The Water Quality Prediction Based on the Gray Model and Curve Fitting

Kefei Cheng\textsuperscript{a,*}, Lei Cheng\textsuperscript{a}, Yongdong Huang\textsuperscript{b}

\textsuperscript{a}College of Computer, Chongqing University of Posts and Telecommunications
Chongqing 400045, China

\textsuperscript{b}Agricultural Quality Standards and Testing Technology Research Institute in Agricultura Academy of Chongqing, Chongqing 400045, China

Abstract

Water quality prediction can be applied to guard against all kinds of emergency events and provide decision support to the relevant departments. Many water quality prediction methods have supplied, such as time series analysis method, fuzzy algorithm, artificial neural network, wavelet analysis. Time series analysis method is suitable for the changing apparent data sequence, the predicted data are very random; fuzzy algorithm are used in forecast of the water quality, establishing the correspondences between predicting factors and predicting objects is more difficult; artificial neural network method is not a good method for the network topology; study and application of wavelet analysis is not very perfect in water quality prediction. Combining the water quality characteristics of the region, this paper uses the combination forecast method of gray model and curve fitting. Before prediction, it pretreats and classifies the corresponding data, using gray model predicts the periodic parameters, using the curve fitting method predicts the cyclical parameters. Then combining forecast result gets the final trend. Experimental data analysis shows that the fusion method has the higher forecast accuracy than the single method, if data are enough, this method may also be made better effect.

Keywords: Water Quality Forecast; Gray Model; Weakening Operator; Residual Sequence; Curve Fitting

1 Water Quality Prediction and the Present Study Situation

Water quality prediction is for a certain extent, to analysis the water quality conditions for the period, to forecast the future development situation, and to determine the condition of the water quality, its changing trend, speed and time of achieve a changing limit, etc. To predict water

*Project supported by the Southwest towns library the poise of surface drinking water security the applicable technology research and demonstration (No. 2009ZX07425-003).

*Corresponding author.

Email address: chengkf@cqupt.edu.cn (Kefei Cheng).
quality can be applied to prevent all kinds of emergencies, provide decision support for relevant departments, maximize reduce the influence of the incident. Water quality prediction is also achieve the water environment planning and management, and water pollution comprehensive prevention and controlling such indispensable foundation work [1, 2]. Water quality prediction model has becoming an important topic that is very popular in the environment staff. Now the water quality prediction establish model through mathematical statistics or other mainly mathematical methods, the common method is time sequence method [3], fuzzy algorithm [4], the artificial neural network [5], wavelet analysis [6], etc. But these methods have their limitations. So these methods are not fit in the area of water quality prediction. Combining the water quality characteristics of the region, this paper uses the combination forecast method of gray model and curve fitting. The specific method summarized as follow:

2 Gray Model in the Application of Water Quality Prediction

The gray system analysis has good applicability for incomplete information (or not completely), the GM (1, 1) model and improved model have already applied in a wide range of water quality prediction application [7]. Grey dynamic model can effectively overcome the bad influence of the information by the fluctuations, and allow to recruit new information to the model, make more accurate and reliable prediction [8]. But the study also found that grey prediction error correction method of dynamic model of the traditional methods to resolve the prediction of the statistical model is taken as the final average prediction to increase the prediction accuracy, and this method in many cases can’t solve ideal error of problems. This paper based on the water quality of the status quo with average choice weakening the buffer operator to generate ash sequence after another, with the residual GM (1, 1) model prediction.

2.1 Operator of Average Weakening the Generation of Buffer

The original data set sequence $x = (x_1, x_2, ..., x_n)$. making $XD = (x_1d, x_2d, ..., x_nd)$, among them $x_{kd} = \frac{1}{n-k+1}[x_k + x_{k+1} + ... + x_n]$, $k = 1, 2, ..., n$. $XD$ is a one order average weakening the buffer operator [9]. making $XD^2 = (x_1d^2, x_2d^2, ..., x_nd^2)$, among them $x_{kd}^2 = \frac{1}{n-k+1}[x_k^2 + x_{k+1}^2 + ... + x_n^2]$, $k = 1, 2, ..., n$. $XD^2$ is a one order average weakening the buffer operator [10]. making $XD^3 = (x_1d^3, x_2d^3, ..., x_nd^3)$, among them $x_{kd}^3 = \frac{1}{n-k+1}[x_k^3 + x_{k+1}^3 + ... + x_n^3]$, $k = 1, 2, ..., n$. $XD^3$ is a one order average weakening the buffer operator [10]. According to the water quality data, using direct raw data sequence, accuracy model [11] is not high, but with an average buffer operator generated weakening the data to predict, accuracy greatly improved, and the higher the order number, the higher the accuracy. But the order number increase, the more greatly increase the amount of calculation, and more
than three order, to improve the accuracy of very few, comprehensive consideration, choose the first three average weakening the buffer operator to forecast is more appropriate.

