Surrogate Model Application to the Identification of Optimal Groundwater Exploitation Scheme Based on Regression Kriging Method—A Case Study of Western Jilin Province

Yongkai An 1,2, Wenxi Lu 1,2,* and Weiguo Cheng 1,2

1 Key Laboratory of Groundwater Resources and Environment, Ministry of Education, Jilin University, Changchun 130021, China; E-Mails: anyongkai1991@163.com (Y.A.); cwg2015@126.com (W.C.)
2 College of Environment and Resources, Jilin University, Changchun 130021, China

* Author to whom correspondence should be addressed; E-Mail: luwenxi@jlu.edu.cn; Tel.: +86-0431-8850-2610.

Academic Editor: Miklas Scholz

Received: 29 June 2015 / Accepted: 24 July 2015 / Published: 30 July 2015

Abstract: This paper introduces a surrogate model to identify an optimal exploitation scheme, while the western Jilin province was selected as the study area. A numerical simulation model of groundwater flow was established first, and four exploitation wells were set in the Tongyu county and Qian Gorlos county respectively so as to supply water to Daan county. Second, the Latin Hypercube Sampling (LHS) method was used to collect data in the feasible region for input variables. A surrogate model of the numerical simulation model of groundwater flow was developed using the regression kriging method. An optimization model was established to search an optimal groundwater exploitation scheme using the minimum average drawdown of groundwater table and the minimum cost of groundwater exploitation as multi-objective functions. Finally, the surrogate model was invoked by the optimization model in the process of solving the optimization problem. Results show that the relative error and root mean square error of the groundwater table drawdown between the simulation model and the surrogate model for 10 validation samples are both lower than 5%, which is a high approximation accuracy. The contrast between the surrogate-based simulation optimization model and the conventional simulation optimization model for solving the same optimization problem, shows the former only needs 5.5 hours, and the latter needs 25 days. The above results indicate that the surrogate model developed
in this study could not only considerably reduce the computational burden of the simulation optimization process, but also maintain high computational accuracy. This can thus provide an effective method for identifying an optimal groundwater exploitation scheme quickly and accurately.

**Keywords:** western Jilin province; simulation model; LHS; regression kriging method; surrogate model; optimization model

1 Introduction

Simulation and optimization is the main technical method for mechanism analysis, simulation prediction and scheme optimization of actual groundwater systems at present [1–3]. In general, the simulation model is used to simulate the actual groundwater system using a numerical method, which can not only express the intrinsic regularity of the actual groundwater system, but also describes the principles and regularities of physics, chemistry and biology followed by the actual groundwater system. As a replica of the actual groundwater system, it can also depict the response relationship between input and output. In a word, the simulation model is applied to predict the results of changes of natural conditions and human activities [4,5].

The simulation model can solve prediction problems in a given decision-making input, but it cannot indicate to us which decision-making input is able to obtain optimal response. Fortunately, the optimization model (operational research model) can solve the optimization problem, by which the optimal decision-making scheme should be obtained by optimizing the decision-making input scheme under the given objectives and constraints.

It should be noted that the optimization model must be based on the simulation model so as to ensure that the optimization process is followed by the intrinsic principle and regularity of the actual groundwater system (presented as simulation model). Therefore, the simulation model needs to be embedded in the optimization model by certain methods to make it become a part of the optimization model. Embedding method, response matrix method and state transition equation method are three methods commonly used to deal with how to embed and invoke the simulation model in the optimization model. However, all of them have their limitations. For example, the embedding method is used to solve small-scale problems generally, but problems of dimension disaster may be produced when using it to solve large-scale problems; the response matrix method is just used to solve linear problems. Comparing these two methods, the computational burden could be reduced to some extent when using the state transition equation method to solve large-scale, multi-period and nonlinear problems, but it still cost a great deal of time to solve the actual problems. Consequently, it is inappropriate for us to use these three methods to solve the large-scale, multi-period and nonlinear groundwater problems [6].

In recent years, some scholars have proposed a surrogate model of the simulation model, where the surrogate model is invoked by the optimization model directly in the process of the iterative solution of the optimization model, which not only solves large-scale nonlinear groundwater problems, but also reduces the huge computational burden and maintains a high computational accuracy. The frequently used surrogate models include the BP neural network model, the RBF neural network model, the
regression kriging model, the support vector machine model and so on [7–11], which have been proved to substitute the simulation model. However, researches using the surrogate model of the simulation model to solve the groundwater flow optimization problem in reality are sparse, and both the high computational accuracy and low computational burden need to be verified in the process of the surrogate model invoked by the optimization model. The interpolation results of the regression kriging method have been proved to be very effective because it is unbiased and has minimum estimation variance [12,13], thus the regression kriging method is generally selected in the surrogate model of the simulation model [14–18].

Groundwater is one of the most important water resources for domestic water and agricultural water in western Jilin province, and exploiting groundwater excessively could cause geological environmental disasters including land collapse, soil salinization, desertification and so on. Therefore groundwater should be exploited in a reasonable way in life, which requires the search for an optimal exploitation scheme with consideration of allowable groundwater withdrawal and economic benefits.

This paper selected the western Jilin province as the study area and first established a numerical simulation model of groundwater flow in order to reduce the continuous decline of the groundwater table of Daan county in western Jilin province in recent years. Then four exploitation wells were set in the Tongyu county and Qian Gorlos county respectively so as to supply water to Daan county. Utilizing the LHS method in the above eight exploitation wells to extract 40 groups of exploitation schemes, they were then introduced into a numerical simulation model of groundwater flow so as to obtain a drawdown dataset of the groundwater table. A surrogate model of the numerical simulation model of the groundwater flow was established by the regression kriging method using the above exploitation schemes and drawdown dataset of the groundwater table. Using the LHS method again 10 groups of exploitation schemes were obtained which were introduced into the groundwater flow numerical simulation model and surrogate model simultaneously so as to verify the computational accuracy of the surrogate model. In view of geological environmental disasters on continuous decline of the groundwater table and the difference of the groundwater exploitation costs of the eight exploitation wells, an optimization model was established to search an optimal groundwater exploitation scheme using the minimum drawdown of groundwater table and minimum cost of groundwater exploitation as multi-objective functions. The optimal exploitation scheme was achieved in the process of the surrogate model invoked by the optimization model.

