Journal of Sustainability Science and Management
Volume 19 Number 8, August 2024: 103-115
eISSN: 2672-7226

THE PREDICTOR-CORRECTOR BACKWARD DIFFERENCE FORMULAS TO ESTIMATE THE TREND OF RIVER WATER QUALITY

NUR FARAH NAJIHA RONIZAM[1], SITI AINOR MOHD YATIM[2] AND ISKANDAR SHAH MOHD ZAWAWI[1]*

[1]School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia. [2]School of Distance Education, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia.

*Corresponding author: iskandarshah@uitm.edu.my
http://doi.org/10.46754/jssm.2024.08.008
Received: 10 March 2023 Accepted: 12 June 2024 Published: 15 August 2024

Abstract: The time-series trend analysis is a mathematical technique that attempts to estimate future data movement based on observed previous and current data. This can include analysing long-terms changes in river water quality in order to emphasise river management in the future. Therefore, in this study, the backward difference formula (BDF) with forward and central difference approximations is derived to estimate a of 20-year trend in water quality index (WQI) data at the Selangor River basin in Malaysia. As an adaptation of the conventional BDF and predictor-corrector scheme, the Euler's method with a central difference formula is introduced as a predictor formula to forecast future values. Subsequently, BDF with central and forward difference formulas are specifically designed as corrector formulas to adjust the predicted future values. The performance of the method is evaluated with absolute error and average error accuracy measure indicators. The numerical results validated the proposed method and proved its reliability in estimating the time series trend analysis owing to its reasonable accuracy. In conclusion, the outcomes of this study can provide the authorities with a technical reference for monitoring, maintaining and improving the river water quality in Selangor and other states.

Keywords: Backward difference formula, predictor, corrector, water quality index, trend analysis.

Introduction

River are a precious source of water for all life on this planet. Apart from being a vital source of fresh drinking water, river water also functions as an aquatic habitat, transportation and as a means of providing irrigation for agricultural purposes. However, negligence in the management, maintenance and improvement of water resources has led to a degradation in water quality, which has put pressure on the Department of Environment (DOE), the Department of Irrigation and Drainage (DID), the Selangor Water Management Board, and the National Water Services Commission, as the authorities responsible for the upkeep of the water sources, including the Selangor River basin. Water quality management and monitoring programmes are vital for obtaining helpful information on the condition of the river. To initiate a serious attempt to handle this issue, careful and detailed management practices are essential while taking cost-effective and financial matters into account (Inyinbor Adejumok *et al.*, 2018).

There are various water quality classification standards that have been introduced worldwide. In Malaysia, the DOE has introduced the water quality standards and index as follows:

- Malaysia Marine Water Quality Standards (MMWQS)
- Malaysia Ground Water Quality Standards (MGWQS)
- The National Water Quality Standards (NWQS)

As stated on the DOE official website, the primary goals of MMWQS are to protect and maintain natural aquatic ecosystems for the benefits they offer to society. Meanwhile, the purpose of the MGWQS is to offer precise, concise and accessible information on the water quality to protect and preserve the needs of groundwater users. The WQI provides a basis for evaluating environmental water quality, whereas the NWQS classifies the benefits of the watercourse using WQI (Najah *et al*., 2021). The DOE uses the water quality index (WQI) and NWQS to measure the water quality and classify rivers in the state (Zawawi *et al*., 2022). There are 72 parameters to be considered in classifying the water quality, but only six parameters, namely Bio-chemical Oxygen Demand (BOD), Dissolved Oxygen (DO) and Chemical Oxygen Demand (COD), Ammonia Nitrogen (AN), Total Suspended Solid (SS), and Acidic and Alkaline (pH) are considered by DOE to determine the WQI. The WQI introduced by DOE is given as follows:

$$\begin{aligned} WQI = {} & 0.22\ (SIDO) + 0.199\ (SIBOD) \\ & + 0.16\ (SICOD) + 0.15\ (SIAN) \\ & + 0.16\ (SISS) + 0.12\ (SIpH) \end{aligned} \tag{1}$$

where "SIDO" is subindex DO, "SIBOD" is subindex BOD, "SICOD" is subindex COD, "SIAN" is subindex AN, "SISS" is subindex SS, and "SIpH" is subindex pH.

