

Using a grey multivariate model to predict impacts on the water quality of the Zhanghe River in China

Feifei Fan^a, Zhengran Qiao^a and Lifeng Wu^{a,b,*}

^a School of Management Engineering and Business, Hebei University of Engineering, Handan, 056038, China

^b Hebei Key Laboratory of Intelligent Water Conservancy, Hebei University of Engineering, Handan 056038, China

*Corresponding author. E-mail: wlf6666@126.com

ABSTRACT

In order to assess the social factors affecting the water quality of the Zhanghe River and predict the potential impact of growth in primary, secondary, tertiary industries and population on water quality of the Zhanghe River in the next few years, a deformation derivative cumulative grey multiple convolution model (DGMC(1,N)) was applied. In order to improve the accuracy of the model, the accumulation of deformation derivatives is introduced, and the particle swarm optimization algorithm is used to solve the optimal order. The DGMC(1,N) model was compared with GM(1,2) and GM(1,1) models. The results show that the DGMC(1,N) model has the highest prediction accuracy. Finally, DGMC(1,N) model is used to predict the potential impact of growth in primary, secondary, tertiary industries and population on water quality in the Zhanghe River (using chemical oxygen demand (COD) as the water quality indicator).

Key words: chemical oxygen demand, deformable grey multivariable convolution, primary industry, secondary industry, tertiary industry, Zhanghe River

HIGHLIGHTS

- The proposed model has the highest prediction accuracy.
- The water quality in Zhanghe is predicted considering the socio-economic development.
- The proposed model can be used in other rivers.

1. INTRODUCTION

Due to the continuous development of the economy, growing human activities increase the potential to pollute waterways and degrade the environment. Water pollution not only affects human health and the health of ecosystems, it also restricts social and economic development. Therefore, it is important to understand the relationship between economic development and the pollution of waterways in order to identify water pollution mitigation strategies. In order to improve water quality, extensive research has been conducted on the relationship between water quality and its influencing factors.

Kyei & Hassan (2019) analyzed the economic and environmental impact of water pollution taxes in the Olifants Basin, South Africa, using an environmentally scalable general equilibrium model. Nguyen *et al.* (2018) developed a model to evaluate the relationship between economic activity and water pollution in Viet Nam in order to identify water pollution mitigation strategies within the context of economic development. In the case of the Samarinda River in East Kalimantan in Indonesia, Vita *et al.* (2018) adopted a field observation method based on interviews of the government of society, industry, public welfare activities along the river and environmental departments, and further used an analytic hierarchy process to establish data, and identify countermeasures for controlling the water pollution. Choi *et al.* (2015) assessed the relationship between the economic growth in four major river basins in Korea and two key water quality indicators. Li & Lu (2020) tested the impact of regional integration on cross-border pollution under the auspices of the Yangtze River Economic Belt by using the difference-in-difference model. The results showed that regional integration could significantly reduce cross-border water pollution. Liu *et al.* (2020) used qualitative and quantitative analyses to study the relationship between water pollution and economic growth in the Nansihu River basin in China.

Cullis *et al.* (2019) discuss the increasing risks to water quality in the Begg River basin in South Africa as a result of climate change and rapid urban development, as well as the direct and indirect economic impacts this may have on the agricultural

sector. Based on the threat to water quality in the American Midwest posed by agricultural runoff, Floress *et al.* (2017) proposed and tested a structural equation model based on the dual interest theory to test whether, and to what extent, the relationship between awareness and agribusiness attitudes is regulated by management attitudes. Qualitative assessments of the Lake Merrill basin by Pires *et al.* (2020), which were performed using discriminant analysis methods, concluded that seasonality mainly affected anthropogenic sources such as agricultural activities and household emissions. An assessment of the current state of water quality in Lake Wadi El-Rajan, particularly following the increase in uncontrolled economic activity within its borders, was presented by Goher *et al.* (2019). Similarly, De Mello *et al.* (2020) outline the relationship between land use/land cover and water quality in Brazil and its impact on water quality. Kuwayama *et al.* (2020) examined long-term trends in surface water quality, nutrient pollution and its potential economic impact in Texas, USA while Du Plessis *et al.* (2015) quantified the complex relationship between land cover and specific water quality parameters and developed a unique model equation to predict water quality in the Grootdraai Dam catchment due to the importance of water quality within the basin to the country's future economic growth.

While researchers have analyzed the impact of the economy on water pollution from different perspectives, few predict the impact of economic development on water resources in the future.

The accurate prediction of levels of water pollution can inform the identification of countermeasures that are needed in response to the direction of future economic development. Extreme learning machine was used by Saberi-Movahed & Mehrpooya (2020) to predict longitudinal dispersion coefficients and evaluate the pollution status of water pipelines. Najafzadeh & Emamgholizadeh (2019) estimated the biochemical oxygen demand, dissolved oxygen and chemical oxygen demand (COD) using gene expression programming, evolutionary polynomial regression and a model tree while Najafzadeh *et al.* (2019) estimated biochemical oxygen demand and chemical oxygen demand using multivariate adaptive regression splines and least squares support vector machines. Mustafa *et al.* (2021) used support vector machines to build prediction models of water quality in the Kelantan River based on historical data collected from different sites. Xinzi *et al.* (2020) applied correlation analysis and path analysis to identify the causal relationship between urbanization and water quality indicators, and then comprehensive water quality indicators and related urbanization parameters were input into a back-propagation neural network for water quality prediction. Bao *et al.* (2020) predicted the water quality index for free surface wetlands by using three soft computing techniques, namely adaptive neurofuzzy systems, artificial neural networks and group data processing.

Liu & Wu (2021) used a new adjacent non-homogeneous grey model to predict renewable energy consumption in Europe Bilgaev *et al.* (2020) analyzed the environmental and socio-economic development indicators of Baikal Island region with the method of constructing time series and structural transfer.

Zhang *et al.* (2020) used the grey water footprint to estimate the different water bodies in 31 provinces (autonomous regions) in China. The poor information principle of the grey system theory was used to predict the rural water environment with a network search method to provide support for rural water environmental governance. Shen *et al.* (2020) compared a residual correction grey model with a grey topology prediction method in order to predict the water quality of the artificial reef area in Haizhou Bay. Yuan *et al.* (2019) used a fractional grey scale power model to predict water consumption while Jiang *et al.* (2019) used grey multivariate forecasting models to predict the long-term electricity consumption of power companies.

Şahin (2019) combines linear and non-linear metabolic models with optimization techniques to accurately predict Turkey's greenhouse gas emissions. Zhong *et al.* (2017) use the grey model of particle swarm optimization algorithm to predict short-term photovoltaic power generation, which improves the prediction accuracy compared with the traditional grey model. Utkucan (2021) uses genetic algorithms to optimize parameters, and proposes a new nonlinear grey Bernoulli model to study and analyze energy trumpets. Hu (2020) uses the grey multivariate forecasting model to predict bankruptcy, and uses genetic algorithm to avoid the influence of time on the result. Wang & Hao (2016) nonlinearly optimized the background value of the grey convolution model and compared the prediction of industrial energy consumption with the traditional model. Although the research is extensive, few researchers have used a grey multivariable model to analyze the impact of future economic development on water quality. A key feature of a grey model is that it can be used when there is little information.

