Models of Hydraulic Factors Analysis Based on Genetic Programming for South-to-North Water Diversion Middle Route Project

Xiaonan Chen 1*, Haitao Chen 2, Yanguo Jin 1, Xiaobo Feng 3, Yanjun Ma 1, and Fang Guo 1

1 Construction and Administration Bureau of South-to-North Water Diversion Middle Route Project, Beijing, 100038, China
2 School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou, Henan Province, 450045, China
3 Supervision Centre of South-to-North Water Diversion Project, Beijing, 100038, China
* Corresponding author’s e-mail: chenxiaonan@nsbd.cn

Abstract. The South-to-North Water Diversion Middle Route Project has long water transmission line, complicated operation conditions for water conservancy dispatch, and high water diversion demands. It is important to calculate and analyse the correlation among hydraulic factors for water diversion. At present, the traditional hydraulic empirical formulas are used to evaluate relation of hydraulic elements in Middle Route Project, the parameters need to be manually corrected by measured data during operation, and the flexibility is poor. Model based on genetic programming is suggested for data mining of water conveyance dispatch. Correlativity function can be established automatically by genetic operations including selection, crossover and mutation. The model is applied into the discharge calculation and analysis of water surface curve for typical gate station and canal pool in Middle Route Project. In discharge calculation, the relation function between head difference, gate opening and flow coefficient can be found automatically with genetic programming. Similarly, the relation between downstream water depth, flow and upstream water depth can be achieved by the suggested model. It is shown that the proposed model based on genetic programing would be effective in nonlinear regression for hydraulic factors.

1. Introduction

The South-to-North Water Diversion Project that is planned east, middle, and west three lines is a major strategic infrastructure to relieve the severe water shortages in northern China, and promote the optimal allocation of water resources. In November 2013 and December 2014, the East Route Project and the Middle Route Project were put into operation respectively. The Middle Route Project is 1432 km long, and diverts water from Danjiangkou Reservoir, and provides high quality living water to Henan Province, Hebei Province, Tianjin and Beijing. Since the beginning of water conveyance, comprehensive benefits produced from project have become increasingly apparent. By the end of February 2019, the Middle Route Project has transported more 20 billion cubic meters of water.

At present, there are many research results about long distance water diversion at home and abroad. Corriga et al. (1982) [1] improved the constant volume control method for open channel operation.
David et al. (1998) [2] reviewed the commonly used canal control methods. Clemmens et al. (2005) [3] derived simulation of automatic canal control system. Shahrokhnia et al. (2006) [4] analysed dimensionless stage-discharge relationship about radial gates. Cui et al. (2009) [5] studied the canal control of large scale water transfer project. Guan et al. (2008) [6] calculated and analysed the difference between steady flow method and unsteady flow method for long canal system. Yao et al. (2008) [7] employed discharge active compensation to study the operation of canal system. Zhang et al. (2014) [8] developed the partition schedule to deal with the control of large water transfer channel.

In recent years, a great deal of research has been focused on the operation for the South-to-North Water Diversion Middle Route Project. Fang et al. (2014) [9] developed one-dimensional hydraulic and water quality model to simulate the water pollution emergency. Cao et al. (2016) [10] improved the control strategy by coupling water level and flow objects for the Middle Route.

The project is a linear project with many diversion gates and no reservoirs along the line. Water supply targets are achieved through close coordination among gates, and the technology for water diversion is difficult. Hydraulic factors analysis is the foundation of water conveyance dispatching model, such as gate discharge calculation, water surface curve computation and so on. The models of factors analysis are usually based on the traditional hydraulic methods, and empirical parameters in formula are revised according to the measured data during operation. Liu et al. (2009) [11] analysed and compared the existing calculating formulas for radial gate flow in Middle Route Project, and made suggestions for application. Jiang et al. (2008) [12] used radial basis function neural network to deal with discharge calculation. As for the water curve calculation, it is made usually by hydraulics and numerical analysis. For example, Zhang et al. (2005) [13] used a new iteration method to calculate water level of gradually varied steady flow. Wen et al. (2014) [14] applied numerical integration method into water curve calculation in parabola shaped canal.

