Research into the Optimal Regulation of the Groundwater Table and Quality in the Southern Plain of Beijing Using Geographic Information Systems Data and Machine Learning Algorithms

Chen Li 1, Baohui Men 1, Shiyang Yin 1,*, Teng Zhang 1 and Ling Wei 2

1 College of Water Resources and Hydropower Engineering, North China Electric Power University, Beijing 102206, China
2 School of Water Resources and Environment, China University of Geosciences (Beijing), Beijing 100083, China
* Correspondence: yinshiyang@ncepu.edu.cn

Title: The purpose of this paper is to provide new ideas and methods for the sustainable use of groundwater in areas with serious groundwater overexploitation and serious groundwater pollution. Geographic information systems (GIS) were combined with machine learning algorithms, water resources optimization technology, and groundwater numerical simulation to optimize the regulation of the groundwater table and quality beneath the Daxing District in the southern plain of Beijing. By collecting local consumption and supply data and observations of the groundwater table and quality in the connected aquifer beneath Daxing for the years 2006–2020, the corresponding water demands and groundwater impact were extrapolated for the years 2021–2025 based on the existing development model. Through the combination of GIS and machine learning algorithms, the NO₃-N concentration of local groundwater monitoring points in wet years, normal years, and dry years were predicted. With respect to NO₃-N pollution, three new groundwater exploitation regimes were devised, which we numbered 1 to 3. The optimal allocation of water resources was then calculated for wet year, typical year, and dry year scenarios for the year 2025. By comparing the water shortage, groundwater utilization rate, and NO₃-N pollution under the new groundwater exploitation regimes, the optimal groundwater exploitation mode for the three different types of hydrological year was determined. The results indicate that NO₃-N pollution was greatly reduced after the adoption of the optimal regimes and that the groundwater table demonstrated rapid recovery. These results can be of great help in realizing the management, supervision, and regulation of groundwater by combining GIS with machine learning algorithms.

Keywords: GIS; machine algorithm; optimal allocation of water resources; groundwater pollution; groundwater exploitation; China

1. Introduction

With the rapid growth of urban populations in recent years, urban water consumption has increased rapidly [1,2]. In northern China, where surface water is limited, urban water supplies are mostly sourced from groundwater. Growing urban water consumption in northern China has led to the over-exploitation of local groundwater resources, and this in turn has led to problems such as a continuous decline of the water table, land subsidence, and deterioration of groundwater quality [3,4]. To address the environmental problems caused by long-term groundwater overexploitation in the north, the Chinese government has launched the South-to-North Water Diversion Project.

Located in the northern part of the North China Plain, the nation’s capital city of Beijing is one of the cities affected by groundwater overexploitation. Since 2014, South-to-North water transfer to Beijing has replaced some groundwater, but groundwater still accounts for 30–40% of Beijing’s total water supply. Moreover, a considerable mass of pollutants is discharged into surface water and from there into the groundwater, a fact
which magnifies the problem of water supply security [5]. Therefore, how best to realize the sustainable utilization of groundwater resources while ensuring economic and social development has become an urgent problem to be solved.

GIS (Geographic Information Systems) is a technology that is widely used to collect, store, manage, analyze, and express spatial data [6]. GIS spans the intellectual realms of computer science, surveying, cartography, and geography, and has a wide range of applications. Not only does GIS play an important role in environmental resource management and planning, it has become an important tool in other fields such as urban management, engineering construction, commercial decision-making, and strategic analysis for the military, and has therefore become an essential working system for many institutions [7–9]. As a part of the global hydrological cycle, groundwater has obvious geographical distribution characteristics. As GIS has powerful geospatial analysis capabilities, it can be combined with machine prediction models, water resource optimization technology, and groundwater numerical simulation to achieve optimal regulation of the groundwater table and quality.

Water quality prediction is a basic function of water supervision and management [10]. At present, water quality prediction methods mainly involve time series prediction [11], gray-box prediction [12], fuzzy logic prediction [13,14], solute transport simulation [15,16], and machine learning algorithms [17,18]. While the time series method has a relatively sound theoretical basis, only the temporal response of the predicted water quality index is considered [19]. Water quality prediction is a very complex problem, and there are many factors affecting the prediction of water quality parameters. The accuracy of prediction in the future is poor only due to changes in the prediction index itself. The gray-box prediction method is easy to operate and suitable for situations where water quality monitoring data is incomplete, but it is easily affected by unstable data, resulting in a large prediction error in such cases [20]. Fuzzy logic prediction can handle uncertainty in the process of water quality prediction, although the calculation effort is large [21]. While the finite difference method can be used to simulate solute transport by flowing groundwater in the aquifer, the boundary conditions are not easy to determine [22,23].

The machine learning algorithm approach has a clear meaning, and adds artificial thinking and judgment along with high precision [24]. At present, machine learning algorithms mainly include neural networks and statistical learning methods. Commonly used neural network algorithms include back-propagation neural networks [25], radial basis function neural networks [26], recurrent neural networks [27], and long short-term memory neural networks (LSTMs) [28–30]. Statistical learning methods include decision trees [31,32], random forests (RFs) [33,34], and support vector machines (SVMs) [35,36]. With such a variety of powerful techniques on offer, effective groundwater quality prediction and control can be realized by combining GIS data with machine learning algorithms.

In this paper, Daxing District in the southern plain of Beijing is adopted as a case study. Thiessen polygons were generated around groundwater quality monitoring points. A GIS regional analysis tool was then used to extract the water consumption and economic development of each township. Finally, the groundwater quality in the study area from 2021 to 2025 was predicted using machine learning algorithms. According to the obtained results, the groundwater quality and the water table in the study area can be regulated by optimizing the allocation of water resources and by numerical simulation of the aquifer. The research methods in this paper can provide a theoretical foundation for the sustainable utilization of water resources in areas where groundwater is the main water source.

2. Materials and Methods

2.1. Study Area

Daxing District is located in the southern plain of Beijing, in a range of map coordinates that spans 116°13′–116°43′ E and 39°26′–39°51′ N. The district is divided into fourteen townships (Figure 1), and the total area of the district is approximately 1036 km². This part of China has a warm temperate semi-humid continental monsoon climate with four distinct seasons. The average annual rainfall is 510.1 mm; the rainfall is unevenly distributed within
the year and varies greatly from year to year. The average annual temperature is 11.7 °C, and the maximum frozen soil depth is 69 cm. The district is on the alluvial–proluvial plain of the Yongding River. Its terrain is flat, with an elevation of 9–73 m (the elevation reference system is the 1956 Yellow Sea Elevation System of China) and a topographic gradient of 0.5–2.0‰. Its soil type is mainly sandy loam, which gradually changes from coarse to fine as one travels from west to east. The thickness of the groundwater aquifer beneath is 80–120 m (Figure 2). The aquifer is mainly composed of medium and fine sand layers of the quaternary loose layer. Because the aquifer is generally permeable, the groundwater table is uniform. There are four main rivers in the study area, namely, the Xinfeng River, the Fenghe River, the Xiaolong River, and the Dalong River. The rivers in the district mainly originate within it; the main source of recharge is atmospheric precipitation. The course of the rivers is shown in Figure 1. Due to the deep groundwater table in the study area, the river and the groundwater table are disconnected. The rivers recharge groundwater mainly through seepage, forming a river with a suspended saturation zone under the riverbed, an unsaturated zone, and a groundwater aquifer system (Figure 3). From 2006 to 2020, the average total recharge of groundwater was $2.29 \times 10^8$ m$^3$, and the total discharge of groundwater was $2.21 \times 10^8$ m$^3$. It can be seen that the total recharge of groundwater was slightly larger than the total discharge. However, because the recharging of well irrigation depends on the groundwater generated after the seepage of farmland irrigation water, it should not be included in the total amount of groundwater resources. Therefore, the amount of groundwater water resources should be the total recharge amount minus the return recharge from well irrigation, meaning that the average groundwater resource amount from 2006 to 2020 was $2.00 \times 10^8$ m$^3$. It can be seen that the groundwater discharge amount from 2006 to 2020 was greater than the groundwater resource amount.

