Noah model. Then, the downscaling performance of the machine learning (MLR and GBDT) can be validate by comparing the downscaled GWSA with the groundwater measurements.

3.4.2. Module #2 for Data Fusion

The data fusion model is proposed to incorporate some climate variables (P, R, and ET) information with the in-situ well measurements, which are utilized to control other grids in the Module #3. Monthly groundwater level data of 18 observation wells are collected from the Haihe River Basin Water Resources Bulletin. In this module, we select 12 wells as the control wells based on the distribution of shadow groundwater wells. As for each well, the corresponding shuffled dataset, used in the data fusion model, is divided into two parts for training (70%) and testing (30%). At last, we determine 12 most reasonable models by continuously adjusting the structure of the GBDT model based on these metrics (e.g., RMSE, NSE, and CC). The model design flowchart is shown in Figure 4b (Module #2).

3.4.3. Module #3 for Prediction

To get high-quality GWLA products, the prediction module is developed by using the results from the first two modules and other climate variables (e.g., SM, SWE, and in-situ measurements), as shown in Figure 4c (Module #3). In the module, 12 observation wells are selected as the training wells, and the remaining 6 wells are used to evaluate the performance of the prediction model. In order to construct the prediction model, we select 15 characteristics as the predictors, including four kinds of variables. As for each training well, we construct a sub-dataset, which includes 12 fixed variables (12 fused GWLA) and 3 changed water-budget variables (SM, SWE, and TWSA). Then, 12 sub-datasets are stacked into a training matrix with the dimensions of 1440×15. By continuous adjusting and training, an ideal prediction model is developed in the NCP. It assumes that the model is applicable within all study areas; we can obtain all GWLA by constructing and employing sub-datasets into the model for each grid.

3.5. Model Evaluation and Data Analysis Standards

In order to evaluate the performance of the fusion model, four indices are used as the evaluation criteria, namely, RMSE, MAE, NSE and CC. The specific expressions are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i)^2}, \tag{4}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - X_i|, \tag{5}$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}, \tag{6}$$

$$\text{CC} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}, \tag{7}$$

where X_i and Y_i represent two independent datasets with the mean values of $\overline{X}$ and $\overline{Y}$, X_i represents the simulated value, Y_i represents the measured value, n means the total number of samples. As for the RMSE and MAE, the smaller the values are, the higher the accuracy of the model. Similarly, the closer the values of CC and NSE are to 1, the more consistent the simulated and measured values are.
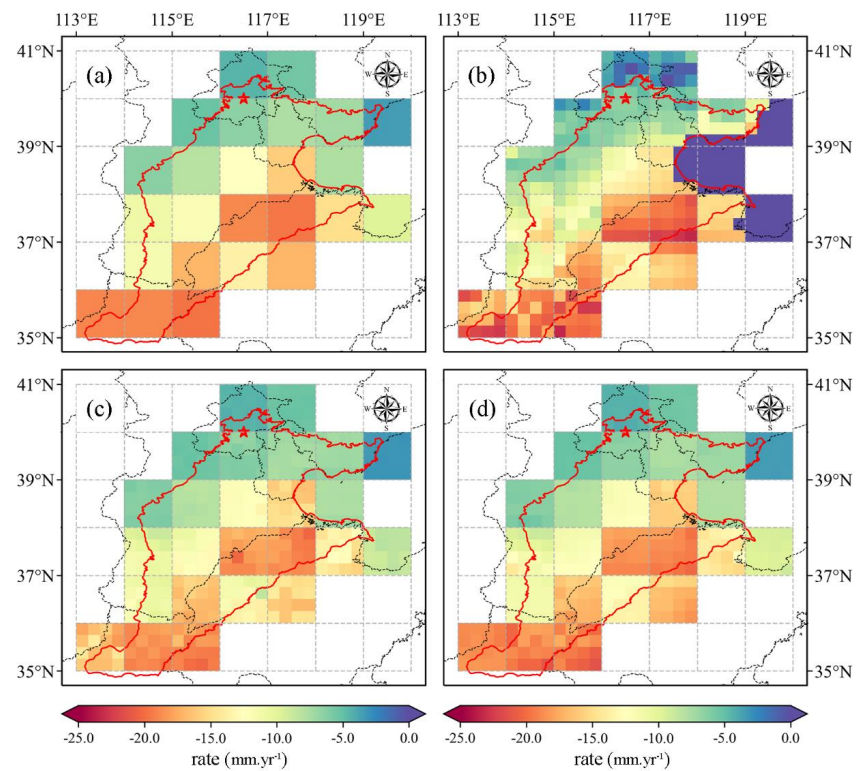
4. Results

4.1. Evaluation of Downscaling Models

The performance of three downscaling models is evaluated from two perspectives, which are spatial and temporal resolution. The detailed descriptions and results are as follows:
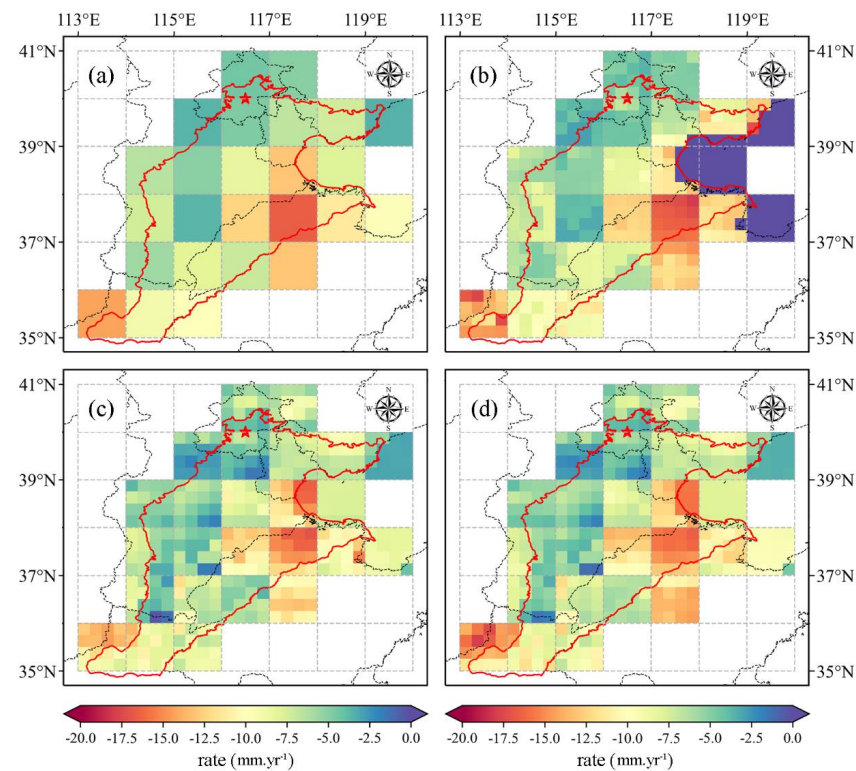
4.1.1. Spatial Resolution