Traditional hydraulic methods are mature, but hydraulic parameters such as roughness, flow coefficient need to be selected empirically, and parameters need to be constantly revised according actual data during operation, which is inconvenient to operate. The neural network model needs to be based on a large of complete data. However, it is impossible to achieve the hydrological data that include all range to want in practice. The representative data is few or the range of the measured data is small actually. If the network is directly applied into fitting the hydraulic relation, the model’s generalization ability is not high.

In view of the above problems, combined with the experience about water diversion in Middle Route, this paper proposes to analyse and calculate hydraulic factors based on genetic programming. With the updating of measured data, model is automatically revised and refreshed.

2. Genetic programming method
Genetic programming, as a new automatic modelling technology, has been widely used in many fields such as control, predication, and recognition and so on. Li et al. (2005) [15] applied genetic programming into statistical modelling. Chen et al. (2006) [16] proposed the model of water production function with genetic programming. The basic frame of the classic genetic algorithm is applied into genetic programming, and the optimal function structure can be found automatically by genetic operations of selection, crossover and mutation. The effect in non-linear fitting through genetic programming is obviously superior to traditional methods in accuracy. Function expressions are considered chromosomes that are described by tree structure. The elements that make up the expression come from the terminal set T and function set F. The function set includes operation symbols and functions, etc., and the terminal set includes variables and constants.

Similar to genetic algorithm, genetic programming generates randomly a group of chromosomes as initial population, evaluates the each chromosome with evaluation function, and makes the chromosome evolve in the direction of superiority with the evolution process by genetic operations. The main steps of genetic programming can be described as follows:

Step 1: Establish object function. Suppose that the dimension of input samples is m, the dimension of output samples is 1, the number of input samples is n, then input samples set and output samples set
can be described separately as $X=\{(x_{11}, x_{12}, \ldots, x_{1m}), (x_{21}, x_{22}, \ldots, x_{2m}), \ldots, (x_{n1}, x_{n2}, \ldots, x_{nm})\}$, and $Y=\{y_1, y_2, \ldots, y_n\}$. The aim is to find the optimal expression $G(c, x_{k1}, x_{k2}, \ldots, x_{km})$ which minimizes the fitting error. Object function can be established as

$$f = \sum_{k=1}^{n} |G(c, x_{k1}, \ldots, x_{km}) - y_k|$$

Where $c$ is a constant.

Step 2: Encode genetic programming. Generate function set $F$ and terminal set $T$. In this paper, the elements in function set are set to arithmetic operators including $[+, -, \times, /]$, trigonometric and antitrigonometric functions $\{\sin{x}, \cos{x}, \arctan{x}, \arccot{x}\}$. The terminal set elements are selected variable $x$ and constant $c$. Encode elements of the union $D=T\cup F$ uniformly, and the encoding scheme is indicated in table1. When code 0 is selected, generate randomly a real number form interval $[0, 1]$. If code is 1, choose randomly $x_i$, $i=1, 2, \ldots, m$. Expression that is encoded is described by tree structure in computer.

| Elements | Code |
|----------|------|
| $c$      | 0    |
| $x$      | 1    |
| $+$      | 2    |
| $-$      | 3    |
| $\times$ | 4    |

| Elements | Code |
|----------|------|
| $/$      | 5    |
| $\sin{x}$ | 6    |
| $\cos{x}$ | 7    |
| $\arctan{x}$ | 8    |
| $\arccot{x}$ | 9    |

For example, expression $b^2-4ac$ can be described as follows:

Figure 1. Chromosome described by tree structure

Step 3: Initialize population. Suppose that the count of individuals is $N$. Generate $N$ trees by selecting randomly root node from function set $F$, middle nodes from union $D$, leaf nodes from terminal set $T$. In order to facilitate the analysis and calculation, the depth of the tree is generally controlled at 4–6 layers.