2. Study Area and Methods

2.1. Study Area

The western Jilin province is located in the southwest of Songnen Plain where it is situated in the transitional zone from semi-humid to semi-arid area. The geographic coordinate lies between 123°09′–124°22′ east longitude and 44°57′–45°46′ north latitude. The whole region is mainly affected by the inland climate of Inner Mongolia and has typical features of the continental climate. The annual average air temperature is 4.6 °C, and the temperature in the southwest is higher, while the north and east are relatively lower. The average annual precipitation in this region is about 400–500 mm, and its temporal-spatial distribution is extremely uneven due to the influences of geographical location and topography.
The study area is a huge aquifer system which has multiple aquifers, including pore unconfined aquifers and pore confined aquifers (shallow and middle-deep), pore-fracture aquifers of Daan and Taikang formations of Neogene (deep), fracture-pore aquifers of lower and upper cretaceous (deep). The recharge sources of groundwater include precipitation infiltration, river leakage, irrigation, infiltration, and lateral groundwater runoff, in which precipitation infiltration is in the dominant position. The discharge of groundwater includes groundwater evaporation, discharge of river, lateral groundwater runoff and artificial exploitation, in which the groundwater evaporation and artificial exploitation are in the dominant position.

2.2. Methods

2.2.1. Latin Hypercube Sampling Method

LHS is a kind of homogeneous stratified sampling method, developed from the monte-carlo method [19,20]. The basic principle of this method is to divide the whole sample space into several subintervals, and choose randomly a sample in these subintervals. In this way, the sampling results can cover the entire sample space and will be more representative [21–24]. The detailed sampling process is described as follows:

Suppose the dimension of random variable is $\kappa$, $x^i \in [x^i_l, x^i_u]$, $i = 1, 2, \ldots, k$, where $x^i$ is the $i^{th}$ variable, $x^i_l$ and $x^i_u$ are the lower and upper limits of $i^{th}$ variable respectively. Then the process of stratified sampling for a multi-dimensional random variable is described as follows [25–28]:

1. Determining the sampling scale of random variable ($N$).
2. Dividing each variable into $N$ equiprobable intervals, $x^i_l = x^i_0 < x^i_1 < \cdots < x^i_j < x^i_{j+1} < \cdots < x^i_N = x^i_u$, and the probability of each interval is $1/N$.
3. Extracting a random sample from each interval of variable $x_i$, $\kappa$ refers to variables, and then there are $\kappa \times N$ samples.
4. The $N$ samples extracted respectively from variable $x^1$ and $x^2$ are matched randomly without repetition. Then let the matching process go on until the samples extracted from all the variables $x^i$ are completely matched. The eventual matched form is as follows:

$$ X = \begin{bmatrix} x^1_j, x^2_j, \ldots, x^k_j \\ x^1_j, x^2_j, \ldots, x^k_j \\ \vdots \quad \vdots \quad \vdots \\ x^1_j, x^2_j, \ldots, x^k_j \end{bmatrix}, j = 1, 2, \ldots, N \tag{1} $$

2.2.2. Regression Kriging Method

The kriging method is a geostatistics technique which has many different types [29–31], and each type has its own special features. The regression kriging method is a type of kriging method, which was first introduced as a surrogate model by Sacks [32–33]. Many researchers now use the regression kriging method to establish the surrogate model.

The form of regression kriging model is [34–37]
\[
y(x) = f^T(x) \beta + z(x) = \sum_{j=1}^{k} f_j(x) \beta_j + z(x)
\]

where \( \sum_{j=1}^{k} f_j(x) \beta_j \) is the term of deterministic functions, \( \beta_j \) refers to the coefficients of the deterministic function, and \( f_j(x) \) is a known regression function of \( p \)-order, which is usually a polynomial function of 0-order, 1-order, or 2-order.

0-order: \( p = 1, \quad f_1(x) = 1 \)

1-order: \( p = n + 1, \quad f_1(x) = 1, \quad f_2(x) = x_1, \ldots, \quad f_{n+1}(x) = x_n \)

2-order: \( p = (n+1)(n+2)/2, \quad f_1(x) = 1, \quad f_2(x) = x_1, \ldots, \quad f_{n+1}(x) = x_n \)

\[
f_{n+2}(x) = x_1^2, \ldots, \quad f_{2n+1}(x) = x_1x_n \quad \ldots, \quad f_p(x) = x_n^2
\]

\( z(x) \) is a stochastic process with zero-mean, variance \( \sigma^2 \), and covariance

\[
\text{Cov}\left[ z(x_i), z(x_j) \right] = \sigma^2 R(x_i, x_j)
\]

where \( R(x_i, x_j) \) is the correlation function, depending only on the distance vector of \( x_i \) and \( x_j \), not on their locations. The common types of correlation functions are as follow [38]:

- **Exponential function:**
  \[
  R(x_i, x_j) = \exp\left[ -\sum_{k=1}^{n} \theta_k \left| x_i^k - x_j^k \right| \right]
  \]

- **Gauss function:**
  \[
  R(x_i, x_j) = \exp\left[ -\sum_{k=1}^{n} \theta_k \left| x_i^k - x_j^k \right|^2 \right]
  \]

- **Cubic-spline function:**
  \[
  R(x_i, x_j) = \begin{cases} 
  1 - 15\xi_k^2 + 30\xi_k^3 & \text{for } 0 \leq \xi_k \leq 0.2 \\
  1.25(1-\xi_k)^3 & \text{for } 0.2 \leq \xi_k \leq 1, \quad \xi_k = \theta_k \left| x_i - x_j \right| \\
  0 & \text{for } \xi_k \geq 1
  \end{cases}
  \]

where \( \xi_k \) are the unknown parameters, \( x_i^k \) and \( x_j^k \) are the \( k \)th component of sample points \( x \). The Gauss function had been proved feasible in many researches [15,39], thus it was selected as a correlation function in this paper.