Figure 5 shows the long-term trends of GRACE-derived TWSA and three downscaled results in the NCP. It can be found that the spatial distribution characteristics of TWSA are basically consistent with that of downscaled results. In general, the downscaled results can capture the sub-grid heterogeneity, while preserving the TWSA characteristic at the original scale. An obvious downtrend is observed in the southwestern region of the NCP, which is located in the conjunction area of Hebei and Henan province. This is mainly caused by intensive agricultural activities, which is in accordance with previous studies [16,81]. Furthermore, the downtrend becomes serious from the northern to the southern parts, with the trend of −4.96 mm/yr and −18.87 mm/yr, respectively. Additionally, there are some outlier values near the Bohai region in the GRACE-Noah model. The possible reason is that there are large uncertainties in the forcing data of the GLDAS-Noah model.

Additionally, the long-term trends of GRACE-based and downscaled GWSA are shown in Figure 6 during the period from 2005 to 2015. The larger decreasing trend is also detected in the Southern regions, which is consistent with that of TWSA. It is worth noting that the range of trends varies from −16.61 mm/yr to −1.41 mm/yr, which is only a bit smaller than that of TWSA (from −29.70 mm/yr to −2.08 mm/yr), indicating that the slope of TWSA is mainly caused by GWSA estimates. What is more, we can find that the East Central Plain is the most serious region, and it may be due to the over-explorati6on of deep groundwater storage [82].

Figure 5. Spatial distribution of TWSA trends from 2005 to 2015. (**a**) GRACE-derived TWSA, (**b**) downscaled results based on GRACE-Noah, (**c**) downscaled results based on GBDT, and (**d**) downscaled results based on MLR.


Figure 6. Spatial distribution of the GWSA trend from 2005 to 2015. (**a**) GRACE -derived TWSA, (**b**) downscaled results based on GRACE-Noah, (**c**) downscaled results based on GBDT, and (**d**) downscaled results based on MLR.

4.1.2. Temporal Resolution

In order to evaluate the downscaled results more intuitively, the time series of TWSA and GWSA are plotted in Figure 7. Similar decreasing trends can be observed in TWSA and GWSA from 2005 to 2015, with the trend of −9.89 mm/yr and −8.45 mm/yr, respectively. With respect to GWSA, the downtrend intensifies with the slope increasing from −5.94 mm/yr to −10.21 mm/yr. The downscaled water storage estimates based on the GLDAS-Noah model are well correlated with the results at the original resolution, with the correlation up to 0.99, and the acceptable RMSE value of 1.49 mm (Table 1). The worse performance can be found in the MLR model, and the possible reason is that there is some information missing during the process of downscaling. Based on these above discussions, we choose the downscaled results of GRACE-Noah as representative values, thus used in the following discussions.
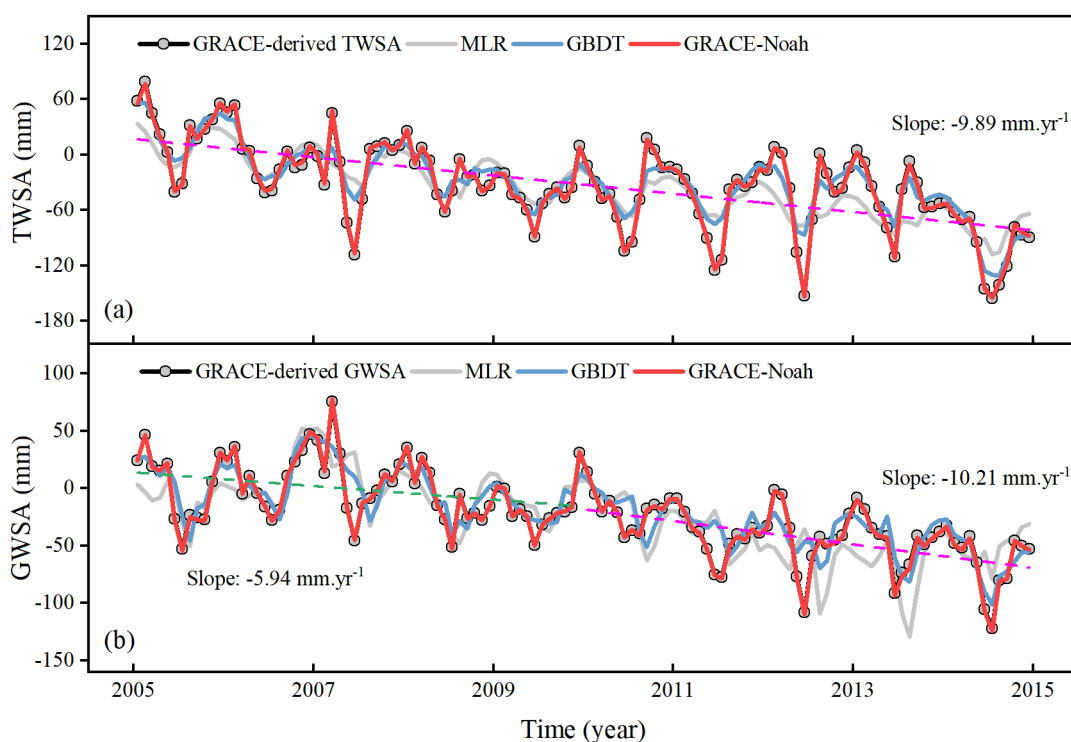
Figure 7. Time-series trends of (**a**) TWSA and (**b**) GWSA before and after downscaling in the whole NCP. The black line shows the time-series trend of GRACE-derived products, while grey, blue, and red lines repBresent downscaled results (MLR, GBDT, and GRACE-Noah).