Step 4: Evaluate individual. Calculate function value of individual according to objective function $G$, and get sequence $f(i)$, $i=1, 2, \ldots, N$. Sort sequence $f(i)$ by ascending, and the fit degree of individual can be calculated as follows:

$$F(i) = (f^2(i) + 0.001)^{1/3}$$

Step 5: Genetic operations. Selection operation can be realized by optimizing individuals according to a certain probability as follows:

$$s(i) = \frac{F(i)}{\sum F(i)}$$

Let

$$p(0) = 0, p(i) = \sum_{k=1}^{i} s(k) \quad i=1, 2, \ldots, N$$

Generate a random number $u$ form interval $[0, 1]$. If $u$ is in the range $[p(i-1), p(i)]$, The $i$th chromosome will be selected. In order to enhance the global search ability of algorithm, the five top individuals will be selected directly in each evolution calculation.

Crossover operation can be realized by exchanging the subtrees of tow parent trees. Generate intersections in tow parent trees, and then exchange the subtrees with the intersections as the root nodes.
Mutation operation can be achieved by changing subtrees. Select randomly a node in parent tree as mutation point, and generate a new subtree with mutation point as root node according to step 3.

During evolution process, genetic programming performs genetic operations with corresponding probability as follows:

Set selection probability is $p_s$, crossover probability equals $p_c$, then mutation probability $p_m=1-p_s-p_c$. Generate randomly a number $u$ in [0, 1]. If $u \leq p_s$, selection operation will be done. If $p_s < u \leq p_c$, crossover operation will be executed. If $u > p_s+p_c$, mutation operation will be performed. Repeat these procedures $N$-5 times and finish one time evolution.

Step 6: Record the best individual. Take new generation as parent generation, and go to step 4. Repeat these steps until the number of evolutionary iterations is greater than the default or the value of the objective function is satisfied with the default accuracy. Take the optimal individual as the final solution.

3. Calculation model of hydraulic factors with genetic programming

3.1. Discharge calculation model based on genetic programming

At present, the discharge form of radial gate in Middle Route is usually orifice flow with submerged discharge, and corresponding formula with traditional hydraulics can be described as follows:

$$Q = \sigma_s \mu b e \sqrt{2g\Delta H}$$

(5)

where $Q$ is water discharge through sluice, m$^3$/s; $\sigma_s$ is submergence coefficient; $\mu$ is flow coefficient; $b$ is width of sluice, m; $e$ is gate opening, m; $\Delta H$ is hydraulic head difference, m.

Combine submergence coefficient and flow coefficient, and the comprehensive flow coefficient $m$ is obtained. The calculation formula is changed as follows:

$$Q = mbe \sqrt{2g\Delta H}$$

(6)

For sluice in Middle Route Project, the engineering parameters such as sluice width and gate bottom elevation are known. In process of dispatch and operation, water levels at the upstream and downstream of sluice gate, gate opening and flow can be acquired in real time by automation system.

As for control sluice, its chamber width is fixed, and the comprehensive flow $m$ is a function of gate opening and hydraulic head difference. During the operation of water conveyance, the data of water level, gate opening, and discharge can be achieved in real time by monitoring system, and the comprehensive flow coefficient can be calculated with the following equation:

$$m = \frac{Q}{be \sqrt{2g\Delta H}}$$

(7)

Based on genetic programming, the discharge calculation model is established by the data of gate opening, water level and the comprehensive flow coefficient obtained. In the model, there are two input variables corresponding to hydraulic head difference and gate opening, and one output variable corresponding to comprehensive flow coefficient. The regression model is set up based on genetic programming according to sample data. When the gate opening and water depth are given, the comprehensive flow coefficient can be computed with regression function established by genetic programming, and then the discharge can be obtained by formula (6).