![Figure 1. Geographic map of the study area.](image-url)
2.2. Data Sources

2.2.1. Groundwater Table and Quality Data

The groundwater table data were mainly derived from a long-term monthly monitoring data series generated by 32 observation wells from 2006 to 2020. The locations of the monitoring wells are shown in Figure 1. The groundwater quality data were mainly derived from the annual monitoring data of 25 shallow groundwater monitoring wells from 2006 to 2020. The locations of these monitoring wells are shown in Figure 1 as well. NO$_3$-N pollution is the main pollutant of groundwater in Beijing, affecting 5% of the total area of the plain. NO$_3$-N pollution has the potential to cause great harm to the health of both humans and livestock [37], and is greatly affected by human activities. Therefore, NO$_3$-N was selected as the evaluation index.

2.2.2. Statistical Data

The statistical data used in this paper were mainly obtained from the following publications, of which the originals are all in Chinese: the water statistical yearbook of Beijing from 2006 to 2020; the statistical yearbook of Daxing District from 2005 to 2020; the data compilation of the third agricultural census of Daxing District in 2016; the second national pollution source census bulletin of Beijing (2017); the regional statistical yearbook of Beijing from 2012 to 2020; the 2020 environmental statistical yearbook of China; the 2018...
comprehensive water resources plan of the Daxing District; the 2017 water supply plan for the Xinfeng River Basin of the Daxing District; and the 2019 main data bulletin of the third national land survey of the Daxing District.

2.3. Research Method

2.3.1. Forecast of Water Demand

According to the Beijing 14th Five-Year Plan for Water-saving Society Construction and related research materials [38–40], we determined the domestic water demand, industrial water demand, agricultural water demand, and ecological water demand in the study area from 2021 to 2025. The specific calculation process was as follows:

- **Forecast of domestic water demand**
  
  Daxing District’s domestic water demand from 2021 to 2025 was calculated from the per capita water consumption and the population using the following formula:
  
  \[ R_1 = S \cdot (1 + a)^t \cdot G \cdot (1 + \beta)^t \cdot 365/1000 \]  

  In Equation (1), \( R_1 \) is Daxing’s domestic water consumption in units of \( 10^4 \) m\(^3\); \( S \) is the current water consumption in L/d·per person; the current water consumption is based on the actual water consumption in 2020, which was 129.73 L/d·per person; \( a \) is the annual growth rate of water use per person, which is currently 2%; \( G \) is the current population in units of \( 10^4 \) persons; the current population is \( 1.994 \times 10^6 \) person; and \( \beta \) is the annual growth rate of the population, currently 4%; \( t \) is elapsed time in years.

- **Forecast of industrial water demand**
  
  Daxing District’s industrial water consumption from 2021 to 2025 was calculated from the water consumption required to generate CNY 10,000 worth of industrial output and the total industrial output using the following formula:
  
  \[ R_2 = g_0 \cdot (1 + v)^t \cdot M_0 \cdot (1 + v')^t \]  

  In Equation (2), \( R_2 \) is Daxing’s industrial water consumption in units of \( 10^4 \) m\(^3\); \( g_0 \) is the current water consumption required to yield CNY 10,000 worth of industrial output in m\(^3\)/(10\(^4\) CNY); \( M_0 \) is the total industrial output value in units of \( 10^8 \) CNY; \( v \) represents the annual growth rate of water consumption per 10\(^4\) CNY of industrial output, currently at −5%; and the annual growth rate of industrial development \( v' \) is 6%.

- **Forecast of agricultural water demand**
  
  The rate of water consumption for animal husbandry and aquaculture in Daxing District is very small. The district’s agricultural water demand is used for the irrigation of crops, mainly wheat. Therefore, only the water used for crop irrigation is incorporated into our calculations, as a proxy for total agricultural water demand in Daxing.
  
  \[ R_3 = \frac{1}{\eta} \cdot \omega \cdot A \]  

  In Equation (3), \( R_3 \) is Daxing’s agricultural water consumption in units of \( 10^4 \) m\(^3\). The symbol \( \eta \) denotes the effective utilization coefficient of irrigation water, meaning the degree of efficiency with which this water is used. The effective utilization coefficient of irrigation water increases by 0.01 every year and is thus expected to be 0.75 in 2025. The symbol \( \omega \) denotes the annual irrigation water quota, which locally amounts to 2850 m\(^3\)/hm\(^2\) in a wet year and 3300 m\(^3\)/hm\(^2\) in a dry year. \( A \) denotes the irrigated area in hm\(^2\). According to the main data bulletin of the third national land survey for Daxing (2019), the area of agricultural land in the study area is 26,371.65 hm\(^2\); thus, we have assigned this value to \( A \).

- **Forecast of ecological water demand**
From 2006 to 2020, the consumption of water in Daxing for ecological purposes has grown continuously. For the purposes of this paper, the district’s demand for ecological water is forecast on the basis of an annual increase of 1%.

\[ R_4 = e_0 \cdot (1 + \gamma)^t \]  

In Equation (4): \( R_4 \) represents the forecast ecological water demand in units of \( 10^4 \) m\(^3\), \( e_0 \) represents the year-2020 demand of \( 1.43 \times 10^8 \) m\(^3\), \( \gamma \) represents the annual growth rate (1%), and \( t \) represents the elapsed time since 2020 in years.

### 2.3.2. Machine Learning Algorithms

In this paper, three machine learning algorithms, namely, long short-term memory (LSTM), RF, and support vector regression (SVR), were selected to predict the change of NO\(_3\)-N concentration in Daxing groundwater from 2021 to 2025. By comparing the accuracy of the prediction models during a validation period, the best prediction model was selected.

The long short-term memory neural network (LSTM) method was first proposed by Hochreiter in 1997 [41]. This algorithm is an improved version of the recurrent neural network (RNN), and is intended to solve the problems of gradient disappearance and gradient explosion caused by the single structure of backpropagation when RNN training a model based on RNN [41]. On the RNN foundation, LSTM erects an input gate, an output gate, a forgetting gate, and a memory unit. These additional gating mechanisms help to overcome gradient problems [42].

The governing equations of the LSTM model are as follows:

- **Input gate:**
  \[ i_t = \sigma(W_{yi}y_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \]  

- **Output gate:**
  \[ f_t = \sigma(W_{yf}y_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f) \]  

- **Neuron activation function:**
  \[ c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{cy}y_t + W_{cm}m_{t-1} + b_c) \]  

- **Forgetting gate:**
  \[ o_t = \sigma(W_{yo}y_t + W_{om}m_{t-1} + W_{oc}c_{t-1} + b_o) \]  

- **Memory cell state input:**
  \[ m_t = o_t \odot \tanh(c_t) \]  

- **Model output:**
  \[ Z_t = W_{2m}m_t + b_z \]  

In these equations, key terms are defined as follows: \( y_t(t = 1, 2, \ldots, T) \) is the model input sequence; \( i_t, o_t, \) and \( f_t \) represent the output values of the input gate, forgetting gate, and output gate at time \( t \), respectively; \( c_t \) and \( m_t \) represent the activation state of the neuron and memory cell state inputs at time \( t \), respectively; \( W \) is the weighting coefficient matrix between different layers; \( \sigma \) is the sigmoid activation function; tanh is the hyperbolic tangent activation function; and \( Z_t(t = 1, 2, \ldots, T) \) is the model output sequence.