In this study, a grey multivariable model was used to analyze and forecast the water pollution in the upper reaches of the Zhanghe River by the added value of the first, second and third industries and the population added value. A new accumulation method was adopted to improve the prediction accuracy of the original grey multivariable model.

In this paper, Section 2 describes the study area and the indicators. Section 3 outlines the forecasting method while Section 4 analyzes the impact of local socioeconomic factors on the water quality of an upstream reach of the Zhanghe River in China.

2. STUDY AREA AND DATA

2.1. The study area

The Zhanghe River is a tributary of the Hai River. Its source is located in the Shanxi Province and flows through Shanxi, Hebei and Henan Provinces. The upper reaches of the river mainly comprise two tributaries, namely the Qingzhanghe River and the Zhuozhanghe River. The study area is shown in Figure 1.



Figure 1 | Study area.

2.2. Water quality indicator and data

When assessing the level of pollution of a river, chemical oxygen demand is an important and fast-determinable indicator of organic pollution. The COD is a measure of the water quality of a river. The higher the COD level, the greater the degradation of water quality. The COD value is reported once a month. This article takes the average of the 12-month values as the research object. The COD levels in the upper reaches of the Zhanghe River from 2013 to 2018 have been reported by Handan Ecological Environment Bureau.

Primary industry is the foundation of the national economy, while secondary industry is a leading industry of the national economy and tertiary industry is the key to providing employment in China. Population is the main indicator of social factors. The data on socio-economic indicators was obtained from the 'China County Statistical Yearbook' and 'Hebei Economic Yearbook' from 2013 to 2018.

Consequently, the relationship between the socio-economic indicators and COD levels was analyzed.

2.3. Deformable grey multivariable convolution (DGMC) model

A non-negative original sequence is $X_1^{(0)} = \{x_1^{(0)}(1), x_1^{(0)}(2) \dots x_1^{(0)}(n)\}$. $x_1^{(1)}(k) = \sum_{i=1}^k x_1^{(0)}(i)$ is a first order accumulation sequence. Considering the definition of the deformable derivative (Wu & Zhao 2019), the $\alpha(0 \leq \alpha \leq 1)$ -order accumulation is

$$x_1^{(\alpha)} = x_1^{(0)}(k) + \alpha x_1^{(1)}(k - 1) \tag{1}$$

The DGMC(1,N) modelling process is described below.

(1) A non-negative sequence is

$$X_1^{(0)} = \{x_1^{(0)}(1), x_1^{(0)}(2) \dots x_1^{(0)}(m)\}.$$

The sequences of the related factors are:

$$X_2^{(0)} = \{x_2^{(0)}(1), x_2^{(0)}(2) \dots x_2^{(0)}(m)\}$$

.....

$$X_N^{(0)} = \{x_N^{(0)}(1), x_N^{(0)}(2) \dots x_N^{(0)}(m)\}.$$

(2) The α -order of DGMC(1,N) is

$$x_1^{(0)}(k) + b_1 z_1^{(\alpha)}(k) = \sum_{i=2}^N b_i z_i^{(\alpha)}(k) + u \tag{2}$$

$b_1, b_2, \dots b_N$ and u is the parameter to be estimated.

The parameters can be obtained by using the least squares method which minimizes the sum of the squared residuals. The unknown parameters can be solved by the following formulas:

$$[\hat{b}_1, \hat{b}_2, \dots \hat{b}_N, \hat{u}]^T = (B^T B)^{-1} B^T Y \tag{3}$$

where

$$Y = \begin{bmatrix} x_1^{(\alpha)}(2) - x_1^{(\alpha)}(1) \\ x_1^{(\alpha)}(3) - x_1^{(\alpha)}(2) \\ x_1^{(\alpha)}(4) - x_1^{(\alpha)}(3) \\ \vdots \\ x_1^{(\alpha)}(m) - x_1^{(\alpha)}(m-1) \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{x_1^{(\alpha)}(1) + x_1^{(\alpha)}(2)}{2} & \frac{x_2^{(\alpha)}(1) + x_2^{(\alpha)}(2)}{2} & \dots & \frac{x_N^{(\alpha)}(1) + x_N^{(\alpha)}(2)}{2} & 1 \\ \frac{x_1^{(\alpha)}(2) + x_1^{(\alpha)}(3)}{2} & \frac{x_2^{(\alpha)}(2) + x_2^{(\alpha)}(3)}{2} & \dots & \frac{x_N^{(\alpha)}(2) + x_N^{(\alpha)}(3)}{2} & 1 \\ \frac{x_1^{(\alpha)}(3) + x_1^{(\alpha)}(4)}{2} & \frac{x_2^{(\alpha)}(3) + x_2^{(\alpha)}(4)}{2} & \dots & \frac{x_N^{(\alpha)}(3) + x_N^{(\alpha)}(4)}{2} & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \frac{x_1^{(\alpha)}(m-1) + x_1^{(\alpha)}(m)}{2} & \frac{x_2^{(\alpha)}(m-1) + x_2^{(\alpha)}(m)}{2} & \dots & \frac{x_N^{(\alpha)}(m-1) + x_N^{(\alpha)}(m)}{2} & 1 \end{bmatrix}$$

(3) Assume $\hat{x}_1^{(1)}(1) = x_1^{(0)}(1)$, then the solution is

$$\hat{x}_1^{(\alpha)}(k) = x_1^{(0)}(1)e^{-\hat{b}_1(k-1)} + \int_1^k e^{-\hat{b}_1(k-t)} f(t) dt. \tag{4}$$

where

$$f(t) = b_2 x_2^{(\alpha)}(t) + b_3 x_3^{(\alpha)}(t) + \dots + b_N x_N^{(\alpha)}(t) + u. \tag{5}$$

The time response formula obtained from the Gaussian formula is:

$$\hat{x}_1^{(\alpha)}(k) = x_1^{(0)}(k)e^{-\hat{b}_1(k-1)} + \sum_{t=2}^k \left\{ e^{-\hat{b}_1 \left(k-t + \frac{1}{2} \right)} * \frac{1}{2} [f(t) + f(t-1)] \right\}. \tag{6}$$

(4) Therefore, the sequence

$$\hat{x}_1^{(\alpha)} = \{\hat{x}_1^{(\alpha)}(1), \hat{x}_1^{(\alpha)}(2) \dots \hat{x}_1^{(\alpha)}(m)\}.$$