For a given set of sample points, \( x = [x_1, x_2, \ldots, x_i, x_j, \ldots, x_m]^T \) (each sample point \( x_i^k \) \( k = 1, 2, \ldots, n \)) and the corresponding response \( y = [y_1, y_2, \ldots, y_i, y_j, \ldots, y_m]^T \), the prediction of the unsampled point’s response \( Y(X) \) can be represented as a function of the unknown parameters \( \beta \) and \( \theta_k, \quad k = 1, 2, \ldots, n \) [40].

\[
Y(X) = f^T(X) \beta^* + r^T(X) R^{-1} \left( Y - f(X) \beta \right)
\]
The Equation (10) is a general form of regression kriging model, which is established by the following steps:

1. \( r(x) = \left[ R(x, x_1), R(x, x_2), \ldots, R(x, x_m) \right] \)

2. \( f(X) \), referring to the known regression functions of \( p \)-order, is calculated through equation 5.

3. \( \beta^* \) is the estimated value of \( \beta \), which is obtained by the generalized least-squares method.

4. The estimated value of \( \sigma^2 \) is obtained by the following equation.

5. The parameter \( \theta_k \) is obtained when the following equation achieves its maximum value, and this method is named as the maximum likelihood estimation method. The basic idea of this method (Maximum Likelihood, ML) is that the most reasonable parameter estimator is determined when extracting an \( n \) group sample observation value from the sample population of the model randomly and making the \( n \) group sample observation value selected from the overall model have a maximum probability.

\[
MLE = -\frac{1}{2} \left( m \ln (\sigma^2) + \ln |R| \right)
\]
2.2.3. Numerical Simulation of Groundwater Flow

The aimed for aquifer located in western Jilin Province is a pore aquifer which is composed of unconfined aquifer and confined aquifer. In the middle of these two aquifers, there is a weakly permeable clayey soil layer.

The top of the simulation area is the unconfined aquifer’s upper boundary where such activities pertaining to water exchange mainly occur as precipitation infiltration, irrigation leakage, evaporation, artificial exploitation, etc. The bottom boundary of the simulation area is the floor of the confined aquifer which is a clayey soil layer and almost has no water exchange. The lateral boundary is generalized based on the boundary of unconfined aquifer (Figure 1), because the unconfined aquifer is found relatively thicker.

The groundwater flow system of the simulation area can be generalized as non-homogeneous, isotropic, and two-dimensional unsteady flow system, which can be shown as follows [41, 42]:

![Figure 1. Types of lateral boundary in study area.](image-url)
\[
\frac{\partial}{\partial x} k \left( H(x, y, t) - Z_b \right) \frac{\partial H(x, y, t)}{\partial x} + \frac{\partial}{\partial y} k \left( H(x, y, t) - Z_b \right) \frac{\partial H(x, y, t)}{\partial y} + W = \mu \frac{\partial H(x, y, t)}{\partial t}
\]

\[\left. H(x, y, t) \right|_{t=0} = H_0(x, y) \]

\[\left. H(x, y, t) \right|_{\Gamma_1} = H_1(x, y, t) \]

\[\left. k(H - Z_b) \frac{\partial H}{\partial n} \right|_{\Gamma_2} = q(x, y, t) \]

where \( H(x, y, t) \) is the groundwater table (m), \( H_0(x, y) \) is the initial water table (m), \( Z_b \) is the elevation of aimed for aquifer floor (m), \( k \) is the hydraulic conductivity (m·d\(^{-1}\)), \( \mu \) is the specific yield (dimensionless), \( W \) is the vertical recharge, discharge strength of unconfined aquifer (m·d\(^{-1}\)), \( \Gamma_1 \) is the boundary of Dirichlet condition, \( \Gamma_2 \) is the boundary of Newman condition, \( q(x, y, t) \) is the recharge and discharge quantity of aquifer per unit width (m·d\(^{-1}\)), \( \vec{n} \) is the direction of outward normal on the boundary, \( D \) is the area for simulation computation.

The groundwater flow direction and parameters partitions of the study area are shown in Figure 2, in which the study area is divided into 13 subareas, and the parameters values of study subareas are in Table 1.

Figure 2. Groundwater flow direction and parameters partitions of study area.
Table 1. Parameters values of study subareas.

| Partitions Parameters | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 |
|-----------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Hydraulic conductivity (m/d) | 14 | 135 | 27 | 17 | 20 | 26 | 9  | 11 | 12 | 13 | 28 | 44 | 15 |
| Specific yield        | 0.09 | 0.23 | 0.10 | 0.12 | 0.18 | 0.08 | 0.08 | 0.10 | 0.11 | 0.08 | 0.09 | 0.10 | 0.15 |
| Specific storage (m$^{-1}$) | 0.008 | 0.008 | 0.009 | 0.008 | 0.008 | 0.008 | 0.009 | 0.008 | 0.008 | 0.008 | 0.007 | 0.008 | 0.008 |

The Groundwater Modeling System (GMS) is made of several modular (MODFLOW, FEMWATER, MT3DMS, RT3D and so on) designed by Environmental Model Laboratory of Brigham Yong University and Test Station of America Army Drainage Engineering. It was used to model groundwater flow and groundwater quality widely [43,44]. MODFLOW modular of GMS (version 9.2.2) software is used to solve the numerical simulation model of groundwater flow, and the algorithm of MODFLOW is a finite difference method [45–47].