RF is a statistical learning method proposed by Breiman in 2001 on the foundation of the Bagging algorithm [43]. The algorithm mainly consists of a sub-training sample set and a sub-decision tree model. The sub-training sample set uses the bootstrap resampling method to extract multiple sub-training sample sets with the same sample size as the original sample from the original sample, then constructs a sub-decision tree model for each sub-training sample set. In this way, the RF model is generated. Each sub-decision tree model generates a prediction, and the final prediction is obtained by voting. Because
the algorithm has good tolerance for outliers and noise and is not prone to overfitting, it shows high simulation accuracy across many different theories and examples [44,45].

SVM is a machine learning algorithm proposed by Vapnik in 1995 [46]. It is a supervised learning model based on statistical learning theory and is mainly used to solve classification and regression problems. The basic idea of SVM is to use a kernel function to convert low-dimensional nonlinear problems into high-dimensional linear problems and then to use linear methods to solve the original nonlinear problems in high-dimensional feature space. Because SVM has the advantages of simple structure and theoretical global optimality, it is the best of the three methods at addressing problems such as small samples, nonlinearity, and high dimensionality.

2.3.3. Optimal Allocation Model for Water Resources

- **Objective function**

  The degree of water scarcity has an important impact on social stability, while the cost of water determines the economic benefits of water use. The objective function that we have chosen aims, as far as possible, to minimize total water shortage and total water cost at the same time.

  (1) Minimum total water shortage

  \[
  f_1(q) = \min \sum_{j=1}^{4} \left[ R_j - \sum_{i=1}^{4} q_{ij} \right] 
  \]

  In Equation (11), the terms are defined as follows. The function \( f_1(q) \) is the total water shortage in units of \( 10^4 \) m\(^3\); the parameter \( q_{ij} \) is the annual water supply of the \( i \)th water source to the \( j \)th type of water user, in units of \( 10^4 \) m\(^3\); \( R_j \) is the annual water demand of the \( j \)th type of users, in units of \( 10^4 \) m\(^3\); the parameter \( i \) represents the \( i \)th water source, whereby \( i = 1, 2, 3, 4 \) are numbers that respectively denote surface water, groundwater, reclaimed water, and transferred water; and the parameter \( j \) represents the \( j \)th water user, whereby \( j = 1, 2, 3, 4 \) are numbers that respectively denote domestic water, industrial water, agricultural water, and ecological water.

  (2) Minimum water cost

  \[
  f_2(q) = \min \sum_{i=1}^{4} \left( \sum_{j=1}^{4} q_{ij} \cdot p_i \right) 
  \]

  In Equation (12), the terms not already defined above are defined as follows: the function \( f_2(q) \) is the cost of water in units of \( 10^4 \) CNY; \( p_i \) is the unit price of water supply for the \( i \)th water source, in CNY/m\(^3\); and the unit price of water supply for each water source is determined according to the existing unit price of water supply in Beijing, which is 0.48 CNY/m\(^3\) for surface water, 5.00 CNY/m\(^3\) for groundwater, 1.00 CNY/m\(^3\) for reclaimed water, and 7.44 CNY/m\(^3\) for transferred water.

- **Constraint condition**

  (1) Water supply capacity constraints

  \[
  \sum_{i=1}^{4} \sum_{j=1}^{4} q_{ij} \leq Q_z
  \]

  \[
  \sum_{j=1}^{4} q_{ij} \leq Q_i
  \]

  In Equations (13) and (14), \( Q_z \) is the total available annual volume of all the water sources in units of \( 10^4 \) m\(^3\), while \( Q_i \) is the available annual volume of each water source in units of \( 10^4 \) m\(^3\). According to the comprehensive plan for of water resources in Daxing,
the maximum available volume of surface water in the study area \( Q_1 \) is \( 1572 \times 10^4 \) m\(^3\) in a normal year or a wet year and \( 1001 \times 10^4 \) m\(^3\) in a dry year. The maximum usable water amount of groundwater \( Q_2 \) is determined in accordance with whether the year is wet, dry, or normal. The maximum usable volume of groundwater in a wet year is \( 22,450.66 \times 10^4 \) m\(^3\), the maximum usable volume of groundwater in a normal year is \( 18,086.54 \times 10^4 \) m\(^3\), and the maximum usable volume of groundwater in a dry year is \( 14,931.44 \times 10^4 \) m\(^3\). The maximum available volume of transfer water \( Q_3 \) is \( 2600.00 \times 10^4 \) m\(^3\). According to the water supply plan for the Xinfeng River Basin of Daxing District, the maximum usable volume of reclaimed water \( Q_4 \) is \( 15,987.00 \times 10^4 \) m\(^3\).

(2) water consumption constraints

The four different water uses need to fall between their respective maximum and minimum water demand constraints, as follows:

\[
Q_{\text{min},j} \leq Q_j \leq Q_{\text{max},j}
\]  

(15)

In Equation (15), \( Q_{\text{min},j} \) and \( Q_{\text{max},j} \) represent the minimum water demand and maximum water demand of the \( j \)th water user \( Q_j \), respectively; the minimum water demands \( Q_{\text{min},1} \) and \( Q_{\text{min},2} \) for domestic water and industrial water are in each case 99% of the respective nominal water demand, and the maximum demands \( Q_{\text{max},1} \) and \( Q_{\text{max},2} \) are each 110% of the respective nominal water demand; the minimum water demands \( Q_{\text{min},3} \) and \( Q_{\text{min},4} \) for agricultural water and ecological water are respectively 75% and 70% of each nominal water demand; and the maximum values \( Q_{\text{max},3} \) and \( Q_{\text{min},4} \) are each 110% of the respective nominal water demand.

In addition, the water supply and water demand in the above formula should meet non-negativity constraints.

- Model solution

The water resource optimization model constructed in this paper is a multi-objective optimization problem with constraints. For a solution algorithm, we have chosen the Non-dominated Sorting Genetic Algorithm-II (NSGA-II), which shows good performance in dealing with constrained multi-objective optimization problems [47]. NSGA-II was proposed by Deb in 2002 as an improved version of NSGA. Because NSGA-II adopts the strategy of fast non-dominated sorting and crowding distance, the calculation speed and effect of this algorithm are greatly improved in comparison to NSGA [48].

2.3.4. Numerical Simulation of Groundwater Flow

We made use of known drilling and hydrological data for Daxing to establish a hydrogeological model of a conceptual nature. A numerical simulation package called GMS (Groundwater Modeling System 7.1) was used to generalize the aquifer into a two-dimensional non-steady-flow phreatic aquifer with heterogeneity and isotropy. Although parts of this aquifer contain micro-confined water, the medium-fine sand layer in the quaternary loose layer is the main water-bearing medium; permeability extends between the vertically stable water-retaining layers. Thus, the aquifer has the same groundwater table throughout, yielding a thickness in the range of 80–120 m.

- Zoning generalization of hydrogeological parameters

In accordance with the drilling data, Daxing was divided into three hydrogeological parameter divisions; this division is shown in Figure 4. The parameters used in the calculation of the water flow model were the permeability coefficient \( K \) and the specific yield \( \mu \), with their respective value ranges shown in Table 1.
Table 1. Hydrogeological parameters of Daxing.