In Malaysia, the water quality and parameter classifications are shown in Table 1 and Table 2.

Table 1: The WQI status and index range

WQI Status	Index Range
Clean	81-100
Slightly polluted	60-80
Very polluted	0-59

Class I requires no water treatment. Class II only requires conventional treatment while Class III requires extensive water treatments; the aquatic is labelled as common and tolerant species. Class IV is for irrigation. For rivers with parameters that fall below the Classes I through IV; will fall into Class V or labelled as polluted.

Previous studies on river water quality have been done using various methods, including a compilation of artificial intelligence (AI) models (Tiyasha *et al*., 2020) such as fuzzy inference system, genetic algorithm and knowledge-based system which was applied in modelling water quality system as well as the autoregressive integrated moving average (Katimon *et al*., 2018). Then, Salih *et al*. (2021) simulated the water quality parameters by using stochastic model based on vector autoregression. The Bayesian network (BN) model is also commonly applied when it comes to forecasting. Nodoushan (2018) used BN and artificial neural network (ANN) models in predicting the water quality parameters in Honolulu, Pacific Ocean. Both models were chosen due to their proficiency in developing models based on existing data which will further predict water quality parameters in coastal waters. The BN is better than ANN due to its ability to handle missing data. Kadam *et al*. (2019) predicted the WQI value of the Shivganga River basin in India using a multiple linear regression model and compared it with the ANN model. In recent year, Yan *et al*. (2020) combined BN and regression models to build the particle swarm optimisation-DBN-least squares support vector regression model, a prediction model on water quality.

Table 2: The WQI and its parameters classification in Malaysia

Parameter	Unit	Class			
		I	II	III	IV
BOD	mg/l	1	3	6	12
COD	mg/l	10	25	50	100
DO	mg/l	7	5-7	3-5	< 3
AN	mg/l	0.1	0.3	0.9	2.7
pH	-	6.5-8.5	6-9	5-9	5-9
SS	mg/l	25	50	150	300
WQI	-	< 92.7	78.5-92.7	51.9-76.5	31.0-51.9

Investigating the pattern of long-term water quality changes may help better risk management in the future. To achieve this, predicting the direction of data using a mathematical approach is convenient as it can minimise the maximum prediction error (Kliestik, 2018). However, there is a limitation on statistical or mathematical methods where these analytical methods may not be able to solve complicated real-life problems. This is where the numerical methods are handy since they provide an accurate approximation of the solution in which the initial value problem (IVP) exists. Numerical methods can be divided into one-step and the multistep methods. The Euler's method and the Runge-Kutta method are categorised as one-step method, while Adams-Bashforth and Adams-Moulton methods are classed as multistep methods. The advantage of the multistep method over the one-step method is that it computes multiple points at several steps, improves accuracy by using previously calculated points, and permits the solution to achieve the same precision with fewer computations.

Explicit and implicit methods are two fundamental approaches that are indispensable in numerical analysis, especially when they describe the behaviour of dynamical systems. The explicit method predicts the initial approximation and the implicit method corrects the prediction. In a simulation with time dependency, both implicit and explicit methods are typically used due to their stability for time steps (Suárez-Carreño & Rosales-Romero, 2021). The explicit approach does not necessitate iterative steps at each time increment, but it mandates a smaller step size to maintain conditional stability. In contrast, implicit methods offer unconditional stability, permitting the use of larger time increments (Kim, 2020). Among the commonly employed explicit techniques are Forward Euler's method, along with other explicit methods like Heun's formula, Improved Euler's formula, the Runge-Kutta method, and the Trapezoidal method.