2.2.4. Genetic Algorithm

The genetic algorithm (GA) is a computational model on the basis of Darwin's biological evolution theory genetic mechanism, used to search for the optimal solution by simulating natural evolution. It includes three genetic operators of selection, crossover and mutation [48–50]. A flowchart for solving a general problem through the genetic algorithm is shown in Figure 3.

![Figure 3. Process of the genetic algorithm.](image-url)
3. Results and Discussions

3.1. Numerical Simulation of Groundwater Flow

The calibration phase of simulation was selected in the dry season for 181 days from October 1, 2006 to March 31, 2007, taking into consideration that less source and sink are beneficial to identify hydrogeology parameters. The verification phase was selected in the wet season for 182 days from April 1, 2007 to September 30, 2007, on account that more source and sink are beneficial to verify the effectiveness of hydrogeology parameters. The fitting results of computed groundwater table and the actual measured groundwater table are shown in Figures 4 and 5 at the end of the model calibration and verification stage respectively. The equipotential lines of the groundwater table are also shown in Figures 6 and 7 at the end of the model calibration and verification stage respectively.

From Figures 4 and 5 it can be seen that the slopes of the straight lines fitted by the actual measured groundwater table values and computed groundwater table values are all close to 1. From Figures 6 and 7 it can be seen that the fitting results of equipotential lines between the actual groundwater table and the computed groundwater table are very good. The above description means that the actual measured groundwater table values are very close to the computed groundwater table values, the direction of computed groundwater flow field is in accordance with the actual groundwater flow field, the selected hydrogeological conceptual model generalization, partial differential equations and algorithm are reasonable and feasible, and the established numerical simulation model of groundwater flow can objectively and accurately describe the groundwater flow characteristics of the study area. The research results concluded above can give a good foundation for the establishment of a surrogate model.

![Figure 4](image-url)  
**Figure 4.** The fitting chart of the groundwater table between measured and computed of each observation well at the end of the model calibration stage.
Figure 5. The fitting chart of the groundwater table between measured and computed of each observation well at the end of the model verification stage.

Figure 6. The actual and computed equipotential lines of groundwater table at the end of the model calibration stage.
Figure 7. The actual and computed equipotential lines of groundwater table at the end of the model verification stage.

Figure 8. Exploitation wells distribution.
Table 2. Training and validation samples of surrogate model ($q$: m$^3$/d, $s$: m).

| Exploitation Scheme | $W_1$ | $W_2$ | $W_3$ | $W_4$ | $W_5$ | $W_6$ | $W_7$ | $W_8$ |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| $q$ | $s$ | $q$ | $s$ | $q$ | $s$ | $q$ | $s$ | $q$ |
| 1 | 3240 | 0.808 | 7070 | 0.925 | 1551 | 0.609 | 2383 | 0.625 |
| 2 | 5713 | 0.896 | 88 | 0.565 | 7889 | 0.967 | 1480 | 0.588 |
| 3 | 1159 | 0.558 | 5852 | 0.799 | 1677 | 0.584 | 3829 | 0.743 |
| 4 | 3786 | 0.844 | 3100 | 0.911 | 4463 | 1.016 | 6899 | 1.194 |
| 5 | 9330 | 1.178 | 1742 | 0.823 | 367 | 0.732 | 4703 | 0.794 |
| 6 | 3701 | 0.994 | 7724 | 1.151 | 3603 | 0.921 | 4171 | 0.958 |
| 7 | 9759 | 1.732 | 7895 | 1.600 | 4834 | 1.413 | 6581 | 1.405 |
| 8 | 204 | 0.395 | 838 | 0.535 | 6831 | 0.902 | 5248 | 0.960 |
| 9 | 9174 | 1.262 | 2532 | 0.793 | 4706 | 0.854 | 393 | 0.451 |
| 10 | 4479 | 0.933 | 3469 | 0.931 | 5121 | 1.035 | 5637 | 1.084 |
| 11 | 6170 | 1.108 | 4151 | 0.851 | 7308 | 0.975 | 22 | 0.514 |
| 12 | 2540 | 0.798 | 6649 | 1.051 | 1914 | 0.815 | 6084 | 1.080 |
| 13 | 7677 | 1.306 | 4740 | 1.200 | 3494 | 1.136 | 6789 | 1.250 |
| 14 | 4990 | 0.896 | 4252 | 0.836 | 2507 | 0.731 | 3374 | 0.726 |
| 15 | 4106 | 0.919 | 7451 | 1.081 | 71 | 0.680 | 4881 | 0.903 |
| 16 | 477 | 0.497 | 5428 | 0.664 | 4359 | 0.597 | 964 | 0.469 |
| 17 | 1275 | 0.404 | 1098 | 0.399 | 5253 | 0.630 | 1961 | 0.508 |
| 18 | 1887 | 0.562 | 2336 | 0.583 | 5738 | 0.773 | 2653 | 0.661 |
| 19 | 7164 | 1.123 | 3718 | 1.053 | 710 | 0.907 | 7674 | 1.209 |
| 20 | 6440 | 1.314 | 7240 | 1.238 | 6091 | 1.143 | 2830 | 0.916 |
| 21 | 2993 | 0.683 | 2633 | 0.643 | 5531 | 0.796 | 2437 | 0.652 |
| 22 | 4699 | 0.701 | 1290 | 0.568 | 2016 | 0.610 | 3710 | 0.653 |
| 23 | 1570 | 0.510 | 2824 | 0.632 | 3314 | 0.682 | 4471 | 0.802 |
| 24 | 8262 | 1.050 | 277 | 0.794 | 518 | 0.826 | 7581 | 1.100 |
| 25 | 6686 | 1.173 | 3845 | 1.103 | 3868 | 1.118 | 7098 | 1.261 |
| 26 | 2492 | 0.802 | 5630 | 0.867 | 6446 | 0.896 | 1652 | 0.675 |
### Table 2. Cont.