Table 1. Performance of downscaling models (RMSE, MAE, NSE, and CC).

|	Models	RMSE (mm)	MAE (mm)	NSE	CC
	GRACE-Noah	1.49	1.17	0.99	0.99
TWSA	GBDT	18.00	10.20	0.85	0.93
	MLR	28.32	16.84	0.67	0.79
	GRACE-Noah	1.24	0.81	0.99	0.99
GWSA	GBDT	17.08	9.78	0.75	0.87
	MLR	27.23	15.81	0.36	0.68

4.2. Results of Data Fusion

Based on previous studies, the GBDT algorithm is employed to construct the regression model with the advantages of robustness, efficiency, and simplicity [42,70]. Specifically, the GBDT model is designed for 12 in-situ wells, which are selected based on their spatial

distribution and data quality. Each model is developed for incorporating some climate variables (P, R, and ET) into in-situ observations.

The performance of models and time-series comparison results before and after fusion are shown in Table 2 and Figure 8, respectively. It can be found in Table 2 that 12 data fusion models show good performances with the average RMSE, MAE, NSE, and CC values of 1.10 m, 0.87 m, 0.91, and 0.97, respectively.

Table 2. Performances of 12 GBDT models (RMSE, MAE, NSE, and CC).

Model	Grid	RMSE (m)	MAE (m)	NSE	CC
M01	T1	0.72	0.59	0.85	0.95
M02	T2	0.55	0.44	0.91	0.97
M03	T3	0.23	0.18	0.93	0.98
M04	T4	0.85	0.68	0.90	0.97
M05	T5	0.74	0.57	0.90	0.96
M06	T6	1.58	1.20	0.94	0.98
M07	T7	3.03	2.38	0.94	0.98
M08	T8	1.40	1.16	0.87	0.96
M09	T9	0.61	0.47	0.91	0.97
M10	T10	1.26	1.05	0.95	0.97
M11	T11	0.72	0.58	0.88	0.96
M12	T12	1.51	1.19	0.88	0.96
Mean		1.10	0.87	0.91	0.97
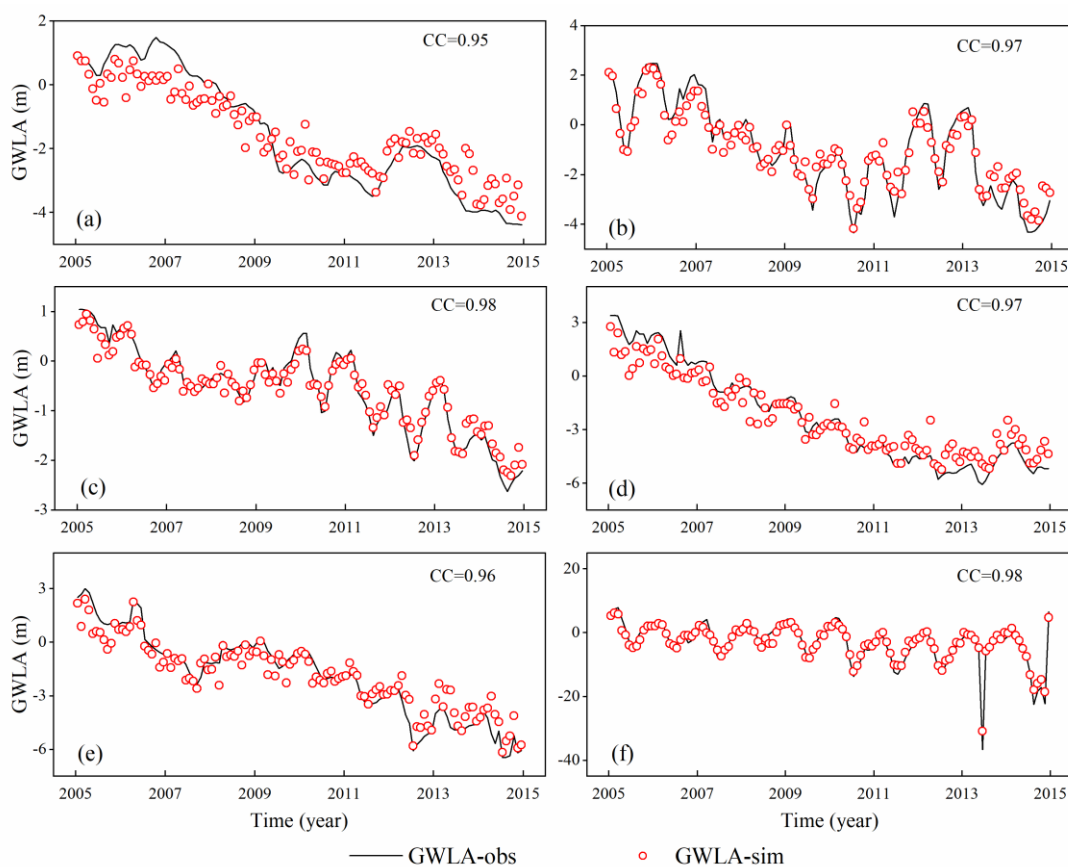
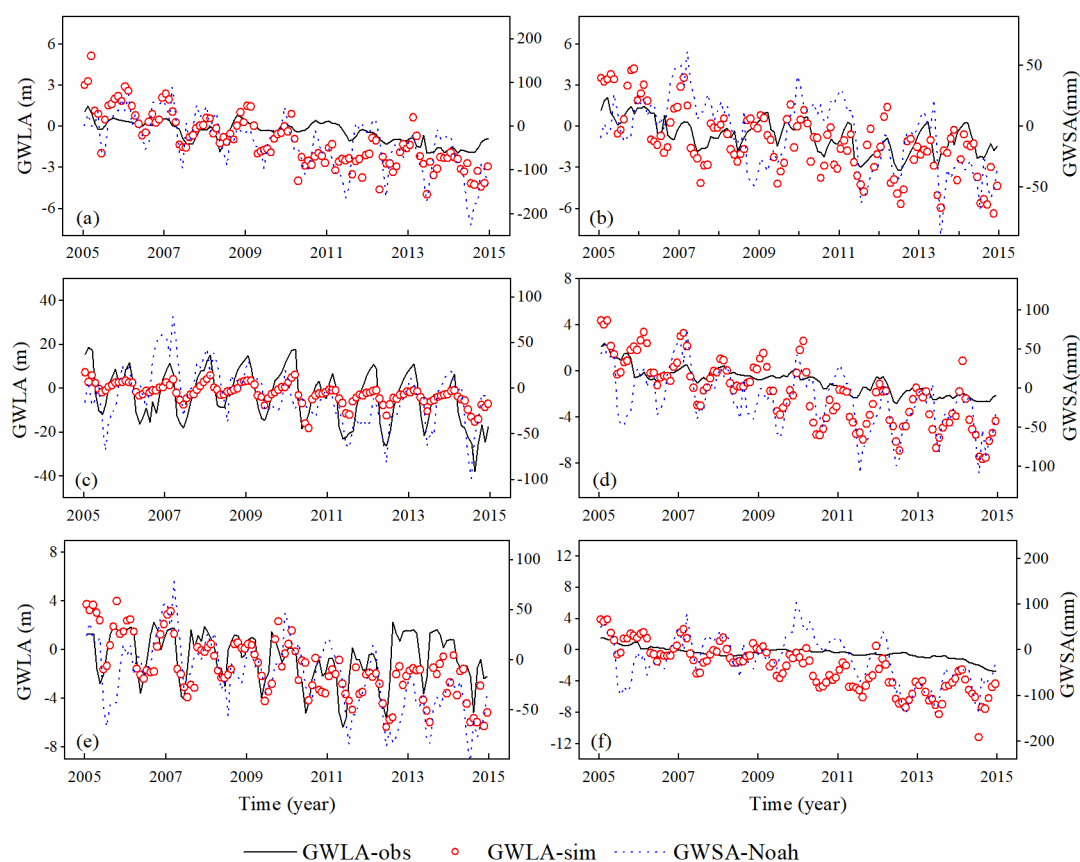**Figure 8.** Time-series comparison of in-situ and simulated GWLA data. (**a**–**f**) represent the corresponding wells (T1–T6).