<table>
<thead>
<tr>
<th>Parameter Partition Number</th>
<th>Zone Area (km²)</th>
<th>Permeability Coefficient K (m/d)</th>
<th>Specific Yield μ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone I</td>
<td>532.10</td>
<td>15–25</td>
<td>0.20–0.26</td>
</tr>
<tr>
<td>Zone II</td>
<td>384.99</td>
<td>8–10</td>
<td>0.13–0.20</td>
</tr>
<tr>
<td>Zone III</td>
<td>118.91</td>
<td>10–15</td>
<td>0.10–0.15</td>
</tr>
</tbody>
</table>

• Boundary condition generalization

The direction of groundwater flow in Daxing is generally from west to east. In accordance with the contour distribution of the groundwater table in the Beijing plain and the study area (see Figure 5), the boundaries of the aquifer beneath Daxing were divided into two categories: a constant-flux boundary and a zero-flux boundary. As shown in Figures 4 and 5, the groundwater table contours in the north, northwest, and southwest are parallel to the boundary in ways that create lateral inflow boundaries, while the eastern boundary was a lateral outflow boundary. The groundwater table contours in parts of the west and southeast are perpendicular to zero-flux boundaries. The upper boundary of the aquifer beneath Daxing mainly receives atmospheric precipitation, river supply, and farmland irrigation supply, and is the diving boundary. As average depth of groundwater below the surface of the soil was 17.31 m, evaporation could be ignored. According to the drilling data, most of the aquifer exists above a depth of 120 m, and there is a thicker clay layer beneath it.
Mathematical model establishment

In accordance with the conceptual model that has just been described, the groundwater flow beneath Daxing can be generalized into a non-homogeneous and isotropic unsteady flow model for the purposes of more detailed numerical analysis. The mathematical model for simulating the aquifer is as follows:

\[
\frac{\partial}{\partial x} \left[ K(H - B) \frac{\partial H}{\partial x} \right] + \frac{\partial}{\partial y} \left[ K(H - B) \frac{\partial H}{\partial y} \right] + Q_r - Q_d = \mu \frac{\partial H}{\partial t} \tag{16}
\]

\[
H(x, y, t)|_{t=0} = h_0(x, y, t), \quad (x, y) \in D \tag{17}
\]

\[
H(x, y, t)|_{\Gamma_1} = h_1(x, y, t), \quad (x, y) \in D, t>0 \tag{18}
\]

\[
K(H - B) \frac{\partial H}{\partial n} \bigg|_{\Gamma_2} = q(x, y, t), \quad (x, y) \in D, t>0 \tag{19}
\]

In Equations (16)–(19), the new parameters are as follows: \( K \) denotes the permeability coefficient of the aquifer in \( \text{m} / \text{d} \); \( H \) denotes the groundwater table in \( \text{m} \); \( B \) denotes the aquifer floor elevation in \( \text{m} \); \( Q_r \) denotes the groundwater recharge rate in \( \text{m}^3 / \text{d} \); \( Q_d \) denotes the groundwater discharge rate in \( \text{m}^3 / \text{d} \); the parameter \( \mu \) denotes the specific yield of the aquifer; the parameter \( h_1 \) is the groundwater table of the boundary point of class I in \( \text{m} \); the parameter \( q \) denotes the single-width flow rate of the class II boundary in \( \text{m}^3 / (\text{d} \cdot \text{m}) \); the parameters \( x \) and \( y \) are coordinates specified in \( \text{m} \); \( D \) is the calculation area range; the parameter \( n \) is the inner normal on the boundary; and the identifiers \( \Gamma_1 \) and \( \Gamma_2 \) denote class I and II boundaries, respectively.

Equation (16) is the partial differential equation of unsteady flow in the diving plane, which describes the movement of shallow groundwater.

Equations (17)–(19) are the definite solution conditions of Equation (16), for which (17) is the initial condition and (18) and (19) are the boundary conditions.

Mathematical model solution
The aquifer beneath Daxing was discretized in accordance with our hydrogeological conceptual model in a manner that took realistic account of the likely levels of simulation accuracy and data accuracy. Spatially, Daxing was divided into a rectangular horizontal grid with a grid spacing of 500 m × 500 m; 4700 such grid blocks were generated. The parameter identification period was set from January 2015 to December 2019, and the stress period was one month, which was divided into sixty stress periods. The verification period was from January 2020 to December 2020, and the stress period was one month, which was divided into twelve stress periods. The water flow simulation data for the aquifer were input into the Modflow module in GMS (Groundwater Modeling System 7.1) to evaluate the changes in the groundwater flow field.

Model identification and validation

We used the measured data from 32 groundwater level monitoring points from January 2015 to December 2019 to identify groundwater levels, and used the groundwater table data from the same 32 monitoring points from January 2020 to December 2020 to verify the model.

1. Processing of Source and Sink Items

The groundwater recharge sources in Daxing principally comprised the atmospheric precipitation infiltration recharge, river infiltration recharge, well irrigation return recharge, and groundwater lateral inflow.

The principal discharges comprised artificial exploitation (including industrial water, domestic water, agricultural water, and ecological water) and ground water lateral outflow. The source and sink items from 2015 to 2020 were obtained through the groundwater balance calculation. The results of these calculations are shown in Table 2.

Table 2. Calculation results of groundwater source and sink items in Daxing District from 2015 to 2020.

<table>
<thead>
<tr>
<th>Source and Sink Items</th>
<th>2015 (10⁴ m³)</th>
<th>2016 (10⁴ m³)</th>
<th>2017 (10⁴ m³)</th>
<th>2018 (10⁴ m³)</th>
<th>2019 (10⁴ m³)</th>
<th>2020 (10⁴ m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmospheric precipitation infiltration recharge</td>
<td>15,699.30</td>
<td>20,928.10</td>
<td>15,955.23</td>
<td>14,423.30</td>
<td>12,782.17</td>
<td>14,346.22</td>
</tr>
<tr>
<td>River infiltration recharge</td>
<td>2,230.10</td>
<td>3,012.80</td>
<td>2,278.90</td>
<td>2,060.09</td>
<td>1,825.69</td>
<td>2,049.08</td>
</tr>
<tr>
<td>Well irrigation return recharge</td>
<td>2,732.10</td>
<td>2,580.60</td>
<td>2,388.00</td>
<td>2,143.80</td>
<td>2,164.00</td>
<td>1,472.00</td>
</tr>
<tr>
<td>Groundwater lateral inflow</td>
<td>376.20</td>
<td>530.00</td>
<td>403.9024</td>
<td>365.12</td>
<td>323.58</td>
<td>363.17</td>
</tr>
<tr>
<td>Artificial exploitation</td>
<td>21,672.50</td>
<td>21,390.73</td>
<td>20,618.63</td>
<td>18,791.62</td>
<td>19,493.98</td>
<td>15,507.55</td>
</tr>
<tr>
<td>Groundwater lateral outflow</td>
<td>103.00</td>
<td>145.10</td>
<td>108.35</td>
<td>97.95</td>
<td>86.80</td>
<td>97.43</td>
</tr>
<tr>
<td>Changes in groundwater storage</td>
<td>−737.80</td>
<td>5535.67</td>
<td>299.05</td>
<td>102.74</td>
<td>−2485.34</td>
<td>2625.49</td>
</tr>
</tbody>
</table>

2. Model Identification and Verification

The initial flow field for model simulation was that of 1 January 2015. The groundwater flow model was identified and validated using data from the 32 groundwater table monitoring wells from January 2015 to December 2019 and from January 2020 to December 2020, respectively. By changing the model parameters, adjusting the boundary conditions, and repeating the trial calculations, the errors between the calculated and observed values at most of the observation wells could be brought within the simulation accuracy requirements. In the process of groundwater flow simulation and identification, the boundary conditions and model parameters, such as specific yield \(\mu\) and permeability coefficient \(k\), were substantially adjusted. The ratio between the calculated level and the actual level of the groundwater table during the validation period of the groundwater flow model is shown in Figure 6. It can be seen from Figure 6 that the ratio of the calculated level to the observed level during the model verification period was around 1:1. The absolute error between the calculated level and the actual level was less than 1 m for 80.50% of the observations, the maximum absolute error was 1.90 m, and the maximum relative error was 13.16%. It can be seen that the overall degree of fit between the calculated groundwater
table and the measured groundwater table is good, which meant that the model was able to accurately reflect the actual characteristics of the groundwater flow field under Daxing. Therefore, the model can be used to predict future changes in the groundwater flow field. The identified hydrogeological parameters are shown in Table 3.