3.2. Analysis model of water surface curve by genetic programming

Water conveyance dispatch is carried out according to constant downstream depth operation in Middle Route. The part between the two control sluices is called a canal pool. If the water supply flow changes, the water surface curve changes accordingly, and the downstream water level of pool is still near the objective water level, but the upstream depth will change correspondingly according to the change of the flow. The analysis and calculation of water surface curve is important to water diversion dispatch.
For a certain pool, suppose that there is no outward water diversion from the canal section. The size parameters of canal and buildings are fixed. Because the constant downstream water level operation is adopted in Middle Route, the water level before the sluice can be controlled artificially, and the calculation of water surface curve is to compute the upstream water level of pool according to the known downstream water level and flow. The equation of water level variation along the course for steady non-uniform flow is indicated as follows:

$$\frac{dz}{ds} = (\alpha + \xi) \frac{dv^2}{2g} + \frac{Q^2}{K^2}$$

(8)

where \(z\) is water level, m; \(s\) is the length along the canal pool, m; \(v\) is the water velocity, m/s; \(\alpha\) is correction coefficient of kinetic energy; \(\xi\) is local head loss coefficient; \(K\) is discharge modulus, and can be calculated as follows:

$$K = AC \sqrt{R}$$

(9)

where \(A\) is the area of flow cross-section, m\(^2\); \(C\) is Chezy coefficient; \(R\) is hydraulic radius, m.

The above differential equation is discretized by the finite difference method, and the following formula with difference form is obtained.

$$Z_i = Z_{i+1} + \alpha_{i+1} \frac{v_{i+1}^2}{2g} - \alpha_i \frac{v_i^2}{2g} + h_i + h_j$$

(10)

where \(i, i+1\) represent respectively upstream and downstream sections; \(h_i, h_j\) are friction head loss and local head loss respectively.

For canal pool of the Middle Route Project, the parameters of canal bottom width, side slope coefficient, canal bottom slope, and parameters of building are all known conditions, forming the database. In the process of operation, water level and flow data are acquired in real time by automatic system. For the case of no outward water diversion in the pool, the downstream water level and flow are taken as input variables, and the upstream water level of canal section is taken as output variable, and regression function can be achieved by genetic programming according to sample data. If there is water diversion in the canal section, the outward flow of each diversion outlet is also taken as the input variable. The upstream water level can be computed by genetic programming according to known the downstream water level and flow.

4. Application of the genetic programming

4.1. Calculation of discharged flow of control sluice

The South-to-North Water Diversion Middle Route Project is long, and there are many diversion gates. The dispatching of water conveyance is realized through the coordination of gates. When the flow of water conveyance changes, the relevant gates are adjusted to make the water conveyance system gradually change from one original state to another stable target state. There are 64 control gates set up in Middle Route Project. During the water transfer process, the object of water supply can be realized by the analysis of water balance and the calculation of flow.

In this paper, the effectiveness of discharge model with genetic programming is validated by taking the Cihe River sluice as an example. The inverted siphon of Cihe River is one of the most important water conveyance structures in Middle Route. The control sluice of Cihe River is located in the inverted siphon outlet. There are three radial gates in the sluice, each of which has a width of 6 m. The bottom elevation of the sluice is 66.721 m.

As an important gate station, control sluice of Cihe River has been used since 2008, and accumulated abundant monitor data. It is representative to choose Cihe River sluice to analyse.

The partial measured data of water depth, gate opening (the opening value of each gate is the same) and discharge of sluice (flow through 3 gate holes) at a certain period are shown in the following table:
The selected probability is 0.1, the crossover probability is 0.7, and the mutation probability is 0.2. The population size is 100, the evolution generation is 200, the selected probability is 0.1, the crossover probability is 0.7, and the mutation probability is 0.2. The regression function achieved by genetic programming is indicated as follows:

\[
\text{Discharge (m}^3/\text{s)} = \cos(\sin(\text{art} \tan(x_1))) \times 0.5068 + 0.0001
\]

where \(x_1\) and \(x_2\) are normalized head difference and opening data respectively.