| Exploitation Scheme | $W_1$ | $W_2$ | $W_3$ | $W_4$ | $W_5$ | $W_6$ | $W_7$ | $W_8$ |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|
|                     | $q$   | $s$   | $q$   | $s$   | $q$   | $s$   | $q$   | $s$   |
| **Training samples**|       |       |       |       |       |       |       |       |
| 27                  | 2240  | 0.556 | 3392  | 0.684 | 1352  | 0.596 | 5016  | 0.819 |
| 28                  | 8070  | 1.443 | 6986  | 1.430 | 2211  | 1.175 | 7823  | 1.403 |
| 29                  | 8956  | 1.439 | 4502  | 1.037 | 7600  | 1.144 | 401   | 0.625 |
| 30                  | 8525  | 1.266 | 2056  | 0.832 | 7166  | 1.065 | 1217  | 0.629 |
| 31                  | 7758  | 1.252 | 760   | 0.997 | 7616  | 1.387 | 7268  | 1.355 |
| 32                  | 5984  | 1.025 | 5310  | 0.979 | 984   | 0.734 | 4325  | 0.832 |
| 33                  | 979   | 0.588 | 6563  | 0.915 | 1058  | 0.641 | 5425  | 0.937 |
| 34                  | 532   | 0.537 | 4979  | 0.850 | 2987  | 0.780 | 6210  | 1.058 |
| 35                  | 5259  | 0.848 | 1951  | 0.674 | 4033  | 0.778 | 3028  | 0.678 |
| 36                  | 6849  | 1.162 | 6251  | 0.994 | 2756  | 0.758 | 1174  | 0.550 |
| 37                  | 5164  | 0.838 | 554   | 0.627 | 5950  | 0.919 | 3516  | 0.772 |
| 38                  | 7293  | 1.311 | 6169  | 1.084 | 6217  | 1.034 | 755   | 0.641 |
| 39                  | 3372  | 0.700 | 1530  | 0.582 | 6783  | 0.861 | 2087  | 0.630 |
| 40                  | 9738  | 1.513 | 5192  | 1.275 | 3026  | 1.139 | 5848  | 1.157 |
| **Validation samples**|       |       |       |       |       |       |       |       |
| 1                   | 2032  | 0.512 | 952   | 0.573 | 4167  | 0.783 | 5642  | 0.932 |
| 2                   | 5050  | 1.079 | 6829  | 1.137 | 3100  | 0.926 | 4497  | 0.971 |
| 3                   | 553   | 0.346 | 4431  | 0.487 | 1193  | 0.304 | 866   | 0.300 |
| 4                   | 6490  | 1.368 | 7557  | 1.311 | 6734  | 1.237 | 3334  | 1.015 |
| 5                   | 4363  | 1.079 | 6000  | 1.156 | 6113  | 1.167 | 5259  | 1.153 |
| 6                   | 1630  | 0.475 | 3426  | 0.561 | 2283  | 0.503 | 2514  | 0.535 |
| 7                   | 9206  | 1.509 | 5357  | 1.134 | 7365  | 1.181 | 759   | 0.690 |
| 8                   | 3615  | 0.786 | 2909  | 0.869 | 3417  | 0.932 | 7145  | 1.175 |
| 9                   | 8124  | 1.097 | 1635  | 0.904 | 100   | 0.839 | 7783  | 1.148 |
| 10                  | 7193  | 0.970 | 118   | 0.571 | 5252  | 0.822 | 1685  | 0.531 |

$q$ and $s$ represent the training and validation samples, respectively.
3.2. Surrogate Model of Numerical Simulation Model of Groundwater Flow

The four exploitation wells were set in the Tongyu county and Qian Gorlos county respectively so as to supply water to Daan county shown in Figure 8. The Latin hypercube sampling method was used to obtain 40 and 10 groups of exploitation schemes which were introduced into the numerical simulation model of the groundwater flow to obtain groundwater table drawdown datasets respectively (Table 2).

The \( q \) is the groundwater exploitation quantity and \( s \) is the groundwater table drawdown under the exploitation schemes in Table 2.

MATLAB (2013a) procedure was compiled according to the principle of the regression kriging method. Training samples were used to establish the surrogate model (regression kriging model) and validation samples were used to verify the computational accuracy of the surrogate model. The \( \theta \) are a series of coefficients of gauss functions which determine the precision of the surrogate model. The \( \theta \) are calculated by a genetic algorithm through Equation (16) in Table 3.

| Parameter | \( \theta_1 \) | \( \theta_2 \) | \( \theta_3 \) | \( \theta_4 \) | \( \theta_5 \) | \( \theta_6 \) | \( \theta_7 \) | \( \theta_8 \) |
|-----------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Value     | 0.7922        | 0.9961        | 0.5000        | 1.6490        | 0.6476        | 0.4952        | 0.5737        | 1.0513        |

To investigate the validity of the surrogate model, the validation samples were introduced into the groundwater flow numerical simulation model and the surrogate model respectively. Then, the results of the surrogate model and the numerical simulation model of the groundwater flow were compared with the evaluation indexes including relative error and root mean square error. The value and relative error of the groundwater table drawdown of the simulation model and surrogate model are shown in Figure 9, the mean relative error and root mean square error between the simulation model and the surrogate model are shown in Table 4.