2.2 With the Grey Forecasting Model of Residual Model $GM(1, 1)$

The traditional model of the application is very limited, when the precision of the model is not accord with, usable residual sequence [11] model building, the original model is revised, in order to improve the precision. Below is a residual model $GM(1, 1)$ establishment and prediction of steps:

Step 1: the original data initialization The original data series $x = (x(1), x(2), ..., x(n))$ three order average weakening the buffer operator after processing data sequence $XD^3$, $XD^3$ the notes for $x(i)$.

Step 2: will a accumulation data sequence to 1-AGO Will a accumulation data sequence $x^{(0)}(i)$, get 1-AGO generate data sequence $x^{(1)}(k), x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$.

Step 3: 1-AGO next to the mean [12] is generated will be close to generate data sequence $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$ mean, get 1-AGO close to mean sequence $Z^{(1)}(k) = \frac{1}{2} \left( x^{(1)}(k) + x^{(1)}(k-1) \right)$.

Step 4: GM (1, 1) the basic form and development coefficient-a gray and action is the calculation

$$
a_{k} = \frac{1}{Z^{(1)}(k)} \sum_{i=1}^{k} x^{(0)}(i) - b_{k} \left( x^{(1)}(k) - x^{(0)}(k) \right) e^{-ak}.
$$

Step 5: with data processing model $x^{(0)}(k) + aZ^{(1)}(k) = b$ as the primary data sequence, $x^{(1)}(k)$ as 1-AGO generation sequence, GM (1, 1) model of response time to type $x^{(1)}(k)$, $d\hat{x}^{(1)}(k+1) = (-a) \left( x^{(0)}(k) - \frac{b}{a} \right) e^{-ak}$ as derivatives also said the original value. $d\hat{x}^{(1)}(k+1) = (-a) \left( x^{(0)}(k) - \frac{b}{a} \right) e^{-ak}$ as predictive value. Set $\epsilon^{(0)} = \left( \epsilon^{(0)}(1), \epsilon^{(0)}(2), \cdots, \epsilon^{(0)}(n) \right)$, which $\epsilon^{(0)}(k) = x^{(1)}(k) - \hat{x}^{(1)}(k)$ is residual sequence of $x^{(1)}(k)$. The corresponding residual fixed time response type

$$
\hat{x}^{(0)}(k+1) = \begin{cases} 
(-a) \left( x^{(0)}(k) - \frac{b}{a} \right) e^{-ak}, & k < k_0 \\
(-a) \left( x^{(0)}(k) - \frac{b}{a} \right) e^{-ak} \pm a\epsilon \left( \epsilon^{(0)}(k_0) - \frac{b}{a\epsilon} \right) e^{-a\epsilon(k-k_0)}, & k \geq k_0 
\end{cases}
$$

is called the derivative of residual reduction type correction model.

2.3 Common Forecast Method

Using a web better standard of five data to a test, each data according to the above step calculation, the results as shown in Table 1.
<table>
<thead>
<tr>
<th>Serial number</th>
<th>Development coefficient a</th>
<th>Gray action b</th>
<th>residual</th>
<th>Actual data</th>
<th>Forecast data</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.009</td>
<td>1.778</td>
<td>0.006</td>
<td>1.782</td>
<td>1.463</td>
<td>0.969</td>
</tr>
<tr>
<td>2</td>
<td>0.004</td>
<td>1.997</td>
<td>0.005</td>
<td>1.960</td>
<td>1.818</td>
<td>0.665</td>
</tr>
<tr>
<td>3</td>
<td>0.013</td>
<td>1.855</td>
<td>0.053</td>
<td>1.426</td>
<td>1.375</td>
<td>2.711</td>
</tr>
<tr>
<td>4</td>
<td>0.024</td>
<td>1.783</td>
<td>0.142</td>
<td>1.544</td>
<td>1.033</td>
<td>5.289</td>
</tr>
<tr>
<td>5</td>
<td>0.015</td>
<td>1.715</td>
<td>0.048</td>
<td>1.396</td>
<td>1.222</td>
<td>2.877</td>
</tr>
</tbody>
</table>

3 The Curve Fitting in a Certain Area of the Application of the Water Quality Prediction

For there is little data in a lot of area at present, only one year of data is full, and a periodic strong (one year in a cycle). In view of this, with this year’s data curve fitting, to determine what kind of type of curve, other every year in this curve to forecast, but only in a few months before the data to determine the parameters in the curve, you can take this year of curve painted, thus realized prediction. The specific procedure is as follows:

The first step, looking for fitting curve type.