The genetic programming model is compared with the traditional hydraulic method and BP neural network model established directly by water depth, opening and measured flow. In this paper, the

| Time interval | Water depth before gate (m) | Water depth after gate (m) | Head difference (m) | Gate opening (m) | Discharge (m$^3$/s) | Comprehensive flow coefficient |
|---------------|-----------------------------|---------------------------|---------------------|-----------------|---------------------|-------------------------------|
| 1             | 6.662                       | 5.934                     | 0.728               | 1.050           | 55.920              | 0.783                         |
| 2             | 6.647                       | 5.929                     | 0.718               | 1.050           | 56.840              | 0.802                         |
| 3             | 6.687                       | 5.919                     | 0.768               | 1.000           | 54.930              | 0.787                         |
| 4             | 6.707                       | 5.889                     | 0.818               | 0.850           | 47.070              | 0.768                         |
| 5             | 6.697                       | 5.849                     | 0.848               | 0.650           | 47.560              | 0.997                         |
| 6             | 6.797                       | 5.779                     | 1.018               | 0.650           | 37.390              | 0.715                         |
| 7             | 6.807                       | 5.689                     | 1.118               | 0.500           | 29.800              | 0.707                         |
| 8             | 6.857                       | 5.599                     | 1.278               | 0.500           | 22.800              | 0.698                         |
| 9             | 6.907                       | 5.509                     | 1.438               | 0.500           | 19.800              | 0.698                         |
| 10            | 6.957                       | 5.419                     | 1.598               | 0.500           | 16.800              | 0.698                         |

In order to facilitate the calculation, the input data are normalized by the following formula. As for comprehensive flow coefficient, the value is already in the interval [0, 1], so it is not necessary to transform.

\[
z_1 = \frac{z}{z_{\text{max}}}
\]

where \(z_1\) is the normalized input data; \(z\) is the original data before conversion; \(z_{\text{max}}\) is the maximum value for conversion according to the range of input data. As for head difference, the corresponding maximum is selected 1.2 m, and the value of \(z_{\text{max}}\) can be chosen 1.5 m for the gate opening. The sample data for validation are shown as follows:

| Time interval | Water depth before gate (m) | Water depth after gate (m) | Head difference (m) | Gate opening (m) | Measured discharge (m$^3$/s) | Comprehensive flow coefficient |
|---------------|-----------------------------|---------------------------|---------------------|-----------------|-----------------------------|-------------------------------|
| 1             | 6.542                       | 5.501                     | 1.041               | 0.080           | 5.330                       | 0.819                         |
| 2             | 6.509                       | 5.489                     | 1.020               | 0.140           | 8.570                       | 0.761                         |
| 3             | 6.627                       | 5.604                     | 1.023               | 0.500           | 27.890                      | 0.692                         |
| 4             | 6.607                       | 5.604                     | 1.003               | 0.500           | 28.230                      | 0.707                         |
| 5             | 6.617                       | 5.609                     | 1.008               | 0.500           | 28.130                      | 0.703                         |
| 6             | 6.617                       | 5.599                     | 1.018               | 0.450           | 25.760                      | 0.712                         |
| 7             | 6.512                       | 6.006                     | 0.506               | 1.340           | 59.160                      | 0.779                         |
| 8             | 6.612                       | 5.589                     | 1.023               | 0.420           | 23.210                      | 0.686                         |
| 9             | 6.647                       | 5.929                     | 0.718               | 1.050           | 56.840                      | 0.802                         |
| 10            | 6.697                       | 5.849                     | 0.848               | 0.650           | 47.560                      | 0.997                         |