Over the years, many researchers have used a famous implicit multistep method known as the backward difference formula (BDF) with more advancement and precision. Several different approaches such as Taylor series expansion, Newton's backward difference, and Lagrange interpolation can be used to derive this method. Regardless of the technique used, the general BDF equation is expressed as follows:

$$\sum_{i=1}^{k} A_i y_{n+1} = hf_{n+k}\,, \tag{2}$$

where k is the order of the BDF method, A is the constant of a solution obtained from the current and previous points, and h as the step size. By changing the $y_n = y(x_n)$ and assuming $k = 1$, the first order BDF is given as follows:

$$y_{n+1} = y_n + hf(x_{n+1}, y_{n+1}), \tag{3}$$

where y_{n+1} is the approximated future value, $f(t_{n+1}, y_{n+1})$ is the approximated solution of the IVP of the ordinary differential equations (ODEs). The first order BDF is also known as the backward Euler's method while the second order BDF is as follows:

$$y_{n+1} = \frac{4}{3}y_n - \frac{1}{3}y_{n-1} + \frac{2}{3}hf(x_{n+1}, y_{n+1}). \tag{4}$$

BDF has been used to estimate the time derivative as a linear function of the solution at one or more previous points. This is possible due to its time-stepping scheme that may increase the approximation accuracy if the time and function of the problem are given. Some applications of the BDF can be observed solving parabolic equations (Akrivis & Li, 2016), wind-induced vibration problems (Chen *et al*., 2017), time-delay differential equations (Zhang *et al*., 2019), stiff ODEs (Ibrahim & Zawawi, 2019), and chemical kinetics (Chen *et al*., 2023). Second order BDF and backward Euler's method were applied by Cai *et al*. (2019) to solve the multi-group neutron space-dependent kinetics equation.

The implementation of BDF with central and forward difference approximation was proposed by Momoniat *et al*. (2013) in improving the prediction of the direction of market related data. The methods are applied to correct the predicted values and ensure the values approximate the actual data. As a result, the accuracy of the prediction improved by approximately 30%.

The essence of Momoniat *et al*. (2013) leads to the development of predictor-corrector scheme based on BDF which provides approximate values for the WQI data. This paper aims to estimate the trend of water quality in Selangor River using the two-steps BDF with forward and central difference approximations for the corrector formulas while three-steps backward formula and backward Euler's method are employed as the predictor formulas. This is appropriate because it depends on previous values of dataset, which allow producing the corrected future value. This characteristic makes the method ideal for estimating actual WQI data and improving upon already existing trends. It is also crucial to investigate the accuracy of the proposed methods in terms of absolute and average errors as it has never been studied before.

Materials and Methods

Data Information

Selangor is one of the most populous states in Malaysia as a result of its socio-economic and infrastructural development, which has resulted in a rapid decline in river water quality. The main four river basins in Selangor as stated in the State of River (SOR) report published on the LUAS official website are Sungai Langat, Sungai Klang, Sungai Bernam, and Sungai Selangor with Selangor River basin being the second largest river basin. The Selangor River has closed Phases 13 and Rantau Panjang stations due to the pollution of the natural water sources. The Selangor River basin is located at the north of the state. The river flows from Fraser's Hill to the coast in Kuala Selangor. The 110 km long river crosses three districts: Gombak, Kuala Selangor, and Hulu Selangor. In addition, the river has an area of 2,200 km^2 with over 290 streams across the three districts. The river contains ten sub-basins and covers about 28% of the state. The river also has 13 main tributaries and eight lesser tributaries. Figure 1 illustrates the Sungai Selangor River basin and its sub-basin.

Data Collection

The WQI data of Selangor River in Selangor, Malaysia is accessed from the Environmental Quality Report (EQR) year 1999 to 2018, as summarised in Table 3.

The data contains the WQI value, the status of the river, and the classification of the river. For the status of the river, where SP stands for "Slightly Polluted" status and "C" stands for "Clean". The data in the class attribute refers to the river level that may require action.