Matlab toolbox of curve fitting water quality data test, import a year of data, and in all the curve of the experiment, found that the gaussian function $y = a_1 \exp\left(\frac{(x - b_1)/c_1)^2}{2}\right) + a_2 \exp\left(\frac{(x - b_2)/c_2)^2}{2}\right)$ high accuracy. This is determined by this function fitting.

The second step, about next year to carry on the forecast, determine the function parameters.

For this function has six parameters, at least six data to determine the out all the parameters, so as to determine the curve. By the forecast of the year in the first six months of the above data into function expression, confirmed the parameters $a_1 = 0.1534, b_1 = 30.29, c_1 = 38.75, a_2 = 0.1521, b_2 = 5.431, c_2 = 0.8832$.

The third step, with certain parameters on the back a few months forecast. Painting 12 months of graphic, the graphics can behind to six months of prediction.

According to the above step, it forecast other data.

4 The Reasonable Selection of Water Quality Prediction Method

After the experiment, a new method that is the combination of the above two methods to forecast is more appropriate, and at the same time water quality of a certain area is mainly effect by the agricultural non-point and upstream of the water quality of water, so choose 6 important properties to forecast, respectively is ammonia nitrogen and phosphorus, total nitrogen, total escherichia coli, PH and CODCR. Six attributes have three attributes ammonia nitrogen and phosphorus and e. coli periodic strong, with gray model accuracy of prediction is not enough, so the three attributes choose to use the curve fitting to carry on the forecast, the rest of the three
attributes with gray model forecast.

In 5 data of total nitrogen, PH, CODOR of a certain area to a test, each data according to the second part of the calculation procedure is introduced, results as shown in Table 2.

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Development coefficient a</th>
<th>Gray action b</th>
<th>residual</th>
<th>Actual data</th>
<th>Forecast data</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>7.140</td>
<td>0.000</td>
<td>7.09</td>
<td>7.063</td>
<td>0.040</td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>7.128</td>
<td>0.000</td>
<td>7.07</td>
<td>7.080</td>
<td>0.021</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>7.125</td>
<td>0.001</td>
<td>7.13</td>
<td>7.119</td>
<td>0.070</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>7.135</td>
<td>0.001</td>
<td>7.15</td>
<td>7.167</td>
<td>0.090</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>7.082</td>
<td>0.000</td>
<td>7.07</td>
<td>7.052</td>
<td>0.028</td>
</tr>
</tbody>
</table>

In 2010 waters of a total phosphorus test data, the curve fitting is as shown in Fig. 1.

With 2011 years ago six months of data into the above function expressions, confirmed the parameters $a_1 = 0.1534$, $b_1 = 30.29$, $c_1 = 38.75$, $a_2 = 0.1521$, $b_2 = 5.431$, $c_2 = 0.8832$. According to the data, gives annual 12 months curve, the curve is also describes the subsequent six months of forecast data, as shown in Fig. 2 shows.

The curve of the first six months of data is real data, back six months of data is to use the curve fitting out, also can find out 7, 8, 9 on the three months of real data, can use this three months of real data and figure from the three months of data comparison (see Table 3).

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Forecast data</th>
<th>Real data</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0.140</td>
<td>0.150</td>
<td>6.7</td>
</tr>
<tr>
<td>8</td>
<td>0.100</td>
<td>0.094</td>
<td>6.4</td>
</tr>
<tr>
<td>9</td>
<td>0.112</td>
<td>0.107</td>
<td>4.7</td>
</tr>
</tbody>
</table>
In the area of the experimental data, with improved gray model is found and curve fitting, the method of combining a higher accuracy.

5 Last Word

This paper focused on gray model and the curve fitting joint prediction method on research. According to some basic situation of the region, and delete some obvious not reasonable data and attribute, finally choosing with residual with shades of gray model of a certain area of three water quality attributes to carry on the forecast, garnish with curve fitting method for the rest of the three has obvious periodic data to carry on the forecast, than the traditional forecasting method has higher accuracy, now a lot of literature have both a single method the detailed introduction, but the two methods of this new method was combined this paper pointed out for the first time. This paper mainly studies the following aspects: the gray model with the data to the first three average weakening the buffer operator handling, and then residual correction; With curve fitting for a year before micro roughness complete data to find the type of curve, again with forecast that year a few months of the data to find out of the parameters of the curve, and to predict remaining months of data; According to the forecast results found in using gray model and the curve fitting joint prediction method is practicable water quality prediction method.

This article uses the method although already reached very high accuracy, but also can continue to increase the data through modeling, adjust the gray model and of the parameters of the curve fitting, want to get better effect.

References

[1] Jun Wu, Often Used Study and New Method of Researching of the Water Quality Simulation and Prediction, Hefei University of Technology, 2005


[8] Lizhi Cui, Sifeng Liu, Zhengpeng Wu, Study on new weakening buffer operators and their applications, Control and Decision, 24 (2009), 1252-1256