Calculate discharge of Cihe River sluice with genetic programming. The regression model can be established automatically according to training sample. Set the population size is 100, the evolution generation is 200, the selected probability is 0.1, the crossover probability is 0.7, and the mutation probability is 0.2. The regression function achieved by genetic programming is indicated as follows:
water depth before and after sluice, gate opening are taken as the input of BP network, and the input layer has three neurons. The measured flow is taken as the output, and the output layer is one neuron. Set the number of hidden layer neurons is 10, the learning efficiency is 0.5, the training number is 2000, according to many trials. The results of comparison and analysis are shown as follows:

| Time interval | Measured flow (m³/s) | Hydraulic method | Genetic programming | Neural networks |
|---------------|----------------------|------------------|---------------------|----------------|
|               |                      | Calculate d flow (m³/s) | Error (%)           | Calculate d flow (m³/s) | Error (%)           | Calculate d flow (m³/s) | Error (%)           |
| 1             | 5.33                 | 4.90             | 8.07               | 4.42             | 17.07               | 11.81             | 121.58               |
| 2             | 8.57                 | 8.85             | 3.27               | 7.80             | 8.98               | 13.34             | 55.66                |
| 3             | 27.89                | 36.50            | 30.87              | 28.94            | 3.76               | 26.48             | 5.06                 |
| 4             | 28.23                | 36.08            | 27.81              | 28.85            | 2.20               | 26.49             | 6.16                 |
| 5             | 28.13                | 36.18            | 28.62              | 28.88            | 2.67               | 26.48             | 5.87                 |
| 6             | 25.76                | 32.29            | 25.35              | 25.98            | 0.85               | 24.20             | 6.06                 |
| 7             | 59.16                | 74.22            | 25.46              | 66.46            | 12.34              | 83.11             | 40.48                |
| 8             | 23.21                | 29.96            | 29.08              | 24.20            | 4.27               | 22.91             | 1.29                 |
| 9             | 56.84                | 69.19            | 21.73              | 56.86            | 0.04               | 61.13             | 7.55                 |
| 10            | 47.56                | 43.63            | 8.26               | 36.49            | 23.28              | 34.35             | 27.83                |

According to the calculation results above the table, analyze the effectiveness of each method.

(1) Comparing the calculated and measured flow of each method, average error of genetic programming is 7.55%, while that of hydraulic method and BP neural network method is 20.85% and 27.75%, respectively. The error of the model proposed in this paper is lower than that of the other two methods. The average error is calculated by the following formula:

\[ E = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{q_i - q}{q} \right| \]

(13)

where \( E \) is average error; \( n \) is the number of trial; \( q \) is the measured flow, m³/s; \( q_i \) is the calculated flow, m³/s.

(2) According to the calculation results of lines 3, 4, 5, 6 and 8 in the table above, average error of genetic programming is 2.75%, that of BP neural network method is 4.89%, and that of hydraulic method is 28.35%. The calculation error of genetic programming is obviously less than of other methods. The input values in the above samples are very close to the input values in the training sample set, which shows that the genetic programming has good fitting ability. When the input data in validation samples is within the training sample range, it can achieve satisfied output reasoning effect. At the same time, it can also be seen that BP networks has a good fitting effect.

Figure 2. Fitting effect of genetic programming
(3) According to the calculation results of rows 1 and 2 above the table, the average error of the neural networks method is 88.62%, while the average error of the genetic programming is only 13.03%. This is because the input range of the above test samples is quite different from that of the training samples, which shows that BP networks has a good fitting effect, but its generalization ability is not high. In addition, the average error of BP networks calculated by all test samples is the largest, which is mainly caused by the two calculation results of rows 1 and 2. The genetic programming combined with hydraulic method presented in this paper has achieved good results, which indicates that the method of fitting the comprehensive discharge coefficient by genetic programming not only has high accuracy, but also improves the generalization ability of the model.

(4) The genetic programming can automatically train and adjust the model structure with the update of sample data. Compared with the traditional hydraulic method, it can avoid the inconvenience of modifying the parameters in the source code of program. It has strong adaptability, convenience and flexibility, and can save manpower and financial resources in the calibration of hydraulic parameters.

4.2. Calculation of water surface curve
In this paper, canal pool from Huangjinhe River inverted siphon to Caodunhe River aqueduct is selected as the study section. The length of the canal section is 21.7 km. There is no outlet for water diversion, but it contains two buildings, Tuojiahe River inverted siphon and Jiahe River aqueduct.