![Figure 9. Cont.](image-url)
Figure 9. Cont.
Figure 9. Value and relative error of groundwater table drawdown of the simulation model and surrogate model.

From Figure 9, the relative error of each well of each scheme is between 0.01% and 4.93%, less than 5%, and the mean relative error of each scheme is between 0.99% and 2.60%. The mean relative error of the 10 validation schemes is 1.87%, which shows that the computed groundwater table drawdown of each well of each scheme by the kriging model is very close to the simulation model. The root mean square error of each scheme is between 1.06% and 2.93%, and the root mean square error of the 10 validation schemes is 2.27%. The results show that the computed groundwater table drawdown of each well of each scheme by the kriging model is not significantly different close to the simulation model, and each scheme also has no significant difference. The above description demonstrates that the surrogate model could substitute the groundwater flow numerical simulation model effectively.

Table 4. The mean relative error and root mean square error between the simulation model and surrogate model.

| Scheme | Mean Relative Error | Root Mean Square Error |
|--------|---------------------|------------------------|
| 1      | 1.21                | 1.48                   |
| 2      | 0.99                | 1.06                   |
| 3      | 1.90                | 2.37                   |
| 4      | 2.60                | 2.93                   |
| 5      | 1.55                | 1.71                   |
| 6      | 2.31                | 2.60                   |
| 7      | 2.25                | 2.58                   |
| 8      | 2.26                | 2.53                   |
| 9      | 2.16                | 2.60                   |
| 10     | 1.52                | 2.15                   |

3.3. Optimization Model

Considering that the eight pumping wells were used to supply water for Daan county water simultaneously, the exploitation quantity of each well needed to be distributed reasonably according to
the minimum average groundwater table drawdown of the eight exploitation wells. However, the cost of water supply of each exploitation well is different. We selected an optimal exploitation scheme which could make the average drawdown of the groundwater table and the cost of the groundwater exploitation less.

To optimize the conditions of water supply for the scheme of groundwater exploitation, a nonlinear multi-objective optimizations model was developed using the minimum average drawdown of the groundwater table and the minimum cost of groundwater exploitation as multi-objective functions, with the exploitation rates as decision-making variables. The optimization model was constructed as follows:

\[
\begin{align*}
S &= \frac{1}{n} \sum_{i=1}^{n} s_i \\
M &= \sum_{i=1}^{n} x_i \cdot q_i
\end{align*}
\]

\[
\begin{align*}
s_i &= f(q_i) \\
0 &\leq q_i \leq 8000, \quad (i = 2, 3, 4, 7, 8) \\
0 &\leq q_i \leq 10000, \quad (i = 1, 5, 6) \\
\sum_{i=1}^{8} q_i &\geq 60000
\end{align*}
\]  \hspace{1cm} (17)

Subject to:  \hspace{1cm} (18)

where \( S \) is the average groundwater table drawdown (m), \( s_i \) is the groundwater table drawdown of the \( i^{th} \) well (m), \( n \) is the numbers of exploitation wells, \( M \) is the water cost ($), \( x_i \) are cost coefficients in the Table 5 ($ \cdot d \cdot m^{-3} \), \( q_i \) are the exploitation rates of the \( i^{th} \) well (m$^3$ d$^{-1}$).

**Table 5.** Water cost coefficients ($ \cdot d \cdot m^{-3} \$).

| Well | \( x_1 \) | \( x_2 \) | \( x_3 \) | \( x_4 \) | \( x_5 \) | \( x_6 \) | \( x_7 \) | \( x_8 \) |
|------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Cost coefficient | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |

In order to solve the nonlinear multi-objective optimization model, the surrogate model (regression kriging program) was loaded into the genetic algorithm and linked with the exploitation rates. The optimal groundwater exploitation strategy through invoking the surrogate model is in the Table 6.

**Table 6.** The optimal exploitation scheme.

| Exploitation Well | \( W_1 \) | \( W_2 \) | \( W_3 \) | \( W_4 \) | \( W_5 \) | \( W_6 \) | \( W_7 \) | \( W_8 \) |
|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Exploitation quantity \((10^3 \cdot m^3 \cdot d^{-1})\) | 7.597 | 7.585 | 7.737 | 7.592 | 7.593 | 7.724 | 7.585 | 7.596 |
| Groundwater table drawdown \((m)\) | 0.400 | 0.411 | 0.422 | 0.426 | 0.427 | 0.607 | 0.675 | 0.929 |
| Water cost \((10^3 \cdot \$ \cdot d \cdot m^{-3})\) | 15.194 | 15.170 | 15.474 | 15.184 | 22.779 | 23.172 | 22.755 | 22.788 |
The main computational burden of the simulation optimization process is the repeated running of the numerical simulation model. The optimization for the study area requires 55 s of CPU time to run every simulation model on a 2.93 GHz Inter CPU and 2 GB RAM PC platform. A conventional simulation optimization model requires 40000 runs of the simulation model. Thus, it would require 2200000 s (25 days) of CPU time to process the overall simulation model run. However, in this study replacing the simulation model with the surrogate model in the optimization process could reduce the process to only 50 simulation model runs during the training and validation of the surrogate model. Thus, it only required 19800 s (5.5 h) of CPU time to complete the simulation model and optimization model run.

4. Conclusions

(1) The groundwater table values calculated by the numerical simulation model of groundwater flow are very close to the actual measured values both at the stage of model calibration and model verification, which demonstrates that the selected hydrogeological conceptual model generalization, partial differential equations and algorithm are reasonable and feasible in the study area, and the established numerical simulation model of groundwater flow can objectively and accurately describe the groundwater flow characteristics of the study area. These research results can provide a good foundation for establishing a surrogate model.