Six monitoring wells are utilized to evaluate the applicability of the fusion model from the time series trend, and the verification results indicate that all of them perform ideal CC values (0.95, 0.97, 0.98, 0.97, 0.96, and 0.98). Then, the fused GWLA data of 12 wells are

regarded as the control wells and used as the input variables of the prediction model in the Module #3.

4.3. Prediction Performance Analysis

Based on the downscaled TWSA and 12 control wells, the prediction model is developed to forecast the groundwater level at the 0.25° pixel scale. The remaining six wells are utilized to evaluate the accuracy of the prediction model, as shown in Figure 9. Results reveal that the predicted GWLA is reasonable in the first five wells but worse in the last one. This is attributed to the fact that the P6 well is close to the Bohai region with poor quality of in-situ measurements. On the whole, there will be a certain deviation between the predicted value and the in-situ value of all wells, but the overall trend is basically the same.


Figure 9. Comparison of test wells data before and after the prediction model. (**a**–**f**) represent the corresponding wells (P1-P6). GWLA-obs, GWLA-sim, and GWSA-Noah represent the in-situ GWLA, simulated GWLA, and GRACE-derived GWSA, respectively.

4.4. Verification of In-Situ Observations

To further explore the applicability of the machine learning-based fusion model, 18 monitoring wells are collected to evaluate the simulated results, and the comparison is shown in Table 3 and Figure 10. The verification includes two parts, which are the verification of downscaled results and the verification of predicted results.

Table 3. Performances of the machine learning-based fusion model (CC).

	Downscaled Results			Predicted Results
Wells	**MLR**	**GBDT**	**GRACE-Noah**	**GBDT-Pre**
T1	0.60	0.56	0.78	0.82
T2	0.35	0.74	0.75	0.78
T3	0.43	0.63	0.65	0.78
T4	0.67	0.73	0.75	0.87
T5	0.51	0.55	0.69	0.87
T6	0.26	0.42	0.56	0.93
T7	0.22	0.52	0.67	0.95
T8	0.75	0.65	0.59	0.92
T9	0.38	0.45	0.61	0.83
T10	−0.46	0.30	0.41	0.79
T11	0.42	0.45	0.55	0.83
T12	0.43	0.38	0.41	0.83
P1	0.34	0.52	0.55	0.69
P2	0.33	0.49	0.38	0.76
P3	0.09	0.37	0.46	0.83
P4	0.76	0.55	0.58	0.81
P5	−0.07	0.06	0.24	0.50
P6	0.38	0.46	0.36	0.67
Mean	0.36	0.49	0.56	0.80
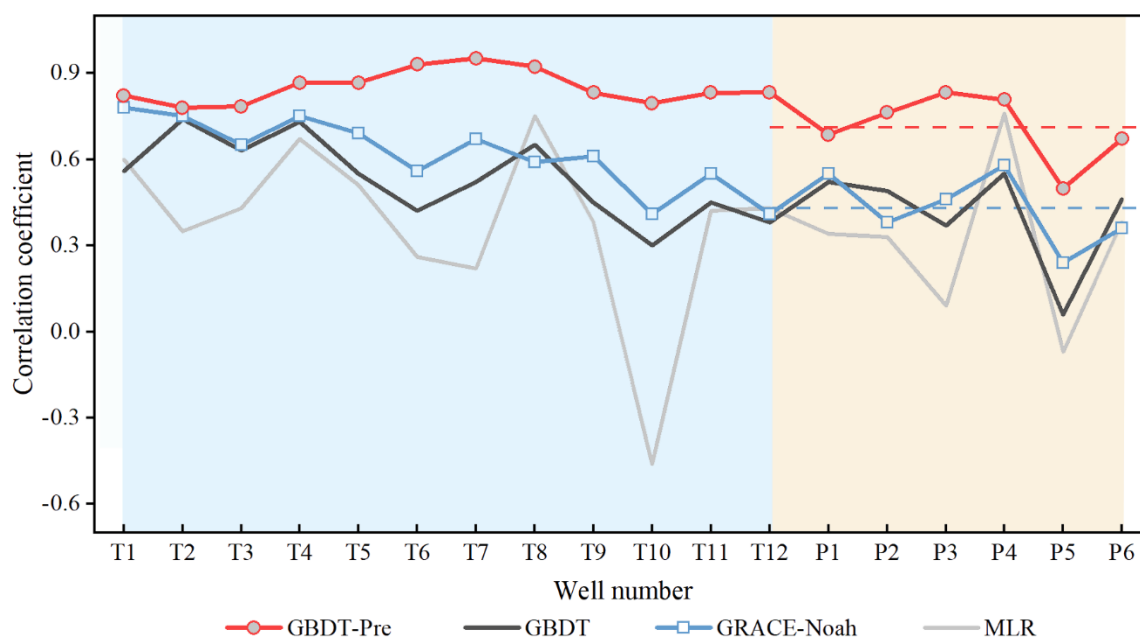
Figure 10. Correlation coefficient comparison before and after prediction. The blue dashed line and the red dashed line show the mean CC values of all wells before and after modeling, respectively.