Figure 3. Sketch of the study canal section

The partial measured data of upstream and downstream water level and flow in a certain period of time are shown in the following table:

| Time interval | Upstream water level (m) | Downstream water level (m) | Water conveyance flow (m$^3$/s) |
|---------------|--------------------------|----------------------------|-------------------------------|
| 1             | 135.960                  | 135.895                    | 76.570                        |
| 2             | 135.960                  | 135.905                    | 76.560                        |
| 3             | 135.960                  | 135.905                    | 73.750                        |
| 4             | 135.970                  | 135.900                    | 77.420                        |
| 5             | 135.970                  | 135.900                    | 73.990                        |
| 6             | 135.980                  | 135.900                    | 72.780                        |
|               |                         |                            |                               |
| 55            | 136.750                  | 136.050                    | 226.290                       |
| 56            | 136.740                  | 136.050                    | 220.840                       |
| 57            | 136.740                  | 136.040                    | 223.000                       |
| 58            | 136.730                  | 136.040                    | 228.500                       |
| 59            | 136.720                  | 136.020                    | 219.520                       |
| 60            | 136.710                  | 136.020                    | 222.770                       |

Normalize the data in the table above by the following formula:

$$S_i = \frac{S - S_{\text{min}}}{S_{\text{max}} - S_{\text{min}}}$$  \hspace{1cm} (14)

where $S_i$ is the normalized input data; $S$ is the original data before conversion; $S_{\text{max}}$, $S_{\text{min}}$ are the maximum and minimum for conversion according to the range of input data. As for upstream water level, the corresponding values are selected 137.27 m and 135.64 m. The maximum and minimum for downstream water level are selected 136.8 m and 135.64 m respectively. For water conveyance flow,
the maximum is 330 m³/s and the minimum is 0 m³/s. The sample data for validation are shown as follows:

| Time interval | Upstream water level (m) | Downstream water level (m) | Water conveyance flow (m³/s) |
|---------------|--------------------------|-----------------------------|-----------------------------|
| 1             | 135.960                  | 135.895                     | 76.570                      |
| 2             | 135.960                  | 135.905                     | 76.560                      |
| 3             | 135.960                  | 135.905                     | 73.750                      |
| 4             | 135.970                  | 135.900                     | 77.420                      |
| 5             | 135.970                  | 135.900                     | 73.990                      |
| 6             | 135.980                  | 135.900                     | 72.780                      |
| 7             | 135.985                  | 135.890                     | 77.210                      |
| 8             | 135.850                  | 135.705                     | 87.910                      |
| 9             | 135.850                  | 135.700                     | 85.810                      |
| 10            | 135.850                  | 135.700                     | 85.810                      |

According to the training samples data, the genetic programming model is used to fit the regression function. In this example, set the population size is 50, the selection probability is 0.2, the crossover probability is 0.7, the mutation probability is 0.1, and evolution generation number is 200. The regression function is follows:

\[ y = \text{arc} \cot(x_2 + 0.0001) \times (x_1 / 0.8746)^2 - (x_2 - 0.0908) \times \text{arc} \cot(-0.8911) \]  

(15)

where \( y \) is the normalized upstream water level of canal section. \( x_1 \) is the transformed downstream water level of canal pool. \( x_2 \) is the water conveyance flow after conversion. According to the validation data, using the equation obtained by genetic programming, the upstream water level is calculated with the downstream water level and water conveyance flow. The results are shown in the following table:

| Time interval | Calculated upstream water level (m) | Measured upstream water level (m) | Absolute error (m) |
|---------------|-------------------------------------|-----------------------------------|--------------------|
| 1             | 135.924                             | 135.950                           | 0.026              |
| 2             | 135.930                             | 135.880                           | 0.050              |
| 3             | 135.989                             | 136.015                           | 0.026              |
| 4             | 135.969                             | 135.985                           | 0.016              |
| 5             | 136.026                             | 136.000                           | 0.026              |
| 6             | 136.034                             | 136.020                           | 0.014              |
| 7             | 136.124                             | 136.160                           | 0.036              |
| 8             | 136.135                             | 136.150                           | 0.015              |
| 9             | 136.197                             | 136.200                           | 0.003              |
| 10            | 136.661                             | 136.710                           | 0.049              |
According to the calculation results of the above table, the maximum error is 0.05 m and the minimum error is only 0.003 m, which can greatly meet the actual needs of operation.