The α -order accumulative reduction

$$\hat{x}^{(0)}(k) = \hat{x}^{(\alpha)}(k) - \alpha \hat{x}^{(1)}(k-1), \tag{7}$$

thus

$$\hat{x}_1^{(0)} = \{\hat{x}_1^{(0)}(1), \hat{x}_1^{(0)}(2) \dots \hat{x}_1^{(0)}(m)\}.$$

(5) Evaluate the model using the mean absolute percentage error (MAPE), as follows:

$$MAPE = 100\% \frac{1}{n} \sum_{k=1}^n \left| \frac{x_1^{(0)}(k) - \hat{x}_1^{(0)}(k)}{x_1^{(0)}(k)} \right|. \tag{8}$$

2.4. GM(1,1)

(1) A non-negative original sequence is (Yin *et al.* 2017)

$$X_1^{(0)} = \{x_1^{(0)}(1), x_1^{(0)}(2) \dots x_1^{(0)}(n)\}, \text{ Accumulate and generate a new sequence } X_1^{(1)} = \{x_1^{(1)}(1), x_1^{(1)}(2) \dots x_1^{(1)}(n)\}.$$

Then the differential equation of GM(1,1) is

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu \tag{9}$$

(2) Assume $\hat{\alpha}$ parameter to be estimated, $\hat{\alpha} = \left(\frac{a}{\mu} \right)$, the least squares estimation minimizes the sum of the squared residuals, we can obtain the parameters by using the least squares method. The unknown parameters can be solved by the following formulas:

$$\hat{\alpha} = (B^T B)^{-1} B^T Y_n \tag{10}$$

(3) The time response formula obtained from the Gaussian formula is

$$\hat{X}^{(1)}(k+1) = \left[X^{(0)}(1) - \frac{\mu}{a} \right] e^{-ak} + \frac{\mu}{a} \quad k = 0, 1, 2, \dots, n \quad (11)$$

2.5. GM(1,2)

(1) GM(1,2) represents a first-order differential equation with two variables, the differential equation is (Li *et al.* 2016)

$$\frac{dx_1^{(1)}}{dt} + ax_1^{(1)} = bx_2^{(1)} \quad (12)$$

(2) Assume \hat{a} parameter to be estimated, $\hat{a} = [a, \quad b]^T$, we can obtain the parameters by using the least squares method. The unknown parameters can be solved by the following formulas:

$$\hat{a} = (B^T B)^{-1} Y_n \quad (13)$$

where

$$B = \begin{bmatrix} -0.5[x_1^{(1)}(1) + x_1^{(1)}(2)] & x_2^{(1)}(2) \\ -0.5[x_1^{(1)}(2) + x_1^{(1)}(3)] & x_2^{(1)}(3) \\ \vdots & \vdots \\ -0.5[x_1^{(1)}(n-1) + x_1^{(1)}(n)] & x_2^{(1)}(n) \end{bmatrix}$$

$$Y_n = [x_1^{(0)}(2), x_1^{(0)}(3), \dots, x_1^{(0)}(n)]^T$$

(3) The predicted value is

$$\hat{x}_1(k+1) = \left[x_1^{(1)}(0) - \frac{b}{a} x_2^{(1)}(k+1) \right] e^{-ak} + \frac{b}{a} x_2^{(1)}(k+1) \quad (14)$$

The original predicted value is

$$\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k), \quad k = 1, 2, \dots, n-1$$

3. COMPARATIVE PREDICTION ACCURACY

DGMC(1,2), GM(1,2) and GM(1,1) models were fitted to the annual COD concentrations for the period 2013 to 2018 reported by the Handan Ecological Environment Bureau. The COD concentrations and the model fitting results are shown in Table 1 and Figure 2. The MAPE value for the DGMC(1,2) model is 4.9% in comparison to 31.8% for the GM(1,2) model and 5.7% for the GM(1,1) model.

Compared with traditional grey models, the DGMC(1,2) model improves the prediction accuracy. Consequently, the DGMC(1,2) was used to predict the annual average COD values.

4. THE INFLUENCE OF SOCIAL DEVELOPMENT ON WATER QUALITY

The next step was to assess the relationship between COD levels and primary, secondary, and tertiary industries and population, respectively.

4.1. Predicting COD under primary industry

The added value of the primary industry from 2013 to 2018 and the annual average COD levels are given in Table 2.

Table 1 | Annual average COD concentrations predicted by GM(1,1), GM(1,2) and DGMC(1,2) models

Year	Annual average COD (mg/L)	GM(1,1) predicted COD	GM(1,2) predicted COD	DGMC(1,2) predicted COD
2013	9.09	9.09	9.09	9.09
2014	6.00	5.79	4.20	6.66
2015	6.67	6.35	9.11	5.95
2016	6.54	6.97	10.31	6.54
2107	7.08	7.65	9.80	7.57
2018	8.92	8.39	8.88	8.85
MAPE		5.7%	31.8%	4.9%

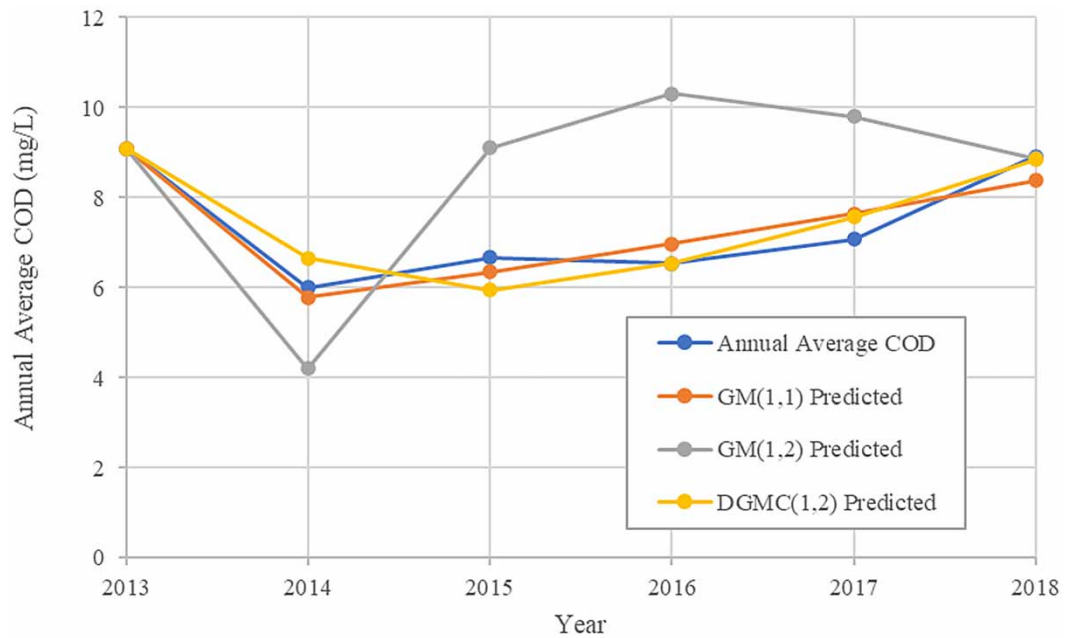


Figure 2 | Comparison of predicted annual average COD.