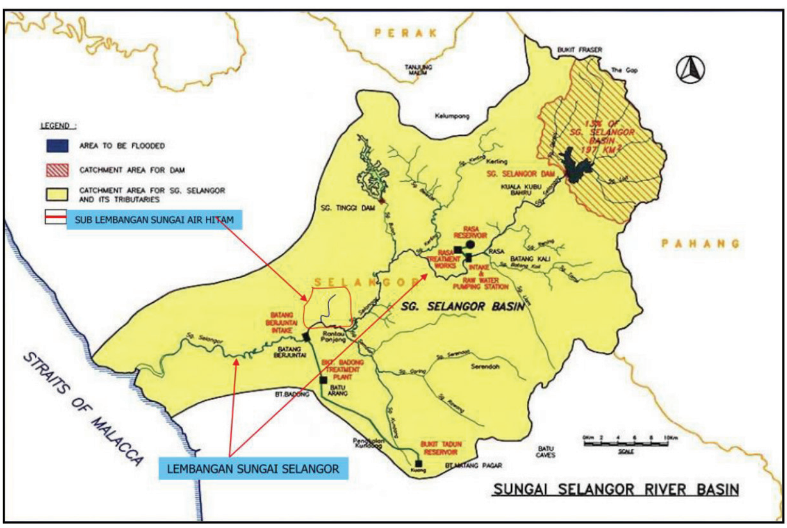

Figure 1: Sketch of Selangor River basin (Baharuddin, 2017)

Table 3: DOE-WQI data classification of Selangor River from 1999 to 2018

Year	n-year	WQI	Status	Class
1999	1	80	SP	II
2000	2	79	SP	II
2001	3	83	C	II
2002	4	84	C	II
2003	5	83	C	II
2004	6	83	C	II
2005	7	81	C	II
2006	8	85	C	II
2007	9	86	C	II
2008	10	86	C	II
2009	11	83	C	II
2010	12	82	C	II
2011	13	83	C	II
2012	14	84	C	II
2013	15	83	C	II
2014	16	81	C	II
2015	17	80	SP	II
2016	18	83	C	II
2017	19	79	SP	II
2018	20	82	C	II

Derivation of the Predictor-corrector BDF Method

The proposed method is based on BDF method with predictor-corrector scheme. The derivation involves three previous values to compute a solution at the future point. This section comprises two subsections: The derivation of two predictor formulas and two corrector formulas.

Derivation of the Predictor Formulas

The application of interpolation approaches makes it possible to produce output at multiple and arbitrary points. In practice, it is often sufficient to find the polynomial that goes through all the points or data set, with less computational difficulty, for instance the Lagrange interpolation polynomial, $P_k(x)$ of degree k, which is defined as follows:

$$P_k(x) = \sum_{j=0}^{k} L_{k,j}(x)\, y\left(x_{n+1-j}\right), \tag{5}$$

where $L_{k,j}(x) = \prod_{\substack{i=0 \\ i\neq j}}^{k} \frac{\left(x - x_{n+1-i}\right)}{\left(x_{n+1-j} - x_{n+1-i}\right)}$ for each

$j = 0, 1, ..., k.$

The first stage of interpolation entails three previous points, x_n, x_{n-1}, and x_{n-2} to compute the future value, y_{n+1} at the point x_{n+1}. The coefficients are determined by substituting the interpolating points x_{n-2}, x_{n-1}, x_n and x_{n+1}, in (5), the resulting polynomial is conventionally given by:

$$P(x) = \frac{(x - x_{n-1})(x - x_n)}{(x_{n-2} - x_{n-1})(x_{n-2} - x_n)} y_{n-2} + \frac{(x - x_{n-2})(x - x_n)}{(x_{n-1} - x_{n-2})(x_{n-1} - x_n)} y_{n-1} + \frac{(x - x_{n-2})(x - x_{n-1})}{(x_n - x_{n-2})(x_n - x_{n-1})} y_n. \tag{6}$$

Replace $x = x_{n+1} + sh$ into (6) where s is the number of step size, h yields:

$$P(x_{n+1}+sh)=\frac{(x_{n+1}+sh-x_{n-1})(x_{n+1}+sh-x_n)}{(x_{n-2}-x_{n-1})(x_{n-2}-x_n)}y_{n-2}+\frac{(x_{n+1}+sh-x_{n-2})(x_{n+1}+sh-x_n)}{(x_{n-1}-x_{n-2})(x_{n-1}-x_n)}y_{n-1}+\frac{(x_{n+1}+sh-x_{n-2})(x_{n+1}+sh-x_{n-1})}{(x_n-x_{n-2})(x_n-x_{n-1})}y_n. \quad (7)$$