![Graph](image)

**Figure 6.** The ratio between the calculated level and the actual level of the groundwater table during the verification period (the elevation reference system is the 1956 Yellow Sea Elevation System of China).

**Table 3.** Hydrogeological parameters derived for the three aquifer zones of Daxing.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Zone</th>
<th>Area (km²)</th>
<th>Permeability Coefficient K (m/d)</th>
<th>Specific Yield μ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone I</td>
<td>532.10</td>
<td>15.4</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Zone II</td>
<td>384.99</td>
<td>9.95</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Zone III</td>
<td>118.91</td>
<td>11.7</td>
<td>0.105</td>
<td></td>
</tr>
</tbody>
</table>

3. Results and Discussion

3.1. Forecast of NO₃-N Concentration

In accordance with NO₃-N monitoring data from the 25 groundwater quality monitoring points from 2006 to 2020, the water quality value of the study area was obtained by interpolation based on the inverse distance weighting method. The maximum number of prediction points within the search radius was 15, and the minimum number of prediction points within the search radius was 10. Through cross-validation, the average error and root mean square error of the NO₃-N concentration were found to be 0.09 and 0.53, respectively. These were both small enough to ensure that the average error and root mean square error met the requirements of our evaluation.

To reduce the influence of fluctuations in NO₃-N pollution concentration in individual years, the period 2006–2020 was divided into three shorter time periods, namely, 2006–2010, 2011–2015, and 2016–2020. These three time periods correspond to the three five-year plans formulated by the Chinese government, namely, the Eleventh Five-Year Plan for National Economic and Social Development of the People’s Republic of China (2006–2010), the Twelfth Five-Year Plan for National Economic and Social Development of the People’s Republic of China (2011–2015), and the Thirteenth Five-Year Plan for National Economic and Social Development of the People’s Republic of China (2016–2020). Using the grid calculator in our GIS package, the average value of the GIS interpolation results in each time period was generated, yielding final calculation results shown in Figure 7.
To further predict the NO$_3$-N pollution in wet years, normal years, and dry years from 2021 to 2025, we created Thiessen polygons in ArcGIS around the water quality monitoring points [49,50]; these polygons are shown in Figure 8. Then, using the regional analysis tool in ArcGIS, the Thiessen polygons were used to extract the values of the following eight parameters in each polygon: industrial water use, domestic water use, agricultural water use, industrial output, population, tertiary industry output, agricultural output, and rainfall. Machine learning algorithms were used to predict the changes of NO$_3$-N concentration in wet years, normal years, and dry years from 2021 to 2025.

It can be seen in Figure 7 that the NO$_3$-N concentration showed an overall increasing trend from 2006 to 2020. Across the whole district, the average concentration of NO$_3$-N from 2006 to 2010 was 5.52 mg/L. On the same basis, the average concentration of NO$_3$-N from 2011 to 2015 was 5.47 mg/L, a slight decrease. However, the average concentration of NO$_3$-N from 2016 to 2020 was 5.92 mg/L, an increase of 0.40 mg/L compared with the average concentration of NO$_3$-N from 2006 to 2010. Overall, these changes indicated that with the continuous increase of urbanization in the study area, NO$_3$-N shows an increasing trend as well.

From a spatial perspective (Figure 7), the average concentration of NO$_3$-N in the northern region of Daxing was greater than that in the central and southern regions. The
average concentration of NO$_3$-N in the northern region increased from 12.62 mg/L in 2006–2010 to 14.05 mg/L in 2016–2020. The average concentrations of NO$_3$-N in the central and southern regions remained at 2.67–2.80 mg/L and 1.03–1.57 mg/L from 2006 to 2020, respectively, indicating that the groundwater quality in the central and southern regions was better than that in the northern regions. This might be related to the higher degree of urbanization in the northern regions than in the central and southern regions. According to the 2020 statistical yearbook for Daxing, the industrial output value, tertiary industry output value, and resident population of the northern region accounted for 73.27%, 72.24%, and 60.84% of the equivalents for the whole district, respectively, showing that the NO$_3$-N pollution was largely caused by the intensity of human activity.

We have modeled the current types of water use in Daxing as domestic water, industrial water, agricultural water, and ecological water. Because the current water supply in the study area relies on groundwater as the single most important of the four sources we modeled, it was necessary to evaluate the utility of local groundwater in view of the observed NO$_3$-N pollutant loads.

The water quality standards for domestic water, industrial water, agricultural water, and ecological water are referred to in the relevant official water quality standard for urban water supply (CJ/T 206-2005), the drinking water hygiene standard (GB 5749-2006), the reuse of urban recycling water—water quality standard for industrial uses (GB 19923-2005), the standard for irrigation water quality (GB 5084-2021), and the reuse of urban recycling water—water quality standard for scenic environmental use (GB/T 18921-2019), respectively. Because only the domestic water standards prescribe limits for NO$_3$-N concentration, while the water quality standards for industrial water, agricultural water, and ecological water do not, the groundwater was considered to meet the requirements for industrial water, agricultural water, and ecological water. However, at the same time it was necessary to further evaluate the utility of the local groundwater for domestic water supply in view of its NO$_3$-N concentrations.

To further evaluate the utility of Daxing’s groundwater for domestic water usage, the evaluation had to be carried out in a manner informed by the current water quality standard for urban water supply (CJ/T 206-2005) and the drinking water hygiene standard (GB 5749-2006). According to the drinking water hygiene standard (GB 5749-2006), the limit of NO$_3$-N for groundwater sources is 20 mg/L. However, according to the water quality standard for urban water supply (CJ/T 206-2005), the limit for NO$_3$-N is 10 mg/L (20 mg/L in special cases). Therefore, in cases where there is sufficient water, NO$_3$-N concentrations in the range of 10–20 mg/L mean that water should not be used as domestic water; under conditions of serious water shortage, the NO$_3$-N concentrations in the 10–20 mg/L range are acceptable for domestic water use; and water with an NO$_3$-N concentration above 20 mg/L should not be used as domestic water.

It can be seen from the above analysis that the groundwater in the northern region was seriously polluted and not suitable for domestic water, while the groundwater quality in the central and southern regions was better and could be used as domestic water.

To further predict the NO$_3$-N pollution in wet years, normal years, and dry years from 2021 to 2025, we created Thiessen polygons in ArcGIS around the water quality monitoring points [49,50]; these polygons are shown in Figure 8. Then, using the regional analysis tool in ArcGIS, the Thiessen polygons were used to extract the values of the following eight parameters in each polygon: industrial water use, domestic water use, agricultural water use, industrial output, population, tertiary industry output, agricultural output, and rainfall. Machine learning algorithms were used to predict the changes of NO$_3$-N concentration in wet years, normal years, and dry years from 2021 to 2025.
The three machine algorithms introduced above, namely, LSTM, RF, and SVR, were used to predict the NO$_3$-N concentration change from 2021 to 2025. By comparing the accuracy of their predictions during a validation period, the best prediction model was selected.