As for downscaled results, all of 18 total wells are utilized to evaluate the performance of models, and three downscaling models present reasonable results with the mean CC values of 0.36 (MLR), 0.49 (GBDT), and 0.56 (GRACE-Noah), respectively. Although the MLR and GRR models may show better performance in several wells, such as P2, P4, and P6 wells, the values are close to the GRACE-Noah model. Moreover, other wells show obviously better performances in the GRACE-Noah model, especially in the T10 well with the CC values of −0.46 (MLR), 0.30 (GBDT), and 0.41 (GRACE-Noah). Consequently, the GRACE-Noah model is considered to be the optimum downscaling model in this study, followed by the GBDT model, which is consistent with the result Section 4.1. Therefore, the

groundwater estimates based on the GRACE-Noah model are used to compare with the predicted products.

As for the predicted results, the CC values before and after prediction are compared in Figure 10. The light blue areas on the left represent the results of 12 control wells, while the light yellow areas stand for the results of 6 predicted wells. It can be seen that the mean CC between GRACE-derived GWSA (GRACE-Noah) and observed GWLA is 0.43, while the predicted results increase to 0.71 against downscaled values. Moreover, all of the CC values between the predicted results and in-situ GWLA are better than the downscaled results, especially in the P6 well, whose performance is higher than the expected result with the CC value of 0.67 (Figures 9f and 10). Overall, the prediction model presents an excellent performance in simulating the changing trend but may be insufficient in numerical prediction.

5. Discussion

5.1. Efficacy of the Fusion Model

Performance metrics from the downscaled and the predicted modules indicate that the machine learning-based fusion model can successfully achieve the purpose of downscaling GRACE-derived TWSA and predicting GWLA. In the Module #1, three kinds of methods are employed to downscale GRACE observations into 0.25°, which are GBDT, MLR, and GRACE-Noah, respectively. Results indicate that the GRACE-Noah model outperforms the other two models, especially in terms of temporal scales (Figure 7). The possible reason may be that the downscaled algorithm based on the Noah model can effectively assign the discrepancies between GRACE and hydrological model into the pixel at higher resolution, thereby preserving the integrity of climate information at the coarse resolution. With respect to the Module #2, it incorporates the information of variables (P, R, and ET) into GWLA based on the 12 models, which are built for each control well using the GBDT machine learning method. As shown in Table 3, each model reveals excellent performance with the CC values ranging from 0.95 to 0.98. In the Module #3, the downscaled TWSA from the Module #1 and fused GWLA from the Module #2 are taken into account in the prediction model. It can be seen in Figure 10 that the prediction model performs reasonably in simulating dynamic changes in GWLA, with the CC values ranging from 0.50 to 0.95. In general, the fusion model developed in this study present satisfactory performance in downscaling and prediction phases within the NCP.

5.2. Limitations and Outlook

Based on these above discussion, the developed fusion model can effectively downscale GRACE observations and predict high-quality GWLA at pixel scales. However, the highest resolution of predicted results is mainly determined by the resolution of climate variables (Module #1) and water storage estimates (Module #3). The water balance variables are the most widely used in previous studies [31,37], provided at the maximum resolution of 0.25°. Similarly, the water storage components are obtained from the GLDAS-Noah model, which provides simulated outputs at the resolutions of 1° and 0.25°. Consequently, the target resolution of downscaled is 0.25° in this study.

Although the fusion method performs promise in this study, we will make improvements from the following aspects in the future. On the one hand, only three kinds of climate variables are taken into account in the downscaling module. Theoretically, more input variables have the potential to improve the accuracy of downscaling results, such as temperature and normalized difference vegetation index. What is more, this study only selects 12 groundwater levels as the control well, restricted by the limited in-situ measurements. If more observed data are obtained, the fusion model may perform better than it does now. Of course, the performance of the fusion model is also limited the actual situation of different study areas.

6. Conclusions

Based on TWSA products derived from GRACE gravity satellites, fruitful results in research on groundwater levels have been achieved in large-scale areas. However, due to the coarse spatial resolution of GRACE observations, the ability to study the changes in groundwater levels is limited in small-scale areas. Consequently, this study conduct meaningful research on downscaling GRACE-derived TWSA and predicting high-quality GWLA based on the machine learning algorithms, and the results are summarized as follows:

(1) The machine learning-based fusion model, including three modules (downscaling module, data fusion module, and prediction module), is proposed in the NCP based on the empirical relationships between GRACE and climate drivers. These modules are both independent and integrated because the first two modules provide input variables for the prediction module while exhibiting their functions.

(2) GRACE-derived TWSA is downscaled from 1° to 0.25° by utilizing three downscaling models (MLR, GBDT, and GRACE-Noah models). From the spatial resolution and temporal resolution, we compare the performances of downscaling models, and the findings indicate that the GRACE-Noah model performs the best performance, with the CC value of 0.99 and RMSE value of 1.49 mm in the whole study area. What is more, the verification results with in-situ observations of 18 wells also indicate the same result, with acceptable CC values ranging from 0.24 to 0.78.

(3) Based on the downscaled and fused results, the prediction model is developed to obtain the GWLA within the whole NCP, and the verification results (CC values ranging from 0.50 to 0.95) indicate that the performance in simulating the long-term trend is ideal but may be insufficient in numerical prediction. Further, the average CC values of 6 test wells are calculated after prediction, which performs that the predicted result (0.71) is 65.12% higher than the downscaled result (0.43).

Overall, the proposed fusion model can effectively implement the downscaling of GRACE products and the prediction of high-accuracy GWLA in the NCP. To some extent, the fusion model can provide some suggestions to obtain and understand the dynamics of water resources for some areas with no or less in-situ measurements. However, the output spatial resolution and accuracy of the fusion model are limited by the climate variables and the water storage components. If higher-resolution and higher-precision climate variables can be obtained in the future, the fusion model may have the potential to obtain higher quality products (TWSA and GWLA).