Since h is assumed to be constant, the following assignments are made: $3h = x_{n+1} - x_{n-2}$, $2h = x_n - x_{n-2}$, $h = x_{n-1} - x_{n-2}$, $-h = x_{n-2} - x_{n-1}$, $-2h = x_{n-2} - x_n$, $h = x_n - x_{n-1}$, $h = x_{n+1} - x_n$, $2h = x_{n+1} - x_{n-1}$, and $-h = x_{n-1} - x_n$. Then, we obtain:

$$P(x_{n+1}+sh)=\frac{2h^2+3sh^2+s^2h^2}{-2h^2}y_{n-2}-\frac{3h^2+4sh^2+s^2h^2}{h^2}y_{n-1}+\frac{6h^2+5sh^2+s^2h^2}{2h^2}y_n \quad (8)$$

Simplify (8) will bring to:

$$P(x_{n+1}+sh)=\frac{1}{2}(2+s)(1+s)y_{n-2}-(3+s)(1+s)y_{n-1}+\frac{1}{2}(3+s)(2+s)y_n. \quad (9)$$

Let $s = 0$ to have the value at the point x_{n+1},

$$P(x_{n+1})=y_{n-2}-3y_{n-1}+3y_n. \quad (10)$$

Set $P(x_{n+1}) = y_{n+1}$ to obtain the predictor formula,

$$y_{n+1}{}^P = y_{n-2}-3y_{n-1}+3y_n, \quad (11)$$

where $y_{n+1}{}^P$ is the predicted value at the point x_{n+1}.

Therefore, the similar technique is employed to the backward Euler's method as the predictor formula:

$$y_{n+1}-y_n = hf_{n+1} \quad (12)$$

Such that f_{n+1} is the solution of the ODE, y_{n+1} is the approximated future value and y_n is the previous value. The central difference formula, $f_{n+1}=\frac{y_{n+1}-y_{n-1}}{2h}$ is chosen to be substituted into (12) since forward and backward difference will not be resulted in the solution of y_{n+1}.

Thus, the Euler-Central formula is obtained as follows:

$$y_{n+1}{}^P = 2\,y_n - y_{n-1} \quad (13)$$

Formula (13) is treated as predictor to produce $y_{n+1}{}^P$ using two previous values, y_n and y_{n-1}.

Derivation of the Corrector Formulas

The polynomial (5) is employed to derive the corrector formula on the basis of BDF method. By interpolating x_{n-1}, x_n, and x_{n+1}, the resulting polynomial is obtained:

$$P(x)=\frac{(x-x_n)(x-x_{n+1})}{(x_{n-1}-x_n)(x_{n-1}-x_{n+1})}y_{n-1}+\frac{(x-x_{n-1})(x-x_{n+1})}{(x_n-x_{n-1})(x_n-x_{n+1})}y_n+\frac{(x-x_{n-1})(x-x_n)}{(x_{n+1}-x_{n-1})(x_{n+1}-x_n)}y_{n+1}. \quad (14)$$

Consider $x = x_{n+1} + sh$ and substitute into (14), yields:

$$P(x_{n+1}+sh)=\frac{(x_{n+1}+sh-x_n)(x_{n+1}+sh-x_{n+1})}{(x_{n-1}-x_n)(x_{n-1}-x_{n+1})}y_{n-1}+\frac{(x_{n+1}+sh-x_{n-1})(x_{n+1}+sh-x_{n+1})}{(x_n-x_{n-1})(x_n-x_{n+1})}y_n+\frac{(x_{n+1}+sh-x_{n-1})(x_{n+1}+sh-x_n)}{(x_{n+1}-x_{n-1})(x_{n+1}-x_n)}y_{n+1}. \quad (15)$$

From (15), the following assignments are made: $h = x_{n+1} - x_n$, $h = x_n - x_{n-1}$, $2h = x_{n+1} - x_{n-1}$, $-h = x_{n-1} - x_n$, $-2h = x_{n-1} - x_{n+1}$, and $-h = x_n - x_{n+1}$, which leads to:

$$P(x_{n+1} + sh) = \left(\frac{s+s^2}{2}\right) y_{n-1} - (2s+s^2) y_n + \left(\frac{2+3s+s^2}{2}\right) y_{n+1}. \quad (16)$$

Differentiate (16) once with respect to s at $x = x_{n+1}$, gives:

$$P'(x_{n+1} + sh) = \left(\frac{1+2s}{2}\right) y_{n-1} - (2+2s) y_n + \left(\frac{3+2s}{2}\right) y_{n+1}. \quad (17)$$