Table 2 | The added value of primary industry and COD

Year	2013	2014	2015	2016	2017	2018
Primary industry added value (10,000 Yuan)	1,424,222	1,492,075	1,543,585	1,471,621	1,346,975	1,238,119
COD (mg/L)	9.09	6.00	6.67	6.54	7.08	8.92

Step 1. The original data sequence is $x_1^{(0)} = (9.09, 6, 6.67, 6.54, 7.08, 8.92)$,

$$x_2^{(0)} = (1424222, 1492075, 1543585, 1471621, 1346975, 1238119)$$

The optimization result $\alpha = 0.07$ is obtained by particle swarm optimization.

$$x_1^{(0.07)} = \{9.09, 6.64, 7.73, 8.07, 9.07, 11.41\}$$

$$x_2^{(0.07)} = \{1424222, 1592477, 1749173, 1786026, 1765124, 1751225\}$$

$$\text{Step 2. } Y = \begin{bmatrix} -2.44919 \\ 1.092977 \\ 0.34021 \\ 1.001045 \\ 2.339113 \end{bmatrix}, B = \begin{bmatrix} -7.86541 & 1508350 & 1 \\ -7.1873 & 1670825 & 1 \\ -7.90389 & 1767599 & 1 \\ -8.57452 & 1775574 & 1 \\ -10.2446 & 1758174 & 1 \end{bmatrix}.$$

By the least square method:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} -0.321 \\ 0.000011 \\ -22.416 \end{bmatrix}$$

Step 3. Then $f(t) = 0.000011x_2^{(0.07)}(t) - 22.416$.
The time response is as follows:

$$\hat{x}_1^{(0.07)}(k) = 6e^{0.321(k-1)} + \sum_{t=2}^k \left\{ e^{0.321 \left(k-t + \frac{1}{2} \right)} * \frac{1}{2} [f(t) + f(t-1)] \right\}$$

Step 4. When

$$\hat{x}^{(0)}(k) = \hat{x}^{(a)}(k) - \alpha \hat{x}^{(1)}(k-1),$$

$$\hat{x}_1^{(0)} = \{9.09, 6.6614, 5.9538, 6.54, 7.5728, 8.8585\}$$

Step 5. Therefore, MAPE = 4.9%.

In order to forecast results, a value for the growth rate of the added value of primary industry was estimated. In the past seven years, the contribution of China's primary industry to GDP has been 0.3%. Noting that Handan City is a fourth-tier city with a large population and mainly relies on agriculture, forestry, animal husbandry and fishery. For the period 2013–2018, the calculated growth rates of the value added to the primary industry were 4.76%, 3.45%, –4.66%, –8.46% and –8.08%, respectively. Consequently, the assumed value-added rate of the primary industry was between 5% and –20%. From the estimated growth rate of the primary industry added value, the primary industry's added value for the period 2019–2022 was estimated, and the DGMC(1,2) model was used to predict the annual average COD value for 2019–2022.

When the growth rate is 5%, $\hat{x}_1^{(0)} = \{11.69, 17.81, 28.38, 45.06\}$.
When the growth rate is –20%, $\hat{x}_1^{(0)} = \{9.53, 8.65, 6.18, 1.95\}$.

As shown in Figure 3, the predicted values of COD rise when the value-added growth rate of the primary industry is 5%. Likewise, the predicted values of COD fall when the rate of the added value of the primary industry falls and the water quality

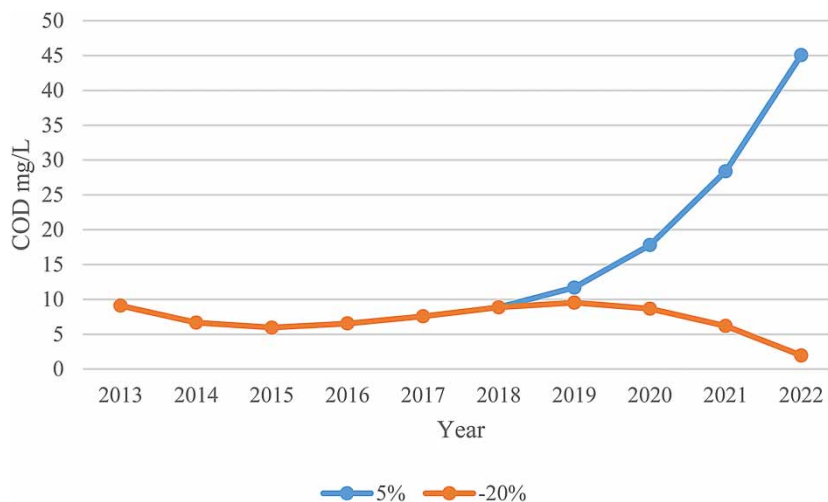


Figure 3 | The predicted COD based on primary industry growth rates.

improves. When the growth level was 1, 3 and 10%, respectively, the COD predicted by the model showed an increasing trend. If the growth rate is 10%, the COD levels would reach 65 mg/L by 2022 (in the absence of any pollution reduction strategies).

In order to further analyze this phenomenon, it is necessary to understand the added value of primary industry in the six counties across the upper reaches of the Zhanghe River, respectively. The added value of the primary industry in the six counties (Shexian, Cixian, Weixian, Daming, Linzhang and Cheng'an) from 2013 to 2018 and the resulting COD levels are shown in Table 3. The COD for the six counties predicted by the new DGMC(1,2) model are given in Table 4. It can be seen from Table 4 that the MAPE values for the six counties are all less than 10%. Assuming that the growth rate of primary industry in the six counties is the same as the overall growth rate, i.e. between 5% and -20%, the predicted impact on COD in the six counties is shown in Figure 4.

Handan is an underdeveloped city, where the proportion of the primary industry is large. Therefore, the development of primary industries will not be reduced in order to reduce the pollution of the rivers. Consequently, there will need to be a focus on optimizing the agricultural industrial structure, adjusting the agricultural structure along the river, strengthening publicity and supervision, and eliminating water pollution at its source, possibly through the reduction of the area of cultivated land along the river and the use of pesticides and fertilizers.

4.2. Predicting COD under secondary industry

The added value of the secondary industry from 2013 to 2018 is shown in Table 5. Following the calculation procedure set out in Section 4.1, the results are shown in Table 6. The MAPE of DGMC(1,2) is 7.35%.