Author Contributions: G.Z., W.Z., and W.Y. were responsible for the overall design of the study. G.Z. performed all the experiments and drafted the manuscript, which was revised by W.Z., W.Y., W.L.; G.Z., W.Z., W.Y., W.L. read and approved the final manuscript through review, editing, and comments. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. Aeschbach-Hertig, W.; Gleeson, T. Regional strategies for the accelerating global problem of groundwater depletion. *Nat. Geosci.* **2012**, *5*, 853–861. [CrossRef]
2. Gleeson, T.; Cuthbert, M.; Ferguson, G.; Perrone, D. Global groundwater sustainability, resources, and systems in the Anthropocene. *Annu. Rev. Earth Planet. Sci.* **2020**, *48*, 431–463. [CrossRef]
3. Morris, B.L.; Lawrence, A.R.L.; Chilton, P.J.C.; Adams, B.; Calow, R.C.; Klinck, B.A. *Groundwater and Its Susceptibility to Degradation: A Global Assessment of the Problem and Options for Management*; UNEP Early Warning & Assessment Report Series RS. 03-3; UNEP: Nairobi, Kenya, 2003.

4. Siebert, S.; Burke, J.; Faures, J.M.; Frenken, K.; Hoogeveen, J.; Döll, P.; Portmann, F.T. Groundwater use for irrigation—A global inventory. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 1863–1880. [CrossRef]

5. Taylor, R.G.; Scanlon, B.; Döll, P.; Rodell, M.; van Beek, R.; Wada, Y.; Longuevergne, L.; Leblanc, M.; Famiglietti, J.S.; Edmunds, M.; et al. Ground water and climate change. *Nat. Clim. Chang.* **2013**, *3*, 322–329. [CrossRef]

6. Wu, W.; Lo, M.; Wada, Y.; Famiglietti, J.S.; Reager, J.T.; Yeh, P.J.F.; Ducharne, A.; Yang, Z. Divergent effects of climate change on future groundwater availability in key mid-latitude aquifers. *Nat. Commun.* **2020**, *11*, 3710. [CrossRef]

7. Bierkens, M.F.P.; Wada, Y. Non-renewable groundwater use and groundwater depletion: A review. *Environ. Res. Lett.* **2019**, *14*, 063002. [CrossRef]

8. Lall, U.; Josset, L.; Russo, T. A snapshot of the world's groundwater challenges. *Annu. Rev. Environ. Resour.* **2020**, *45*, 1–24. [CrossRef]

9. Yin, W.; Li, T.; Zheng, W.; Hu, L.; Han, S.; Tangdamrongsub, N.; Šprlák, M.; Huang, Z. Improving regional groundwater storage estimates from GRACE and global hydrological models over Tasmania, Australia. *Hydrogeol. J.* **2020**, *28*, 1809–1825. [CrossRef]

10. Feng, W.; Shum, C.; Zhong, M.; Pan, Y. Groundwater storage changes in China from satellite gravity: An overview. *Remote Sens.* **2018**, *10*, 674. [CrossRef]

11. Zhang, M.; Hu, L.; Yao, L.; Yin, W. Surrogate Models for Sub-Region Groundwater Management in the Beijing Plain, China. *Water* **2017**, *9*, 766. [CrossRef]

12. Tapley, B.D.; Bettadpur, S.; Ries, J.C.; Thompson, P.F.; Watkins, M.M. GRACE measurements of mass variability in the earth system. *Science* **2004**, *305*, 503–505. [CrossRef] [PubMed]

13. Tapley, B.D.; Bettadpur, S.; Watkins, M.; Reigber, C. The gravity recovery and climate experiment: Mission overview and early results. *Geophys. Res. Lett.* **2004**, *31*, L09607. [CrossRef]

14. Swenson, S.; Wahr, J.; Milly, P.C.D. Estimated accuracies of regional water storage variations inferred from the Gravity Recovery and Climate Experiment (GRACE). *Water Resour. Res.* **2003**, *39*, W1223. [CrossRef]

15. Zhong, Y.; Zhong, M.; Mao, Y.; Ji, B. Evaluation of Evapotranspiration for Exorheic Catchments of China during the GRACE Era: From a Water Balance Perspective. *Remote Sens.* **2020**, *12*, 511. [CrossRef]

16. Yin, W.; Han, S.; Zheng, W.; Yeo, I.; Hu, L.; Tangdamrongsub, N.; Ghobadi-Far, K. Improved water storage estimates within the North China Plain by assimilating GRACE data into the CABLE model. *J. Hydrol.* **2020**, *590*, 125348. [CrossRef]

17. Rodell, M.; Velicogna, I.; Famiglietti, J.S. Satellite-based estimates of groundwater depletion in India. *Nature* **2009**, *460*, 999–1002. [CrossRef]

18. Yin, W.; Hu, L.; Jiao, J.J.; Lo Russo, S. Evaluation of groundwater storage variations in Northern China Using GRACE Data. *Geofluids* **2017**, *2017*, 8254824. [CrossRef]

19. Famiglietti, J.S.; Rodell, M. Water in the balance. *Science* **2013**, *340*, 1300–1301. [CrossRef]

20. Wilby, R.L.; Wigley, T.M.L. Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr Earth Environ.* **1997**, *21*, 530–548. [CrossRef]

21. Tang, J.; Niu, X.; Wang, S.; Gao, H.; Wang, X.; Wu, J. Statistical downscaling and dynamical downscaling of regional climate in China: Present climate evaluations and future climate projections. *J. Geophys. Res. Atmos.* **2016**, *121*, 2110–2129. [CrossRef]

22. Giorgi, F. Simulation of regional climate using a limited area model nested in a general circulation model. *J. Clim.* **1990**, *3*, 941–963. [CrossRef]

23. Xu, Z.; Han, Y.; Yang, Z. Dynamical downscaling of regional climate: A review of methods and limitations. *Sci. China Earth Sci.* **2018**, *62*, 365–375. [CrossRef]

24. Eicker, A.; Schumacher, M.; Kusche, J.; Döll, P.; Schmied, H.M. Calibration/Data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an Ensemble Kalman Filter: First results. *Surv. Geophys.* **2014**, *35*, 1285–1309. [CrossRef]

25. Forootan, E.; Rietbroek, R.; Kusche, J.; Sharifi, M.A.; Awange, J.L.; Schmidt, M.; Omondi, P.; Famiglietti, J. Separation of large scale water storage patterns over Iran using GRACE, altimetry and hydrological data. *Remote Sens. Environ.* **2014**, *140*, 580–595. [CrossRef]