The NO$_3$-N concentration of each water quality monitoring point from 2006 to 2016 was used as the training period sample for the three machine learning algorithms, and the NO$_3$-N concentration of each water quality monitoring point from 2017 to 2020 was used as the validation period sample for the three machine algorithms. The mean absolute error, root mean square error, correlation coefficient, and Nash efficiency coefficient were selected as the evaluation accuracy indicators for the machine algorithms. The ratios of the NO$_3$-N concentration calculation values to the monitored values for the three machine algorithms during the validation period are shown in Figure 9. The mean absolute error, root mean square error, correlation coefficient, and Nash efficiency coefficient between the NO$_3$-N concentration calculated value and the monitoring value of the three machine algorithms during the validation period are shown as numbers in Table S1.

It can be seen from Figure 9 that among the three machine algorithms, LSTM has the highest accuracy; for this algorithm, the ratio of the calculated value of NO$_3$-N concentration to the monitored value was always near the 1:1 line. On the other hand, the NO$_3$-N concentration value calculated by the RF and SVR algorithms showed a large deviation from the monitored value, with the deviation of the RF algorithm being the greatest. Table S1 shows that the average absolute error and root mean square error of NO$_3$-N concentration calculated by LSTM were 0.05 and 0.91, figures which were respectively less than both average absolute errors (1.45 and 0.09) and both root mean square errors (3.59 and 2.98) of the NO$_3$-N concentration as calculated by RF and SVR. At the same time, the correlation coefficient and Nash efficiency coefficient between the NO$_3$-N concentration and the actual value calculated by LSTM were 0.98 and 0.98, which were respectively higher than both correlation coefficients (0.93, 0.94) and both Nash efficiency coefficient (0.62, 0.80) as calculated by RF and SVR.

Figure 8. Thiessen polygons based on the 25 groundwater quality monitoring points in Daxing and the district boundaries.
Figure 9. The ratio between the calculated and actual values of NO\textsubscript{3}-N concentration for the three machine learning algorithms during the validation period: (a) LSTM; (b) RF; (c) SVR.

These results, presented in Figure 9 and Table S1 show that the prediction accuracy of LSTM for NO\textsubscript{3}-N concentration in Daxing was unambiguously better than that of RF and SVR. The order of computational accuracy of the three machine learning algorithms was LSTM > SVR > RF.

Therefore, we chose LSTM as the machine learning system with which to predict the changes of NO\textsubscript{3}-N concentration in Daxing in wet years, normal years, and dry years from 2021 to 2025.

The water demands for a wet year, a normal year, and a dry year in the years 2021–2025 were calculated in accordance with Equations (1)–(4). The rate of economic development was extrapolated from the average rate of economic development between 2006 and 2020. According to the rainfall sequence from 1956 to 2020, the average rainfall in wet year, normal year, and dry year was 598.40 mm, 482.10 mm and 398.00 mm, respectively. In wet, normal, and dry years, the precipitation guarantee rates were 25%, 50%, and 75%, respectively.
Using the trained LSTM model, the changes in the concentration of NO$_3$-N in a wet year, a normal year, and a dry year from 2021 to 2025 were predicted. The results of this prediction are shown in Figure 10.

As compared with the years 2016 to 2020, average district-wide NO$_3$-N concentrations increased from 5.92 mg/L to 6.34 mg/L, 6.61 mg/L, and 6.96 mg/L in a wet year, a normal year, and a dry year, respectively. In more detail, the NO$_3$-N concentration in the northern region of Daxing increased from 14.05 mg/L to 14.87 mg/L, 15.64 mg/L, and 16.54 mg/L, respectively as the model moved from the period 2016–2020 to the period 2021–2025 (wet year), 2021–2025 (normal year), and 2021–2025 (dry year). It can be seen that the NO$_3$-N concentration in the normal year and the dry year increased more sharply than in the wet year. This can be explained by the fact that groundwater recharge is larger in the wet year than in normal and dry years, thus diluting NO$_3$-N pollution more effectively and transporting the NO$_3$-N out of the Daxing aquifer more rapidly.

![Figure 10](image-url)
To ensure the most socially effective use of water, these figures may be used to create a flexible regime of controls on groundwater use in which the extraction of groundwater for domestic consumption is reduced or prohibited in areas with serious groundwater pollution in ways that take recent rates of dilution and recharge due to rainfall into account.

3.2. Optimal Allocation of Water Resources

To reduce the impact of groundwater NO$_3$-N pollution on domestic water supplies, three groundwater exploitation regimes were established.

Groundwater exploitation regime 1: no area with a groundwater NO$_3$-N concentration above 20 mg/L is to be permitted to contribute its groundwater to the domestic water supply.

Groundwater exploitation regime 2: groundwater extracted from areas with NO$_3$-N concentrations above 10 mg/L but below 20 mg/L are not to be used for domestic purposes, and groundwater extraction from areas with NO$_3$-N concentrations above 20 mg/L are prohibited altogether.

Groundwater exploitation regime 3: groundwater extraction from all areas with an NO$_3$-N concentrations above 10 mg/L is prohibited altogether.

According to the spatial distribution map of the predicted NO$_3$-N concentration in the wet year, the normal year, and the dry year from 2021 to 2025 (Figure 10), the areas where NO$_3$-N concentration was 10–20 mg/L and where it was over 20 mg/L were respectively extracted by the regional analysis tool in ArcGIS.

Taking 2025 as an example, the water consumption of the study area in 2025 was calculated according to Equations (1)–(4). Under groundwater exploitation regimes 1–3, the respective available groundwater volumes in a wet, normal, and dry 2025 were calculated. The calculation results are shown in Table 4. It can be seen from Table 4 that under groundwater exploitation regimes 1–3, the available groundwater volumes in the wet year were 214,692,000 m$^3$, 176,519,300 m$^3$, and 168,058,400 m$^3$, respectively. Under groundwater exploitation regimes 1–3, the available groundwater volumes in the normal year were 165,455,900 m$^3$, 132,175,600 m$^3$, and 125,083,400 m$^3$, respectively. Under groundwater exploitation regimes 1–3, the available groundwater volumes in the dry year were 123,817,600 m$^3$, 98,940,400 m$^3$, and 93,196,100 m$^3$, respectively. It can be seen that available groundwater volumes in the wet year were the largest, the available groundwater volumes in the normal year came second, and the available groundwater volumes in the dry year were the least.

Table 4. The available groundwater volumes for Daxing under groundwater exploitation regimes 1–3.

<table>
<thead>
<tr>
<th>Hydrological Year Type</th>
<th>The Maximum Available Groundwater Volume before Reduction (10$^4$ m$^3$)</th>
<th>Groundwater Exploitation Regime 1 (10$^4$ m$^3$)</th>
<th>Groundwater Exploitation Regime 2 (10$^4$ m$^3$)</th>
<th>Groundwater Exploitation Regime 3 (10$^4$ m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wet year</td>
<td>22,450.66</td>
<td>961.46</td>
<td>21,489.20</td>
<td>4798.73</td>
</tr>
<tr>
<td>Normal year</td>
<td>18,086.54</td>
<td>1540.95</td>
<td>16,545.59</td>
<td>4868.98</td>
</tr>
<tr>
<td>Dry year</td>
<td>1,493.44</td>
<td>2549.68</td>
<td>12,381.76</td>
<td>5037.4</td>
</tr>
</tbody>
</table>

In accordance with our optimal allocation model for water resources, the optimal allocation of water consumption in 2025 was carried out for wet, normal, and dry year scenarios. The optimization results are shown in Tables S2–S4. Tables S2–S4 show the results of water resources allocation for the current groundwater exploitation regime and the optimal allocation of water resources for groundwater exploitation regimes 1–3.