The growth rate of the added value of the secondary industry from 2013 to 2018 were 0.17%, -1.09%, 0.15%, 19.94% and -19.96%. The results predicted for -5.0%, 5.0%, 10.0% and 15.0% growth rates are shown in Figure 5. When the growth rate was 5%, the results showed that with the growth rate increasing, the COD value was 10.5 mg/L in 2022, and the overall trend was increasing. When the growth rate is 10%, the COD value is 13.33 mg/L by 2022, which does not exceed the national standard of 20 mg/L. When the growth rate is 15%, the COD value will be 16.46 mg/L by 2022. However, when the growth rate is reduced by 5%, COD shows a downward trend, and will drop to 5.68 mg/L by 2022.

Table 3 | The value added of the primary industry of the six counties

Year	2013	2014	2015	2016	2017	2018
Shexian	89,625	124,122	133,886	107,365	124,266	108,664
Weixian	274,726	292,932	301,808	286,465	327,825	338,960
Daming	316,197	329,260	327,756	349,959	242,375	246,216
Chengan	228,008	235,865	251,843	263,804	211,916	228,233
Cixian	218,516	214,646	215,000	150,057	94,844	97,657
Linzhang	297,150	295,250	313,292	313,971	345,749	218,389
COD	9.09	6	6.67	6.54	7.08	8.92

Table 4 | The predicted COD for the added value of primary industry in six counties

Year	Annual Average COD	Shexian	Weixian	Daming	Chengan	Cixian	Linzhang
2013	9.09	9.09	9.09	9.09	9.09	9.09	9.09
2014	6.00	6.02	7.34	6.64	6.84	7.14	5.85
2015	6.67	6.40	6.39	6.11	5.90	6.14	5.83
2016	6.54	6.60	6.42	6.69	6.54	6.54	5.87
2017	7.08	7.08	7.83	7.09	7.39	7.42	6.50
2018	8.92	8.82	8.92	8.92	8.51	8.66	7.73
MAPE		1.05%	6.50%	3.61%	5.74%	5.77%	7.76%

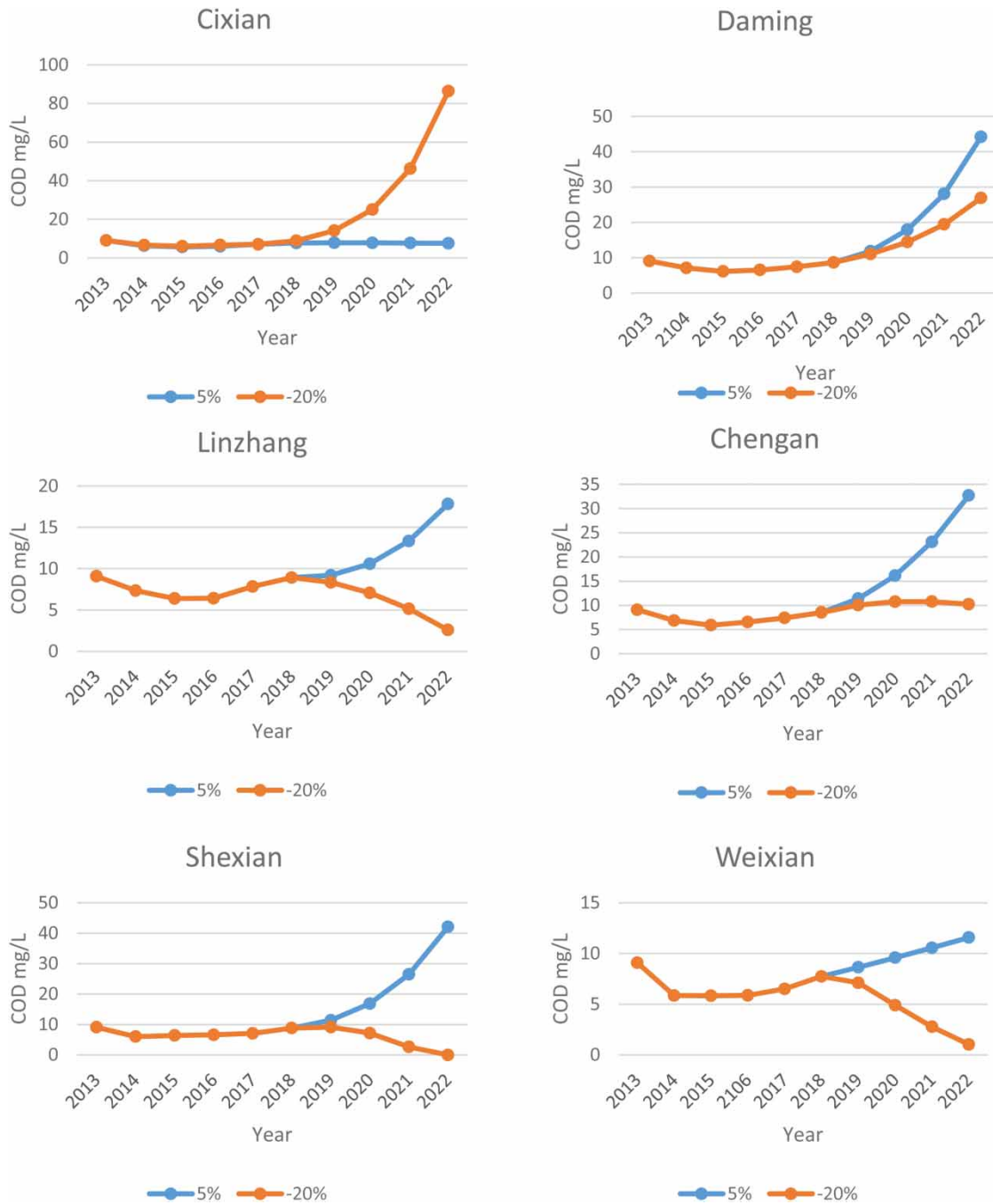


Figure 4 | The predicted COD based on primary industry growth rates in six counties.

Table 5 | The added value of secondary industry and COD

Year	2013	2014	2015	2016	2017	2018
Added value of secondary industry (10,000 Yuan)	4,274,337	4,281,685	4,234,876	4,241,167	5,086,735	4,071,503
COD (mg/L)	9.09	6	6.67	6.54	7.08	8.92

Table 6 | COD under secondary industry predicted by DGMC(1,2)

Year	Annual Average COD (mg/L)	Predicted COD
2013	9.09	9.09
2014	6	6.64
2015	6.67	5.91
2016	6.54	5.93
2107	7.08	7.31
2018	8.92	8.08
MAPE		7.35%

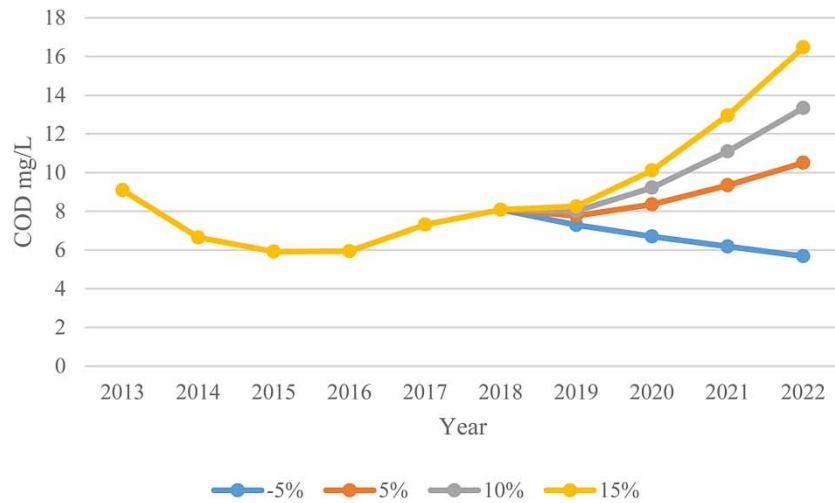


Figure 5 | The predicted COD based on secondary industry growth rates.