26. Sahour, H.; Sultan, M.; Vazifedan, M.; Abdelmohsen, K.; Karki, S.; Yellich, J.A.; Gebremichael, E.; Alshehri, F.; Elbayoumi, T.M. Statistical applications to downscale GRACE-derived terrestrial water storage data and to fill temporal gaps. *Remote Sens.* **2020**, *12*, 533. [CrossRef]

27. Nie, W.; Zaitchik, B.F.; Rodell, M.; Kumar, S.V.; Arsenault, K.R.; Li, B.; Getirana, A. Assimilating GRACE into a land surface model in the presence of an irrigation-induced groundwater trend. *Water Resour. Res.* **2019**, *55*, 11274–11294. [CrossRef]

28. Khaki, M.; Schumacher, M.; Forootan, E.; Kuhn, M.; Awange, J.L.; van Dijk, A.I.J.M. Accounting for spatial correlation errors in the assimilation of GRACE into hydrological models through localization. *Adv. Water Resour.* **2017**, *108*, 99–112. [CrossRef]

29. Wilby, R.L.; Wigley, T.M.L.; Conway, D.; Jones, P.D.; Hewitson, B.C.; Main, J.; Wilks, D.S. Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.* **1998**, *34*, 2995–3008. [CrossRef]

30. Gutiérrez, J.M.; San-Martín, D.; Brands, S.; Manzanas, R.; Herrera, S. Reassessing statistical downscaling techniques for their robust application under climate change conditions. *J. Clim.* **2013**, *26*, 171–188. [CrossRef]

31. Ning, S.; Ishidaira, H.; Wang, J. Statistical downscaling of GRACE-derived terrestrial water storage using satellite and GLDAS products. *J. Jpn. Soc. Civ. Eng. Ser. B1 Hydraul. Eng.* **2014**, *70*, 133–138. [CrossRef]

32. Wan, Z.; Zhang, K.; Xue, X.; Hong, Z.; Hong, Y.; Gourley, J.J. Water balance-based actual evapotranspiration reconstruction from ground and satellite observations over the conterminous United States. *Water Resour. Res.* **2015**, *51*, 6485–6499. [CrossRef]
33. Yin, W.; Hu, L.; Zhang, M.; Wang, J.; Han, S. Statistical downscaling of GRACE-derived groundwater storage using ET data in the North China Plain. *J. Geophys. Res. Atmos.* **2018**, *123*, 5973–5987. [CrossRef]
34. Miro, M.; Famiglietti, J. Downscaling GRACE remote sensing datasets to high-resolution groundwater storage change maps of California's Central Valley. *Remote Sens.* **2018**, *10*, 143. [CrossRef]
35. Kumar, P.S.; Praveen, T.V.; Prasad, M.A. Artificial neural network model for rainfall-runoff—A case study. *Int. J. Hybrid Inf. Technol.* **2016**, *9*, 263–272. [CrossRef]
36. Seyoum, W.M.; Milewski, A.M. Improved methods for estimating local terrestrial water dynamics from GRACE in the Northern High Plains. *Adv. Water Resour.* **2017**, *110*, 279–290. [CrossRef]
37. Chen, L.; He, Q.; Liu, K.; Li, J.; Jing, C. Downscaling of GRACE-derived groundwater storage based on the random forest model. *Remote Sens.* **2019**, *11*, 2979. [CrossRef]
38. Rahaman, M.; Thakur, B.; Kalra, A.; Li, R.; Maheshwari, P. Estimating high-resolution groundwater storage from GRACE: A random forest approach. *Environments* **2019**, *6*, 63. [CrossRef]
39. Fan, J.; Yue, W.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol.* **2018**, *263*, 225–241. [CrossRef]
40. Wu, Z.; Zhou, Y.; Wang, H. Real-Time Prediction of the Water Accumulation Process of Urban Stormy Accumulation Points Based on Deep Learning. *IEEE Access* **2020**, *8*, 151938–151951. [CrossRef]
41. Wu, Z.; Zhou, Y.; Wang, H.; Jiang, Z. Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse. *Sci. Total Environ.* **2020**, *716*, 137077. [CrossRef]
42. Seyoum, W.; Kwon, D.; Milewski, A. Downscaling GRACE TWSA data into high-resolution groundwater level anomaly using machine learning-based models in a glacial aquifer system. *Remote Sens.* **2019**, *11*, 824. [CrossRef]
43. Wang, S.; Shao, J.; Song, X.; Zhang, Y.; Huo, Z.; Zhou, X. Application of MODFLOW and geographic information system to groundwater flow simulation in North China Plain, China. *Environ. Geol.* **2008**, *55*, 1449–1462. [CrossRef]
44. Liang, H.; Qin, W.; Hu, K.; Tao, H.; Li, B. Modelling groundwater level dynamics under different cropping systems and developing groundwater neutral systems in the North China Plain. *Agric. Water Manag.* **2019**, *213*, 732–741. [CrossRef]
45. Gong, H.; Pan, Y.; Zheng, L.; Li, X.; Zhu, L.; Zhang, C.; Huang, Z.; Li, Z.; Wang, H.; Zhou, C. Long-term groundwater storage changes and land subsidence development in the North China Plain (1971–2015). *Hydrogeol. J.* **2018**, *26*, 1417–1427. [CrossRef]
46. Zhang, C.; Duan, Q.; Yeh, P.J.F.; Pan, Y.; Gong, H.; Gong, W.; Di, Z.; Lei, X.; Liao, W.; Huang, Z.; et al. The effectiveness of the South-to-North Water Diversion Middle Route Project on water delivery and groundwater recovery in North China Plain. *Water Resour. Res.* **2020**, e2019WR026759. [CrossRef]
47. Shang, Q.; Liu, X.; Deng, X.; Zhang, B. Downscaling of GRACE datasets based on relevance vector machine using InSAR time series to generate maps of groundwater storage changes at local scale. *J. Appl. Remote Sens.* **2019**, *13*, 1–18. [CrossRef]
48. Zheng, L.; Pan, Y.; Gong, H.; Huang, Z.; Zhang, C. Comparing Groundwater Storage Changes in Two Main Grain Producing Areas in China: Implications for Sustainable Agricultural Water Resources Management. *Remote Sens.* **2020**, *12*, 2151. [CrossRef]
49. Wang, J.; Wang, E.; Yang, X.; Zhang, F.; Yin, H. Increased yield potential of wheat-maize cropping system in the North China Plain by climate change adaptation. *Clim. Chang.* **2012**, *113*, 825–840. [CrossRef]
50. Guo, H.; Zhang, Z.; Cheng, G.; Li, W.; Li, T.; Jiao, J.J. Groundwater-derived land subsidence in the North China Plain. *Environ. Earth Sci.* **2015**, *74*, 1415–1427. [CrossRef]
51. Liu, C.; Yu, J.; Eloise, K. Groundwater exploitation and its impact on the environment in the North China Plain. *Water Int.* **2001**, *26*, 265–272. [CrossRef]
52. Feng, W.; Zhong, M.; Lemoine, J.; Biancale, R.; Hsu, H.; Xia, J. Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements. *Water Resour. Res.* **2013**, *49*, 2110–2118. [CrossRef]
53. Sakura, Y.; Tang, C.; Yoshioka, R.; Ishiashi, H. *Intensive Use of Groundwater in Some Areas of China and Japan. Intensive Use of Groundwater: Challenges and Opportunities*; CRC Press: Boca Raton, FL, USA, 2003.
54. Feng, W. GRAMAT: A comprehensive Matlab toolbox for estimating global mass variations from GRACE satellite data. *Earth Sci. Inform.* **2019**, *12*, 389–404. [CrossRef]
55. Landerer, F.W.; Swenson, S.C. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.* **2012**, *48*, W4531. [CrossRef]
56. Monthly Mass Grids Land. Available online: https://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/ (accessed on 8 December 2020).
57. Sakumura, C.; Bettadpur, S.; Bruinsma, S. Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophys. Res. Lett.* **2014**, *41*, 1389–1397. [CrossRef]
58. Kummerow, C.; Simpson, J.; Thiele, O.; Barnes, W.; Chang, A.T.C.; Stocker, E.; Adler, R.F.; Hou, A.; Kakar, R.; Wentz, F.; et al. The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in Orbit. *J. Appl. Meteorol.* **2000**, *39*, 1965–1982. [CrossRef]
59. NASA Dataset. Available online: https://disc.gsfc.nasa.gov/datasets?keywords=TRMM (accessed on 8 December 2020).