Let $s = 0$ to obtain:

$$P'(x_{n+1}) = \frac{1}{2} y_{n-1} - 2y_n + \frac{3}{2} y_{n+1}. \quad (18)$$

Set $P'(x_{n+1}) = hf_{n+1}$ to obtain y_{n+1}, which produce the following equation:

$$y_{n+1} = -\frac{1}{3} y_{n-1} + \frac{4}{3} y_n + \frac{2}{3} hf_{n+1}. \quad (19)$$

Equation (19) is the BDF method of order two that can be used for solving ODEs, where y_{n+1} is the approximated future value, and f_{n+1} is the approximated solution of the IVP of the ODEs. However, this study does not deal with such problems. The method applied here is an attempt to approximate the solution of y instead of its derivative function, $f(x, y)$. In that case, the most important feature is the forward difference formula, $f_{n+1} = \frac{y_{n+1} - y_n}{h}$, which is substituted into (19). Thus, the BDF-Forward as corrector formula is obtained as follows:

$$y_{n+1}^{\ C} = \frac{2}{3} y_{n+1}^{\ P} + \frac{2}{3} y_n - \frac{1}{3} y_{n-1}, \quad (20)$$

where n refers to the number of years, $y_{n+1}^{\ C}$ is the corrected value, $y_{n+1}^{\ P}$ is the predicted value, whereas y_n and y_{n-1} are the previous values.

Following this, the central difference formula, $f_{n+1} = \frac{y_{n+1} - y_{n-1}}{2h}$ is substituted into (19) to obtain the BDF-Central formula:

$$y_{n+1}^{\ C} = \frac{1}{3} y_{n+1}^{\ P} + \frac{4}{3} y_n - \frac{2}{3} y_{n-1}. \quad (21)$$

It is important to emphasise that both (20) and (21) are similar given in (Momoniat *et al.*, 2013). The goal of the present study is to treat both formulas as corrector formulas and determine the predicted value $y_{n+1}^{\ P}$ from the predictor formula (11) and Euler-Central formula (13) to obtain the corrected value, $y_{n+1}^{\ C}$.

Results and Discussion

To avoid an irrational formulation, the corrector formulas cannot be applied directly for the estimation of data due to its implicit characteristic that requires the prediction of future value. From Table 3, the WQI is predicted by using predictor formula (11) and corrected using BDF-Forward (20) and BDF-Central (21). Then, the Euler-Central formula (13) is used as predictor together with corrector formulas (20) and (21). Thus, the emphasis of this section is to implement the predictor-corrector methods based on the following schemes:

PCBDFF: Predictor formula (11) and BDF-Forward (20)

PCBDFC: Predictor formula (11) and BDF-Central (21)

PCEBDFF: Euler-Central formula (13) and BDF-Forward (20)

PCEBDFC: Euler-Central formula (13) and BDF-Central (21)

To gain insight, the results are tabulated in Table 4 and Table 5 and illustrated in Figure 2 and Figure 3.

Table 4: Estimation of WQI values using PCBDFF and PCBDFC

Year	WQI	Predictor	PCBDFF	PCBDFC
1999	80	-	-	-
2000	79	-	-	-
2001	83	-	-	-
2002	84	92	90.3333	88.6667
2003	83	82	87.2222	90.2222
2004	83	80	81.3704	87.8519
2005	81	84	81.1728	84.9877
2006	85	77	78.3251	80.4156
2007	86	95	88.4925	82.2291
2008	86	84	88.8866	84.0283
2009	83	85	86.4269	85.5517
2010	82	77	79.3224	83.7168
2011	83	83	79.4060	82.2545
2012	84	86	83.8298	82.5282
2013	83	85	86.0846	83.5346
2014	81	80	82.7798	83.0273
2015	80	78	78.4917	81.0134
2016	83	80	78.0678	79.3329
2017	79	90	85.8813	81.7684
2018	82	68	76.5649	78.8025