To select the best groundwater exploitation regime under different hydrological year scenarios, it was necessary to select the best exploitation mode from three aspects: water shortage, groundwater utilization rate as a percentage of the maximum sustainable limit, and the minimization of NO$_3$-N concentration in domestic water. In accordance with
the configuration results in Tables S2–S4, we calculated the water shortage risks and source costs of each groundwater exploitation mode. In accordance with the limited groundwater exploitation amounts under exploitation regimes 1–3, and bearing in mind that water finally delivered would be from mixed sources, the respective maximum NO$_3$-N pollution concentration of each groundwater exploitation regime was calculated. The calculation results are shown in Table 5.

Table 5. Optimal evaluation of groundwater exploitation modes.

<table>
<thead>
<tr>
<th>Hydrological Year Type</th>
<th>Groundwater Exploitation Regime</th>
<th>Water Shortage ($10^4$ m$^3$)</th>
<th>Amount of Groundwater Used ($10^4$ m$^3$)</th>
<th>The Maximum Amount of Groundwater Available ($10^4$ m$^3$)</th>
<th>Groundwater Utilization Rate</th>
<th>NO$_3$-N Maximum Concentration (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wet year</td>
<td>Current groundwater exploitation regime</td>
<td>0</td>
<td>21,456.09</td>
<td>22,450.66</td>
<td>95.57%</td>
<td>27.15</td>
</tr>
<tr>
<td></td>
<td>Groundwater exploitation regime 1</td>
<td>0</td>
<td>16,630.11</td>
<td>22,450.66</td>
<td>74.07%</td>
<td>13.32</td>
</tr>
<tr>
<td></td>
<td>Groundwater exploitation regime 2</td>
<td>0</td>
<td>16,630.11</td>
<td>22,450.66</td>
<td>74.07%</td>
<td>12.00</td>
</tr>
<tr>
<td></td>
<td>Groundwater exploitation regime 3</td>
<td>0</td>
<td>16,630.11</td>
<td>22,450.66</td>
<td>74.07%</td>
<td>10.64</td>
</tr>
<tr>
<td>Normal year</td>
<td>Current groundwater exploitation regime</td>
<td>0</td>
<td>21,456.09</td>
<td>18,086.54</td>
<td>118.63%</td>
<td>28.24</td>
</tr>
<tr>
<td></td>
<td>Groundwater exploitation regime 1</td>
<td>0</td>
<td>16,545.59</td>
<td>18,086.54</td>
<td>91.48%</td>
<td>17.63</td>
</tr>
<tr>
<td></td>
<td>Groundwater exploitation regime 2</td>
<td>0</td>
<td>13,217.56</td>
<td>18,086.54</td>
<td>73.08%</td>
<td>11.52</td>
</tr>
<tr>
<td></td>
<td>Groundwater exploitation regime 3</td>
<td>0</td>
<td>12,508.34</td>
<td>18,086.54</td>
<td>69.16%</td>
<td>10.85</td>
</tr>
<tr>
<td></td>
<td>Current groundwater exploitation regime</td>
<td>0</td>
<td>23,038.39</td>
<td>14,931.44</td>
<td>154.29%</td>
<td>28.55</td>
</tr>
<tr>
<td>Dry year</td>
<td>Groundwater exploitation regime 1</td>
<td>0</td>
<td>12,381.76</td>
<td>14,931.44</td>
<td>82.92%</td>
<td>17.72</td>
</tr>
<tr>
<td></td>
<td>Groundwater exploitation regime 2</td>
<td>−2440.86</td>
<td>9894.04</td>
<td>14,931.44</td>
<td>66.26%</td>
<td>10.85</td>
</tr>
<tr>
<td></td>
<td>Groundwater exploitation regime 3</td>
<td>−3015.29</td>
<td>9319.61</td>
<td>14,931.44</td>
<td>62.42%</td>
<td>10.14</td>
</tr>
</tbody>
</table>

It can be seen from Table 5 that the groundwater utilization rate and the maximum NO$_3$-N pollution concentration of groundwater exploitation regimes 1–3 were lower than the current exploitation regime in all three hydrological year types (wet, normal, and dry), indicating that groundwater exploitation regimes 1–3 are better than the current groundwater exploitation regime.

In the wet year, the water shortages and groundwater utilization rates of groundwater exploitation regimes 1–3 were 0 and 74.07%, while the maximum concentration of NO$_3$-N in groundwater exploitation regime 3 was 10.64 mg/L, which was less than that of groundwater exploitation regimes 1 and 2. Therefore, groundwater exploitation regime 3 was selected for the wet year.

The water shortage of groundwater exploitation regime 3 was 0 in the normal year, and the groundwater utilization rate and maximum concentrations of NO$_3$-N under groundwater exploitation regime 3 were 69.16% and 11.02 mg/L, respectively, which were lower than for groundwater exploitation regimes 1 and 2. Therefore, groundwater exploitation regime 3 was selected for normal water years.

In the dry year, groundwater exploitation regimes 2 and 3 gave rise to water shortages of 24,408,600 m$^3$ and 30,152,900 m$^3$, respectively. Although the groundwater utilization rate and NO$_3$-N maximum concentration of regimes 2 and 3 decreased compared with
regime 1, to ensure social stability and economic development we selected groundwater exploitation regime 1 for the dry year.

3.3. Changes in NO$_3$-N Pollution before and after Optimal Regulation

It can be seen from the above that the best groundwater exploitation modes in wet, normal, and dry years were groundwater exploitation regime 3, groundwater exploitation regime 3, and groundwater exploitation regime 1, respectively. Taking 2025 as the example year, we used ArcGIS to draw the NO$_3$-N concentration distribution map for the optimal groundwater exploitation regime in the wet, normal, and dry year scenarios, respectively (see Figures 11–13).

![Figure 11](image1.png)  
**Figure 11.** Changes of NO$_3$-N concentration in a wet year 2025 (groundwater exploitation regime 3): (a) before optimization and (b) after optimization.

![Figure 12](image2.png)  
**Figure 12.** Changes of NO$_3$-N concentration in a normal year 2025 (groundwater exploitation regime 3): (a) before optimization and (b) after optimization.
3.4. Changes in the Groundwater Table before and after Optimal Regulation

Taking 2025 as the example year, the Modflow model was used to simulate the changes in the groundwater table before and after optimal regulation. Taking 2020 as the initial groundwater table datum, the water resources from 2021 to 2025 were allocated according to groundwater exploitation regime 3, groundwater exploitation regime 3, and groundwater exploitation regime 1 in wet years, normal years, and dry years, respectively. The groundwater consumption before and after optimization was assumed to be the same as the groundwater exploitation volume before and after optimization. Taking 2021–2025 as the simulation period, the groundwater table distribution map for 2025 was obtained (Figures 14–16).

It can be seen from Figures 14–16 that the groundwater exploitation volume in the wet year was based on groundwater exploitation regime 3. Compared with the pre-optimization regime, the average groundwater table rose from 8.32 m to 9.31 m, meaning that the groundwater table rose by 0.99 m. From the perspective of the spatial change in the groundwater table, the groundwater table in the northern region rose from 13.61 m to 14.71 m, meaning that the groundwater table had risen by 1.10 m. The groundwater table in the central region rose by 0.74 m, and the groundwater table in the southern region rose by 1.12 m.
The main use of groundwater is by agriculture.

Changes of groundwater table in normal year 2025 (groundwater exploitation regime 3): (a) before optimization and (b) after optimization.