60. Rodell, M.; Houser, P.R.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The global land data assimilation system. *B. Am. Meteorol. Soc.* **2004**, *85*, 381–394. [CrossRef]

61. Koster, R.D.; Suarez, M.J.; Heiser, M. Variance and Predictability of Precipitation at Seasonal-to-Interannual Timescales. *J. Hydrometeorol.* **2000**, *1*, 26–46. [CrossRef]

62. Dai, Y.; Zeng, X.; Dickinson, R.E.; Baker, I.; Bonan, G.B.; Bosilovich, M.G.; Denning, A.S.; Dirmeyer, P.A.; Houser, P.R.; Niu, G.; et al. The Common Land Model. *B. Am. Meteorol. Soc.* **2003**, *84*, 1013–1024. [CrossRef]

63. Chen, F.; Mitchell, K.; Schaake, J.; Xue, Y.; Pan, H.; Koren, V.; Duan, Q.Y.; Ek, M.; Betts, A. Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res. Atmos.* **1996**, *101*, 7251–7268. [CrossRef]

64. Liang, X.; Lettenmaier, D.P.; Wood, E.F.; Burges, S.J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* **1994**, *99*, 14415–14428. [CrossRef]

65. NASA Dataset. Available online: https://disc.gsfc.nasa.gov/datasets?keywords=GLDAS (accessed on 8 December 2020).

66. Miralles, D.G.; Holmes, T.R.H.; De Jeu, R.A.M.; Gash, J.H.; Meesters, A.G.C.A.; Dolman, A.J. Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 453–469. [CrossRef]

67. Martens, B.; Miralles, D.G.; Lievens, H.; van der Schalie, R.; de Jeu, R.A.M.; Fernández-Prieto, D.; Beck, H.E.; Dorigo, W.A.; Verhoest, N.E.C. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* **2017**, *10*, 1903–1925. [CrossRef]

68. Gleam Dataset. Available online: https://www.gleam.eu/#datasets (accessed on 8 December 2020).

69. HWCC. Available online: http://www.hwcc.gov.cn/hwcc/wwgj/xxgb/szygb/ (accessed on 8 December 2020).

70. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

71. Rao, H.; Shi, X.; Rodrigue, A.K.; Feng, J.; Xia, Y.; Elhoseny, M.; Yuan, X.; Gu, L. Feature selection based on artificial bee colony and gradient boosting decision tree. *Appl. Soft Comput.* **2019**, *74*, 634–642. [CrossRef]

72. Zhang, D.; Liu, X.; Bai, P. Assessment of hydrological drought and its recovery time for eight tributaries of the Yangtze River (China) based on downscaled GRACE data. *J. Hydrol.* **2019**, *568*, 592–603. [CrossRef]

73. Aiken, L.S.; West, S.G.; Pitts, S.C.; Baraldi, A.N.; Wurpts, I.C. *Multiple Linear Regression*; Handbook of Psychology; American Cancer Society: Atlanta, GA, USA, 2012; pp. 511–542.

74. Breiman, L.; Friedman, J.H. Predicting multivariate responses in multiple linear regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1997**, *59*, 3–54. [CrossRef]

75. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]

76. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models Part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]

77. McCuen, R.H.; Knight, Z.; Cutter, A.G. Evaluation of the Nash–Sutcliffe efficiency index. *J. Hydrol. Eng.* **2006**, *11*, 597–602. [CrossRef]

78. Rodgers, J.L.; Nicewander, W.A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **1988**, *42*, 59–66. [CrossRef]

79. 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]

80. Oliver, M.A.; Webster, R. Kriging: A method of interpolation for geographical information systems. *Int. J. Geogr. Inf. Syst.* **1990**, *4*, 313–332. [CrossRef]

81. Zhao, Q.; Zhang, B.; Yao, Y.; Wu, W.; Meng, G.; Chen, Q. Geodetic and hydrological measurements reveal the recent acceleration of groundwater depletion in North China Plain. *J. Hydrol.* **2019**, *575*, 1065–1072. [CrossRef]

82. Huang, Z.; Pan, Y.; Gong, H.; Yeh, P.J.F.; Li, X.; Zhou, D.; Zhao, W. Subregional-scale groundwater depletion detected by GRACE for both shallow and deep aquifers in North China Plain. *Geophys. Res. Lett.* **2015**, *42*, 1791–1799. [CrossRef]