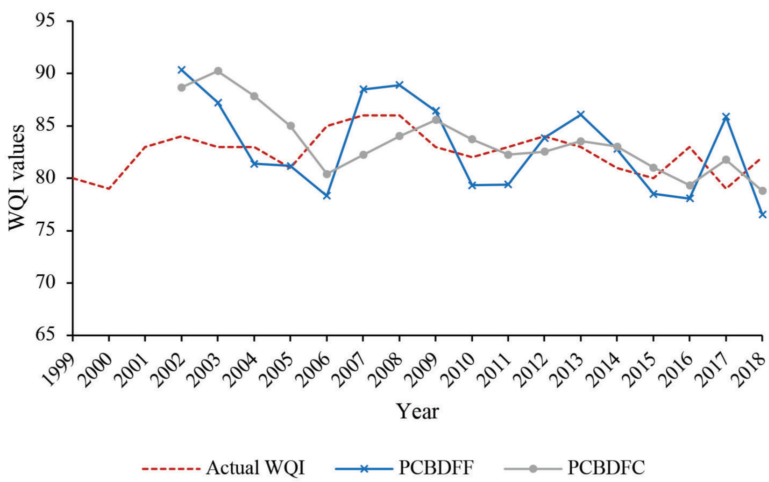

Figure 2: Graph of actual and estimated values of WQI

Table 4 exhibits the predicted and estimated WQI values between 2001 and 2018 while Figure 2 shows the graph of the actual WQI data and its estimated values. The absence of predicted values in the first three years (from 1999 to 2001) is due to the utilisation of real data as initial values in predictor formula (11) to obtain the predicted value in 2002. Then, the

Table 5: Estimation of WQI values using PCEBDFF and PCEBDFC

Year	WQI	Euler-Central (predictor)	PCEBDFF	PCEBDFC
1999	80	-	-	-
2000	79	-	-	-
2001	83	78	78	78
2002	84	87	83.6667	80.3333
2003	83	85	86.4444	83.4444
2004	83	82	84.4074	85.0370
2005	81	83	82.7901	85.4198
2006	85	79	79.7243	83.5350
2007	86	89	84.8861	84.1001
2008	86	87	88.0160	85.4435
2009	83	86	87.7153	86.5246
2010	82	80	82.4715	85.0705
2011	83	81	79.7426	82.7442
2012	84	84	81.6712	81.6120
2013	83	85	84.5333	81.9865
2014	81	82	83.7984	82.2407
2015	80	79	80.3545	81.3299
2016	83	79	78.3035	79.9461
2017	79	86	82.7508	81.0415
2018	82	75	79.0661	79.7580

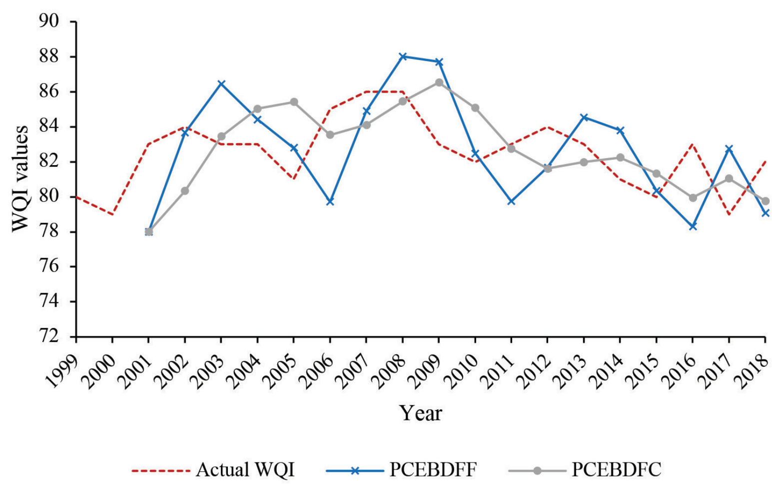

Figure 3: Results of actual and estimated WQI values

PCBDFF and PCBDFC correct the predicted values between 2002 and 2018 to obtain the estimated values of WQI. A similar pattern of trend is observed from 2002 to 2015. It shows an overall downward trend between 2002 and 2005 and has an upward trend from 2006 to 2008.