Figure 15. Changes of groundwater table in normal year 2025 (groundwater exploitation regime 3): (a) before optimization and (b) after optimization.
The groundwater exploitation mode in the normal year was based on groundwater exploitation regime 3. Compared with the preoptimization regime, the average groundwater table rose from 7.09 m to 8.89 m, meaning that the groundwater table rose by 1.80 m. In more detail, the groundwater table in the northern region rose from 12.63 m to 14.38 m, a rise of 1.75 m. The groundwater table in the central region rose from 4.91 m to 6.47 m, a rise of 1.56 m. The groundwater table in the southern region rose from 3.72 m to 5.81 m, a rise of 2.09 m.

The groundwater exploitation mode in the dry year was based on groundwater exploitation mode 1. Compared with the pre-optimization regime, the average groundwater table rose from 6.10 m to 8.17 m, a rise of 2.07 m. The groundwater table in the northern region rose from 11.82 m to 13.80 m, a rise of 1.98 m; The groundwater level in the central region rose from 3.96 m to 5.80 m, a rise of 1.84 m; The groundwater level in the southern region rose from 2.49 m to 4.90 m, a rise of 2.41 m.

It can be seen from the above analysis that the groundwater level in the southern region increased the most, which might be related to the fact that the southern region is mainly agricultural. In this paper, after setting the guaranteed rate of agricultural water use and increasing the water supply of reclaimed water and surface water for agricultural use, the exploitation of groundwater was reduced, resulting in rapid recovery of the groundwater level in the southern region.

4. Conclusions

In this paper, by combining GIS with machine-learning algorithms, water resources optimization technology, and groundwater numerical simulation, the groundwater quality and table in the Daxing District of Beijing were effectively optimized and controlled. The method developed here can provide theoretical and practical guidance for groundwater governance in areas with serious groundwater pollution and overexploitation. In more detail, the main conclusions of this paper are as follows:

(1) ArcGIS was used to generate Thiessen polygons around 25 water quality monitoring points. The regional analysis tool in ArcGIS was then used to extract the water consumption and economic development data for each polygon. Finally, three machine-learning algorithms were used to predict the NO₃-N pollution under three different

![Figure 16. Changes of groundwater table in a dry 2025 (groundwater exploitation regime 1): (a) before optimization and (b) after optimization.](image)
hydrological year scenarios for the Daxing District. The results show that LSTM had better accuracy than RF and SVR. Therefore, further research on the combination of GIS and LSTM has high theoretical and practical value.

(2) According to our model’s predictions of future groundwater NO$_3$-N pollution, this form of pollution shows an increasing trend in Daxing under the existing urban development regime. As compared with 2016–2020, the NO$_3$-N pollution of groundwater over the period from 2021–2025 increased from 14.05 mg/L to 14.87 mg/L, 15.64 mg/L, and 16.54 mg/L in wet year, normal year, and dry year scenarios, respectively. It can be seen that the prediction pollution of pollution in the dry 2025 scenario was the most serious.

(3) Taking 2025 as the example year for which to assess prospective mitigations of NO$_3$-N pollution of groundwater, three new groundwater exploitation regimes were devised. A model for the optimal allocation of water resources was applied to the task of optimizing groundwater use under the three new regimes. The optimization results showed that of these new regimes, groundwater exploitation regime 3 should be adopted in a wet year and a normal year and that regime 1 should be adopted in a dry year.

(4) Taking 2025 as the test year once more, we used ArcGIS to plot the changes in NO$_3$-N pollution before and after optimization in wet year, normal year, and dry year scenarios. The results showed that the concentration of NO$_3$-N decreased significantly after regime optimization. For the northern region of Daxing, which was the most polluted, the concentration of NO$_3$-N decreased from 15.40 mg/L, 15.82 mg/L, and 16.42 mg/L to 6.31 mg/L, 6.61 mg/L, and 7.93 mg/L in the wet year, normal year, and dry year scenarios, respectively.

(5) Taking 2025 as the test year for a third time, the Modflow model was used to simulate changes in the groundwater level before and after optimization and regulation. The results showed that the average groundwater table rose by 0.99 m, 1.80 m, and 2.07 m in the wet year, normal year, and dry year scenarios, respectively.

Because groundwater is buried deep underground, it is more difficult to control than surface water. To reduce groundwater overexploitation while ensuring water safety and adequacy of supply, the use of surface water, reclaimed water, and transferred water should be increased. To ensure groundwater quality, it is necessary to increase the supervision of industries with large emissions of pollutants. To achieve both ends, it is desirable to gradually eliminate or reform industries that currently have high water consumption and/or high pollution emissions. Domestic and industrial water should be fully collected and treated in a centralized manner, and its direct discharge into water bodies should be forbidden.

With the rapid development of both society and the economy, many countries and regions have problems with water shortage and water environment pollution. Especially in areas where surface water is scarce and groundwater is an important water source, achieving sustainable utilization of groundwater resources is of great significance to the development and stability of these areas, especially against the background of the increasing frequency of extreme weather in the world. By combining geographic information technology with machine prediction models, water resource optimization technology, and groundwater numerical simulation, this paper optimizes and controls the groundwater quality and table in the study area, achieving good results. Because of its powerful analytical functions, GIS is of great significance in understanding and grasping temporal and spatial variations in groundwater. In addition, machine learning algorithms have great capacity to generate clear results from complex data. Therefore, it is of great help in realizing the management, supervision, and regulation of groundwater to combine GIS with machine learning algorithms. This paper can provide new ideas and methods for the sustainable utilization of groundwater in areas with both serious groundwater overexploitation and serious groundwater pollution.
Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijgi11101501/s1, Table S1: Errors between the calculated NO3-N concentration values and the monitored NO3-N concentration values for the three machine learning algorithms during the validation period; Table S2: Optimal allocation of water resources under wet year in 2025; Table S3: Optimal allocation of water resources under normal year in 2025; Table S4: Optimal allocation of water resources under dry year in 2025.

Author Contributions: Conceptualization, Baohui Men and Shiyang Yin; methodology, Chen Li; software, Chen Li; Teng Zhang and Ling Wei; validation, Chen Li; formal analysis, Chen Li; investigation, Chen Li; resources, Chen Li; data curation, Chen Li; writing—original draft preparation, Chen Li; writing—review and editing, Chen Li; visualization, Chen Li; supervision, Baohui Men and Shiyang Yin; project administration, Baohui Men and Shiyang Yin; funding acquisition, Baohui Men and Shiyang Yin. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Education of the People’s Republic of China (2019MS028) and the Ministry of Science and Technology of the People’s Republic of China (2016YFC0401406).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be made available upon request. The statistical data can be found here: http://swj.beijing.gov.cn/zwgk (accessed on 20 July 2022).

Acknowledgments: Thanks for the support of the above funding entities and the Laboratory of College of Water Resources and Hydropower Engineering, North China Electric Power University.

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

References
10. Zhou, J.; Wang, Y.Y.; Xiao, F.; Wang, Y.Y.; Sun, L.J. Water Quality Prediction Method Based on IGRA and LSTM. Water 2018, 10, 1148. [CrossRef]
15. Li, J.; Li, X.; Lv, N.Q.; Yang, Y.; Xi, B.D.; Li, M.X.; Bai, S.G.; Liu, D. Quantitative assessment of groundwater pollution intensity on typical contaminated sites in China using grey relational analysis and numerical simulation. Environ. Earth Sci. 2015, 74, 3955–3968. [CrossRef]


45. Nashwan, M.S.; Shahid, S. Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmos. Res.* 2019, 230, 104632. [CrossRef]