(2) Due to the regression kriging method with accurate approximation ability, the surrogate model results are much closer to that of the numerical simulation model of groundwater flow, and could effectively substitute the numerical simulation model of groundwater flow.

(3) The huge computational burden of coupled operations during simulation and optimization hinders the success of the simulation optimization model in groundwater exploitation. According to this study, replacing the simulation models with surrogate models could reduce the huge computational burden effectively and maintain considerably high accuracy so as to obtain an optimal exploitation scheme.

Acknowledgments

The research was jointly funded by China Geological Survey (No. 1212011140027, No. 12120114027401, No. 12120114056201).

Author Contributions

Yongkai An performed the experimental work, the data collection, the model establishing, and prepared the manuscript; Wenxi Lu had the original idea and supervised the research, Weiguo Cheng performed data analysis and manuscript revision.

Conflicts of Interest

The authors declare no conflict of interest.
References

1. Singh, A. Simulation-optimization modeling for conjunctive water use management. *Agr. Water Manage* **2014**, *141*, 23–29.

2. Zhou, H.B.; Chen, T.B.; Gao, D.; Zheng, G.D.; Chen, J.; Pan, T.H.; Liu, H.T.; Gu, R.Y. Simulation of water removal process and optimization of aeration strategy in sewage sludge composting. *Bioresour. Technol.* **2014**, *171*, 452–460.

3. Alizadeh, H.; Mousavi, S.J. Coupled stochastic soil moisture simulation-optimization model of deficit irrigation. *Water Resour. Res.* **2013**, *49*, 4100–4113.

4. Sandow, M.Y.; Bob, A.; Bruce, B.Y.; Millicent, O.A. Simulation of groundwater flow in a crystalline rock aquifer system in Southern Ghana-An evaluation of the effects of increased groundwater abstraction on the aquifers using a transient groundwater flow model. *Hydrol. Proc.* **2014**, *28*, 1084–1094.

5. Feng, S.Y.; Huo, Z.L.; Kang, S.Z.; Tang, Z.; Wang, F.G. Groundwater simulation using a numerical model under different water resources management scenarios in an arid region of China. *Environ. Earth Sci.* **2011**, *62*, 961–971.

6. Lu, W.X. *Simulation Prediction and Optimization of Groundwater System Management*; Science Press: Beijing, China, 1999.

7. Prebeg, P.; Zanic, V.; Vazic, B. Application of a surrogate modeling to the ship structural design. *Ocean Eng.* **2014**, *84*, 259–272.

8. Bao, P.F.; Miao, W.D.; Xie, R.; Shi, Y.J. A BP-ANN based surrogate modeling for predicting engineering analysis of forging press. *Adv. Mat. Res.* **2014**, *915–916*, 987–991.

9. Heng Pau, G.S.; Zhang, Y.Q.; Finsterle, S.; Wainwright, H.; Birkholzer, J. Reduced order modeling in iTOUGH2. *Comput. Geosci.* **2014**, *65*, 118–126.

10. Angelo, C.; Gianni, D.P.; Vittorio L. Support vector machines for surrogate modeling of electronic circuits. *Neural. Comput. Appl.* **2014**, *24*, 69–76.

11. Fukumitsu, K.; Yabe, T.; Ogata, Y.; Oami, T.; Ohkubo, T. A new surrogate modeling technique combining Kriging and polynomial chaos expansions-Application to uncertainty analysis in computational dosimetry. *J. Comput. Phys.* **2015**, *286*, 103–117.

12. Tapoglou, E.; Karatzas, G.P.; Trichakis, I.C.; Varouchakis, E.A. A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation. *J. Hydrol.* **2014**, *519*, 3193–3203.

13. Modis, K.; Sideri, D. Spatiotemporal estimation of land subsidence and ground water level decline in West Thessaly basin, Greece. *Nat. Hazards* **2015**, *76*, 939–954.

14. Hyungil, K.; Seulgi, Y.; Seongim, C. Numerical investigation for erratic behavior of Kriging surrogate model. *J. Mech. Sci. Technol.* **2014**, *28*, 3697–3707.

15. Xia, B.; Ren, Z.; Koh, C.S. Utilizing kriging surrogate models for multi-objective robust optimization of electromagnetic devices. *IEEE Trans. Magn.* **2014**, *50*, 693–696.

16. Bilicz, S.; Lambert, M.; Gyimothy, S.; Pavo, J. Solution of inverse problems in nondestructive testing by a Kriging-based surrogate model. *IEEE Trans. Magn.* **2012**, *48*, 495–498.

17. David, J.J.T.; Andy, J.K. Performance of an ensemble of ordinary, universal, non-stationary and limit Kriging predictors. *Struct. Multidiscip. Optim.* **2013**, *47*, 893–903.
18. Horowitz, B.; Bastos Afonso, S.M.; de Mendonca, P.C.V. Surrogate based optimal waterflooding management. *J. Pet. Sci. Eng.* 2013, 112, 206–219.

19. Goda, T.; Sato, K. History matching with iterative Latin hypercube samplings and parameterization of reservoir heterogeneity. *J. Pet. Sci. Eng.* 2014, 114, 61–73.

20. Jafar, R.; Masoud, E. Latin hypercube sampling applied to reliability-based multidisciplinary design optimization of a launch vehicle. *Aerosp. Sci. Technol.* 2013, 28, 297–304.

21. Ficklin, D.L.; Yuzhou, L.; Minghua, Z. Climate change sensitivity assessment of streamflow and agricultural pollutant transport in California's Central Valley using Latin hypercube sampling. *Hydrol. Process* 2013, 27, 2666–2675.