Given this potential impact on COD, it is meaningful to study the structure of secondary industries. Secondary industries include industry and construction which are heavy polluting industries. The industrial value added and construction value added from 2013 to 2018 (data source: Hebei Economic Yearbook) are given in Table 7. The COD predicted by DGMC(1,2) models are shown in Table 8. The MAPE of COD is less than 10%, for both industry and the construction.

Applying the adopted secondary industry growth rates to industrial growth, it can be seen from Figure 6, that when the industrial growth rate is 5%, COD would reach 14.30 mg/L in 2022, while if the construction growth rate is 5% then COD would reach 8.87 mg/L in 2022. Assuming growth rates of 10% for industry and construction, the COD would reach 19.51 mg/L and 9.80 mg/L in 2022, respectively. Assuming that the growth rates of the industrial and construction industries are both 15%, the COD would reach 25.28 mg/L and 10.84 mg/L in 2022, respectively.

Handan City attaches great importance to the adjustment of industrial structure, vigorously implements an innovation-driven development strategy, and in-depth advances the supply-side structural reforms, focusing on the

Table 7 | The industrial and construction industry added value

Year	2013	2014	2015	2016	2017	2018
Industrial added value	3,758,584	3,655,505	3,554,121	3,111,938	3,623,444	2,820,676
Added value of construction industry (10,000 Yuan)	515,753	626,180	680,755	1,129,229	1,463,291	1,250,827
COD (mg/L)	9.09	6	6.67	6.54	7.08	8.92

Table 8 | The added values of tertiary industry and COD

Year	2013	2014	2015	2016	2017	2018
Added value of tertiary industry (10,000 Yuan)	3,356,311	3,403,518	3,666,135	3,463,011	4,083,542	4,041,458
COD (mg/L)	9.09	6	6.67	6.54	7.08	8.92

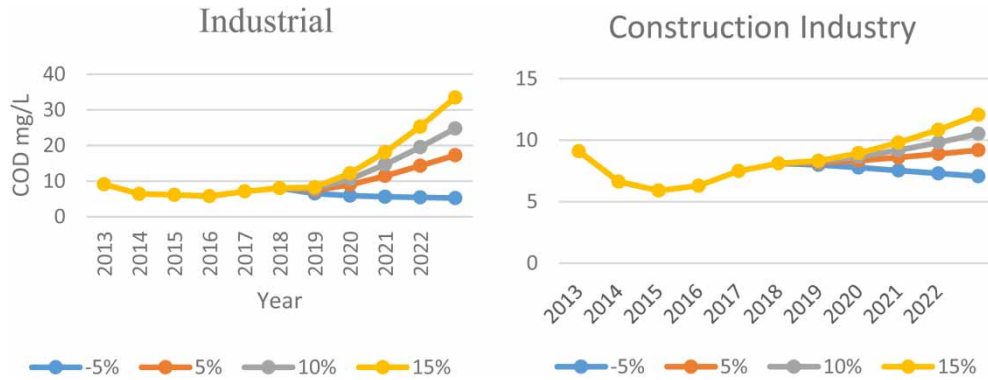


Figure 6 | The predicted COD based on secondary industry growth rates.

transformation of traditional industries, the cultivation of strategic emerging industries, and the development of modern service industries.

These results indicate that the impact construction industry on water quality in the upper reaches of the Zhanghe River is modest and not as great as the impact of industrial growth. From the perspective of water quality, this indicates that it would be beneficial to reduce the proportion of industrial development and that investment in the construction industry could be increased.

4.3. Predicted COD under tertiary industry

The Table 8 shows the added value of tertiary industry from 2013 to 2018. Following the calculation procedure set out in Section 4.1, the results are shown in Table 9. The MAPE of the DGMC(1,2) is 6.66%.

The growth rates of the added value of tertiary industry from 2013 to 2018 were 1.41%, 7.72%, -5.54%, 17.92% and -1.03%. The results predicted for -10%, -5%, 5%, 10% and 15% growth rates are shown in Figure 7. Under a growth rate of -5%, the COD would be 6.24 mg/L by 2022. Under a growth rate of -10%, the COD would be 4.23 mg/L by 2022. The trend in COD is consistent with the growth rate. Under a growth rate of 15%, the predicted COD by 2022 would be 20.86 mg/L, which exceeds national regulations the standard of 20 mg/L. This indicates that from a water quality

Table 9 | COD under tertiary industry predicted by DGMC(1,2)

Year	Annual average COD (mg/L)	Predicted COD
2013	9.09	9.09
2014	6	5.94
2015	6.67	5.80
2016	6.54	5.96
2107	7.08	6.71
2018	8.92	7.82
MAPE		6.66%

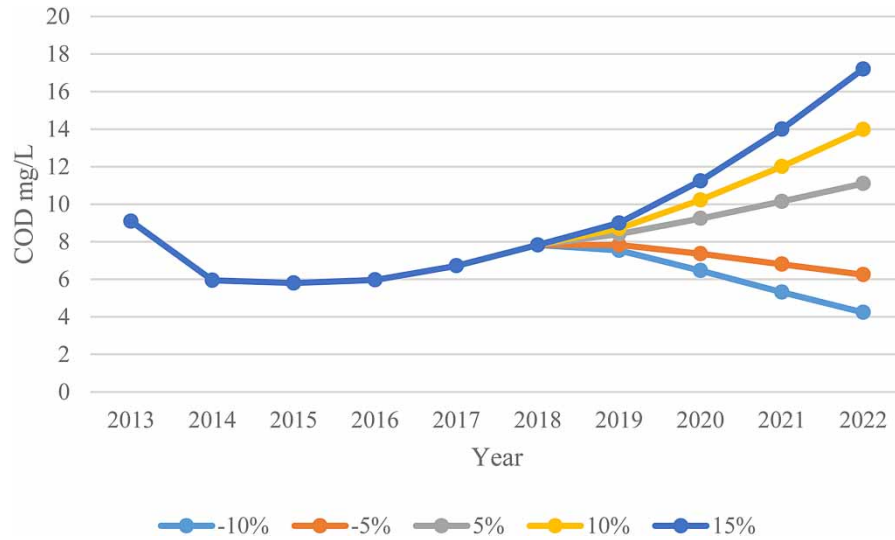


Figure 7 | The predicted COD based on tertiary industry growth rates.

perspective that tertiary industry growth rates of up to 15% could be sustained up to 2023 but that adverse impacts would arise in subsequent years depending on the growth rate.