5. Conclusions
Combined with the practice of the South-to-North Water Diversion Middle Route Project, the model of discharge flow and water surface curve based on genetic programming are put forward. The models realize the fitting of nonlinear relationship between hydraulic factors by genetic programming and conduct reasoning calculation by trained model. The calculation examples of typical gate stations and canal section in Middle Route show that the models with genetic programming are high accurate convenient and adaptable, and have strong practical application value and popularization significance.

References
[1] Corriga G., Fanni A., Sanna S., et al. (1982) A constant-volume control method for open channel operation. International Journal of Modelling & Simulation, 2: 108-112.
[2] David C., Rogers, Goussard J. (1998) Canal control algorithms currently in use. Journal of Irrigation and Drainage Engineering, 124: 11-15.
[3] Clemmens A.J., Bautista E., Wahlin B.T. (2005) Simulation of automatic canal control systems. Journal of Irrigation and Drainage Engineering, 131: 324-325.
[4] Shahrokhnia M.A., Javan M. (2006) Dimensionless stage-discharge relationship in radial gates. Journal of Irrigation and Drainage Engineering, ASCE, 4: 180-184.
[5] Cui, W., Chen W.X., Yao X., et al. (2009) Research on canal of large scale water transfer project. South-to-North Water Transfers and Water Science & Technology, 7: 6-10.
[6] Guan, G.H., Wang C.D. (2008) Difference study of steady flow and unsteady flow simulation for long canal system. South-to-North Water Transfers and Water Science & Technology. 6: 13-17.
[7] Yao, X., Wang C.D., Ding Z.L., et al. (2008) Study on canal system operation based on discharge active compensation. Journal of Sichuan University (Engineering Science Edition). 40: 38-44.
[8] Zhang, C., Li Q.G., Qian J. (2014) Study on partition schedule for large water transfer channel. Journal of Hydroelectric Engineering. 33: 116-121.
[9] Fang Y.M., Zhang D.W., Lei X.H., et al. (2014) Emergency control strategy for sudden water pollution accident in the main channel of Middle Route of South-to-North Water Diversion Project. South-to-North Water Transfers and Water Science & Technology. 12: 133-136.
[10] Cao Y.S., Chang J.X., Chen X.N., et al. (2016) Improved water dispatch model for Middle Route Project of South-to-North Water Transfer. Journal of Hydroelectric Engineering. 35: 95-101.

[11] Liu M.K., Wang C.D., Yan Y.B., et al. (2009) Analysis and comparison of radial gate flows formulas. South-to-North Water Transfers and Water Science & Technology. 7: 18-19.

[12] Jiang W.L., Lei Y.F., Zhang Q.S., et al. (2008) RBFNN modelling based dynamical flow soft sensing method. Chinese journal of Scientific Instrument. 29: 1888-1893.

[13] Zhang J.M., Wang Y.R., Xu W.L., et al. (2005) New iteration method for calculation water level of gradually varied steady flow. SHUILI XUEBAO. 36: 501-504.

[14] Wen H., Li Feng L. (2014) Numerical integration method for calculating water surface profile of gradually varied steady flow in parabola shape channel. Transactions of the Chinese Society of Agriculture Engineering. 30: 82-86.

[15] Li K.S., Li Y.X., Tang M.D., et al. (2005) Application of genetic programming on statistical modelling. Journal of System Simulation. 17: 1597-1600.

[16] Chen X.N., Huang Q., Qiu L., et al. (2006) Crop water production function based on genetic programming. Transactions of the Chinese Society of Agriculture Engineering. 22: 6-9.