22. Shu, Z.; Jirutitijaroen, P. Latin hypercube sampling techniques for power systems reliability analysis with renewable energy sources. *IEEE Trans. Power Syst.* 2011, 26, 2066–2073.

23. Kim, J.B.; Hwang, K.Y.; Kwon, B.I. Optimization of two-phase in-wheel impsm for wide speed range by using the kriging model based on latin hypercube sampling. *IEEE Trans. Magn.* 2011, 47, 1078–1081.

24. Li, C.Q.; Wang, W.; Xiong, J.Z.; Chen, P.Y. Sensitivity Analysis for Urban Drainage Modeling Using Mutual Information. *Entropy* 2014, 16, 5738–5752.

25. Schumaker, M.F.; Kramer, D.M. Comparison of monte carlo simulations of cytochrome b(6)f with experiment using latin hypercube sampling. *Bull. Math. Biol.* 2011, 73, 2152–2174.

26. Van Camp, M.; Walraevens, K. Pumping test interpretation by combination of Latin hypercube parameter sampling and analytical models. *Comput. Geosci.* 2009, 35, 2065–2073.

27. Sallaberry, C.J.; Helton, J.C.; Hora, S.C. Extension of Latin hypercube samples with correlated variables. *Reliab. Eng. Syst. Safety* 2008, 93, 1047–1059.

28. Kweon, K.E.; Lee, J.H.; Ko, Y.D. Neural network based modeling of HfO2 thin film characteristics using Latin Hypercube Sampling. *Expert. Syst. Appl.* 2007, 32, 358–363.

29. Kitanidis, P.K. *Introduction to Geostatistics: Applications in Hydrogeology*; Cambridge University Press: Cambridge, UK, 1997.

30. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: Oxford, UK, 1997.

31. Varouchakis, E.A.; Hristopulos, D.T. Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. *Environ. Monit. Assess.* 2013, 185, 1–19.

32. Kuhnt, S.; Steinberg, D.M. Design and analysis of computer experiments. *Ada Adv. Stat. Anal.* 2010, 94, 307–309.

33. Huang, D.; Allen, T.T.; Notz, W.I.; Miller, R.A. Sequential kriging optimization using multiple-fidelity evaluations. *Struct. Multidiscip. Optim.* 2006, 32, 369–382.

34. Michael, S.W.; David J.P. Use of regression kriging to develop a Carbon: Nitrogen ratio surface for New Zealand. *Geoderma* 2012, 183, 49–57.

35. Hengl, T.; Heuvelink, G.B.M.; Stein, A.A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 2004, 120, 75–93.

36. Hengl, T.; Heuvelink, G.B.M.; Rossiter, D.G. About regression-kriging: from equations to case studies. *Comput. Geosci.* 2007, 33, 1301–1315.
37. Varouchakis, E.A.; Hristopulos, D.T. Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables. *Adv. Water Resour.* **2013**, *52*, 34–49.

38. Ryu, J.S.; Kim M.S.; Cha K.J. Kriging interpolation methods in geostatistics and DACE model. *Ksme. Int. J.* **2002**, *16*, 619–632.

39. Luo, J.N.; Lu, W.X. Sobol' sensitivity analysis of NAPL-contaminated aquifer remediation process based on multiple surrogates. *Comput. Geosci.* **2015**, *67*, 110–116.

40. Indranil, P.; Saptarshi, D. Kriging Based Surrogate Modeling for Fractional Order Control of Microgrids. *IEEE T. Smart Grid.* **2015**, *6*, 36–44.

41. Bear, *J. Dynamics of Fluids in Porous Media*; Dover Publications: New York, NY, USA, 1988.

42. David, K.T. *Groundwater Hydrology*; Wiley: Hoboken, NJ, USA, 2004.

43. Xu, W.; Du, S.H. Information entropy evolution for groundwater flow system: A case study of artificial recharge in Shijiazhuang City, China. *Entropy* **2014**, *16*, 4408–4419.

44. Du, S.H.; Su, X.S.; Zhang, W.J. Effective storage rates analysis of groundwater reservoir with surplus local and transferred water used in Shijiazhuang City, China. *Water Environ. J.* **2013**, *27*, 157–169.

45. Varouchakis, E.A.; Karatzas, G.P.; Giannopoulos, G.P. Impact of irrigation scenarios and precipitation projections on the groundwater resources of Viannos Basin at the island of Crete, Greece. *Environ. Earth Sci.* **2015**, *73*, 7359–7374.

46. Zekri, S.; Triki, C.; Al-Maktoumi, A.; Bazargan-Lari, R.M. An Optimization-Simulation Approach for Groundwater Abstraction under Recharge Uncertainty. *Water Resour. Manag.* **2015**, *29*, 1–15.

47. Spanoudaki, K.; Stamou, A.I.; Nanou-Giannarou, A. Development and verification of a 3-D integrated surface water–groundwater model. *J. Hydrol.* **2009**, *375*, 410–427.

48. Fayyazi, E.; Ghotbi, B.; Najafi, G.; Hosseinzadeh, B.; Mamat, R.; Hosseinzadeh, J. An ultrasound-assisted system for the optimization of biodiesel production from chicken fat oil using a genetic algorithm and response surface methodology. *Ultrason. Sonochemistry.* **2015**, *26*, 312–320.

49. Baizid, K. Yousnadj, A.; Meddahi, A.; Chellali, R.; Iqbal, J. Time scheduling and optimization of industrial robotized tasks based on genetic algorithms. *Robot Comput-Integr. Manuf.* **2015**, *34*, 140–150.

50. Zhang, L.P.; Wong, T.N. An object-coding genetic algorithm for integrated process planning and scheduling. *Eur. J. Oper. Res.* **2015**, *244*, 434–444.