4.4. Predicted COD based on population

Table 10 shows the annual average of population from 2013 to 2018. It can be seen from the data that the population first increases and then decreases, and the data basically tends to be stable. Following the calculation procedure set out in Section 4.1, the results are shown in Table 11. The MAPE of the DGMC(1,2) is 2.24%.

The growth rates of the population from 2013 to 2018 were 1.21%, 1.67%, -4.36%, -0.19% and 0.69%. From the data from 2013 to 2018, it can be concluded that the population has a trend of a slow decline. Because the population of China is influenced by national policies, growth rates of -5% and 5% were assessed. The results are shown in Figure 8. It can be seen from Figure 8 that when the population growth rate is -5%, the COD experiences a stable decline. When the growth rate is 5%, the

Table 10 | The population values and the concentration of COD

Year	2013	2014	2015	2016	2017	2018
Population(10,000 people)	4,149,814	4,200,000	4,270,000	4,083,917	4,076,320	4,104,390
COD (mg/L)	9.09	6	6.67	6.54	7.08	8.92

Table 11 | The simulated values and error of DGMC(1,2)

Year	Actual COD	Simulated value
2013	9.09	9.09
2014	6	6.07
2015	6.67	6.33
2016	6.54	6.95
2017	7.08	7.08
2018	8.92	8.83
MAPE		2.24%

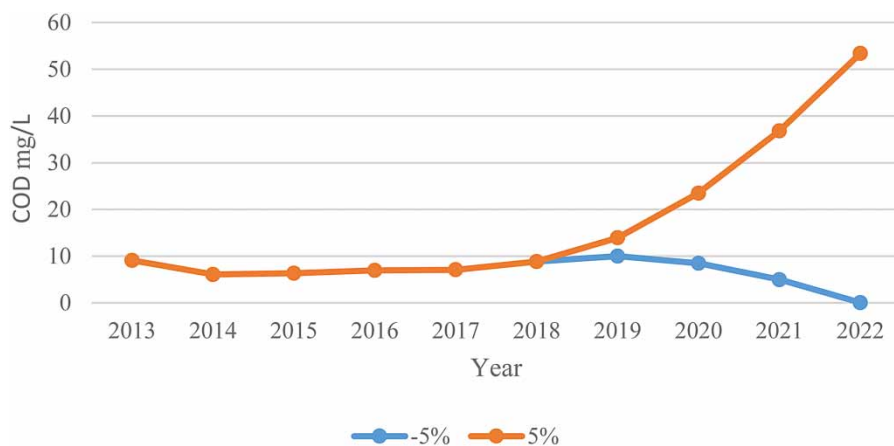


Figure 8 | The predict COD value of population.

COD shows a trend of continuous rise, reaching 53.42 mg/L by 2022. This indicates that if the growth rate of the six counties of Handan city is 5%, then this population growth would have a significant adverse impact on water quality in the Zhanghe River in the absence of any additional pollution reduction strategies.

5. CONCLUSIONS

Most research on water quality issues involves multivariate models. Through comparative analysis of DGMC(1, 2), GM(1, 2) and GM(1, 1) models, it was concluded that a DGMC(1,2) model which was able to analyze and predict the water quality of the Zhanghe River from 2013 to 2022, using COD as the indicator of water quality, to a high level of accuracy.

The DGMC(1,2) model was used to analyze the relationship between COD in the upper reaches of the Zhanghe River and the added value of the primary, secondary and tertiary industries as well as population. It was found that growth of primary, secondary and tertiary industries, as well as population, would all adversely impact on water quality in the in the absence of any additional pollution reduction strategies.

Under an assumed growth rate of 5% the ranking of the adverse impact on COD in 2022 (highest to lowest) would be population (53.42 mg/L), primary industry (45.06 mg/L), industrial development (secondary) (14.30 mg/L), tertiary industry (11.09 mg/L), and the construction industry (secondary) (8.87 mg/L).

While the model can also be used to inform decision makers in other cities of the primary sources of water quality problems in rivers and to help local governments focus broadly on pollution reduction strategies, the uncertainties of the social economy and the limitations of the model mean that more detailed models calibrated to local conditions should be used to develop pollution reduction strategies which have the greatest potential to deliver environmental benefits while sustaining the economy.

ACKNOWLEDGEMENTS

The relevant researches are supported by the National Natural Science Foundation of China (71871084, U20A20316) and the Excellent Young Scientist Foundation of Hebei Education Department (SLRC2019001) and the Natural Science Foundation of Hebei Province (E2020402074).

COMPETING INTERESTS

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Bao, P. Q., Reza, M., Thuy, L. N. T., Meriame, M., Ameneh, P., Sh, S. S. & Van Thai, N. 2020 Application of soft computing to predict water quality in wetland. *Environmental Science and Pollution Research* **28** (1), 185–200.
- Bilgaev, A., Dong, S., Li, F., Cheng, H., Sadykova, E. & Mikheeva, A. 2020 Assessment of the current eco-socio-economic situation of the baikal region (Russia) from the perspective of the Green economy development. *Sustainability* **12** (9), 3767. <https://doi.org/10.3390/su12093767>.
- Choi, J., Hearne, R. R., Lee, K. & Roberts, D. C., 2015 The relation between water pollution and economic growth using the environmental Kuznets curve: a case study in South Korea. *Water International* **40** (3), 499–512.
- Cullis, J. D. S., Horn, A., Rossouw, N., Fisher-Jeffes, L., Kunneke, M. M. & Hoffman, W. 2019 Urbanisation, climate change and its impact on water quality and economic risks in a water scarce and rapidly urbanising catchment: case study of the Berg River Catchment. *H2Open Journal* **2** (1), 146–167. <https://doi.org/10.2166/h2oj.2019.027>.
- de Mello, K., Taniwaki, R. H., de Paula, F. R., Valente, R. A., Randhir, T. O., Macedo, D. R. & Hughes, R. M. 2020 Multiscale land use impacts on water quality: assessment, planning, and future perspectives in Brazil. *Journal of Environmental Management* **2020**, 270. <https://doi.org/10.1016/j.jenvman.2020.110879>.
- du Plessis, Harmse & Ahmed, 2015 Predicting water quality associated with land cover change in the Grootdraai Dam catchment, South Africa. *Water International* **40** (4), 647–663. <https://doi.org/10.1080/02508060.2015.1067752>.
- Floress, K., de Jalón, S. G., Church, S. P., Babin, N., Ulrich-Schad, J. D. & Prokopy, L. S. 2017 Toward a theory of farmer conservation attitudes: dual interests and willingness to take action to protect water quality. *Journal of Environmental Psychology* **2017** (53), 73–80. <https://doi.org/10.1016/j.jenvp.2017.06.009>.
- Goher, M. E., Mahdy, E.-S. M., Abdo, M. H., El Dars, F. M. & Elsherif, A. A. S. 2019 Water quality status and pollution indices of Wadi El-Rayan lakes, El-Fayoum, Egypt. *Sustainable Water Resources Management* **5** (2), 387–400.
- Hu, Y. 2020 A multivariate grey prediction model with grey relational analysis for bankruptcy prediction problems. *Soft Computing* **24** (6). <https://doi.org/10.1007/s00500-019-04191-0>.
- Jiang, H., Liu, X. & Li, Y. 2019 Medium and long term electric quantity prediction of provincial power companies based on multivariable grey prediction model. *Power Demand Side Management* **21** (01), 37–41.
- Kuwayama, Y., Olmstead, S. M., Wietelman, D. C. & Zheng, J. 2020 Trends in nutrient-related pollution as a source of potential water quality damages: a case study of Texas, USA. *Science of the Total Environment* **724**. <https://doi.org/10.1016/j.scitotenv.2020.137962>
- Kyei, C. & Hassan, R. 2019 Managing the trade-off between economic growth and protection of environmental quality: the case of taxing water pollution in the Olifants river basin of South Africa. *Water Policy* **21** (2), 277–290.
- Li, H. & Lu, J. 2020 Can regional integration control transboundary water pollution? A test from the Yangtze River economic belt. *Environmental Science and Pollution Research* **27** (22), 28288–28305.
- Li, J., Li, Z., Zheng, J., Zan, M., Cui, Y. & Liu, X. 2016 Gray GM(1,2) prediction model for mine water inflow. *Journal of Henan University of Technology (Natural Science Edition)* **35** (3), 368–372.
- Liu, L. & Wu, L. 2021 Forecasting the renewable energy consumption of the European countries by an adjacent non-homogeneous grey model. *Applied Mathematical Modelling*. **89**, 1932–1948. <https://doi.org/10.1016/j.apm.2020.08.080>.
- Liu, Y., Yang, L. & Jiang, W. 2020 Qualitative and quantitative analysis of the relationship between water pollution and economic growth: a case study in Nansi Lake catchment, China. *Environmental Science and Pollution Research* **27** (4), 4008–4020.
- Mustafa, H. M., Kurniawan, I. & Hayder, G. 2021 Predicting water quality parameters in a complex river system. *Journal of Ecological Engineering* **22** (1), 250–257. <https://doi.org/10.12911/22998993/128691>.
- Najafzadeh, M. & Emamgholizadeh, S. 2019 Prediction of water quality parameters using evolutionary computing-based formulations. *International Journal of Environmental Science and Technology* **16** (10), 6377–6396. <http://dx.doi.org/10.1007/s13762-018-2049-4>.
- Najafzadeh, M., Ghaemi, A. & Emamgholizadeh, S. 2019 Prediction of water quality parameters using evolutionary computing-based formulations. *International Journal of Environmental Science and Technology* **16** (10), 6377–6396.
- Nguyen, H. T., Aviso, K. B., Kojima, N. & Tokai, A. 2018 Structural analysis of the interrelationship between economic activities and water pollution in Vietnam in the period of 2000–2011. *Clean Technologies and Environmental Policy* **20** (3), 621–638.
- Pires, P. F., Feitosa, T. M., Krüger, B. C., Soares, G. H. A. & DA, S. L. B. P. 2020 Seasonal assessment of water quality parameters in Mirim Lagoon, Rio Grande do Sul State, Brazil. *Anais da Academia Brasileira de Ciências* **92** (3), e20181107–e20181107. <http://dx.doi.org/10.1590/0001-3765202020181107>.
- Saberi-Movahed, F. & Mehrpooya, A. 2020 Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions. *Water Resources Management* **34** (2), 529–561. <https://doi.org/10.1007/s11269-019-02463-w>.
- Şahin, U. 2019 Forecasting of Turkey's greenhouse gas emissions using linear and nonlinear rolling metabolic grey model based on optimization. *Journal of Cleaner Production* **239**. <https://doi.org/10.1016/j.jclepro.2019.118079>.
- Shen, Y., Zhang, S., Wu, L., Lu, J. & Guanghui, F. 2020 Application of grey topological model in water Quality prediction of artificial Reef area in Haizhou Bay. *Southern Fisheries Science* **16** (2), 77–86.
- Utkucan 2021 Future of renewable energy consumption in France, Germany, Italy, Spain, Turkey and UK by 2030 using optimized fractional nonlinear grey Bernoulli model. *Sustainable Production and Consumption* **25**. <https://doi.org/10.1016/j.spc.2020.07.009>.

- Vita, P., Slamet, S. & Setyawan, P. 2018 Strategy of water pollution control base on social economic activity, in Karang Mumus River, Samarinda East Kalimantan, Indonesia. *E3S Web of Conferences* **31**, 04007.
- Wang, Z. & Hao, P. 2016 An improved grey multivariable model for predicting industrial energy consumption in China. *Applied Mathematical Modelling* **40** (11–12), 5745–5758.
- Wu, L. & Zhao, H. 2019 [Discrete grey model with the weighted accumulation](#). *Soft Computing* **23** (23), 12873–12881. <https://doi.org/10.1007/s00500-019-03845-3>.
- Xinzi, W., Kejia, W., Jiamu, D., Xinqi, C. & Wenlong, Z. 2020 [Predicting water quality during urbanization based on a causality-based input variable selection method modified back-propagation neural network](#). *Environmental Science and Pollution Research* **28** (1), 960–973. [https://doi.org/10.1016/0006-291x\(75\)90518-5](https://doi.org/10.1016/0006-291x(75)90518-5).
- Yin, P., Han, L. & Yi, J. 2017 The application of gray forecasting system GM(1,1) model in the drought prediction in Luoyang area. *Groundwater* **39** (6), 191–194.
- Yuan, Y., Zhao, H., Yuan, X. & Lei, X. 2019 [Application of fractional order-based grey power model in water consumption prediction](#). *Environmental Earth Sciences* **78** (8), 1–8. <https://doi.org/10.1007/s12665-019-8257-5>.
- Zhang, L., Zhang, R., Wang, Z. & Yang, F. 2020 [Spatial heterogeneity of the impact factors on gray water footprint intensity in China](#). *Sustainability* **12** (3), 865.
- Zhong, Z., Yang, C., Cao, W., Yan, C. & Sheng, W. 2017 [Short-Term photovoltaic power generation forecasting based on multivariable grey theory model with parameter optimization](#). *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2017/5812394>.

First received 29 March 2021; accepted in revised form 25 June 2021. Available online 8 July 2021