However, between 2015 and 2018, the trend is inconsistent with the actual data. In Table 5, the predicted and estimated values of WQI from 2001 to 2018 are presented. The graph of the WQI data and its estimated values is illustrated in Figure 3. There are no predicted or estimated values available in the first two years (from 1999 to 2000) because actual data is utilised as initial values in Euler-Central formula (13) to obtain the predicted value in 2001. Then, the PCEBDFF and PCEBDFC corrected the predicted values in 2001 to 2018 to obtain the estimated values of WQI.

A similar pattern of trend is observed from 2001 to 2015. In particular, the corrected WQI from the PCEBDFF formula lies on the actual WQI line chart in 2002. Between 2010 and 2015, the estimated WQI values obtained by the PCEBDFF are almost similar to the actual WQI values. It is clear that the highest estimated WQI is 88.0160, obtained by the PCEBDFF while the lowest WQI value from this method is 78. Furthermore, three estimated WQI values produced by the PCEBDFC are observed to approximate the actual WQI in 2003, 2008, and 2011. Figure 3 shows an overall downward trend between 2003 and 2005, and has an upward trend between 2006 and 2008. However, the trend is inconsistent with the actual data from 2016 and 2018. The examination of time series for the period between 2001 and 2018 shows that the trend of estimated values agree with actual data.

Error Analysis

To ascertain the accuracy of the proposed methods in estimating the WQI, the absolute error is calculated as follows:

$$\text{Error} = \left| y^c{}_i - y_i \right|, \tag{22}$$

where $y^c{}_i$ is the estimated values and y_i is the actual values. Meanwhile, the average error (AVER) is determined as:

$$\text{AVER} = \frac{\sum_{i=1}^{n} \text{Error}}{n}, \tag{23}$$

where n is the number of years. The absolute error and average error for PCBDFF, PCBDFC, PCEBDFF, and PCEBDFC are compared in Table 6.

The smallest average error is obtained by PCEBDFC with 2.2028, followed by PCEBDFF and PCBDFC. The PCBDFF reaches the highest average error. It is also discovered that the maximum error is achieved in 2003 obtained by PCBDFC with the value of 7.2222, followed by PCBDFF with 6.6749 and 6.8813 in 2006 and 2017, respectively. Although the PCBDFF has the largest average error, it has minimum error with 0.1728 in 2005, followed by PCEBDFF with 1.7901. Surprisingly, the PCEBDFC that has smaller average error holds the highest error, 4.4198 in the same year. Another proof can be seen in 2012, where the maximum error obtained by PCEBDFC is 2.388 while the minimum error achieved by PCBDFF is 0.1702. Overall, there is no significant difference between the methods in terms of accuracy. It has been accepted in this study that the PCEBDFC indicates the most accurate method in estimating the time series trend analysis due to its smallest average error.

Conclusions

In this study, the second-order BDF method has been successfully employed as corrector formula due to its previous points features. A combination of BDF with forward difference formula and central difference formula is a suitable approach to approximate the actual value of WQI. In a move to ensure better accuracy, the backward Euler's method with central difference formula is proposed as predictor formula. The results reveal that the PCEBDFC is the most accurate method in comparison to the other methods owing to it having the smallest average error. This makes the proposed methods as suitable for the prediction of river water quality trends. The conclusion drawn from the discussion is that the proposed method has effectively estimated the trend of WQI, regardless of the inconsistency in the direction of the values. The reason for this is that the BDF methods assume that the underlying function is smooth and exhibits good

Table 6: The comparison of error between proposed methods

Actual Data		Error			
Year	**WQI**	**PCBDFF**	**PCBDFC**	**PCEBDFF**	**PCEBDFC**
1999	80	-	-	-	-
2000	79	-	-	-	-
2001	83	-	-	5.0000	5.0000
2002	84	6.3333	4.6667	0.3333	3.6667
2003	83	4.2222	7.2222	3.4444	0.4444
2004	83	1.6296	4.8519	1.4074	2.0370
2005	81	0.1728	3.9877	1.7901	4.4198
2006	85	6.6749	4.5844	5.2757	1.4650
2007	86	2.4925	3.7709	1.1139	1.8999
2008	86	2.8866	1.9717	2.0160	0.5565
2009	83	3.4269	2.5517	4.7153	3.5246
2010	82	2.6776	1.7168	0.4715	3.0705
2011	83	3.5940	0.7455	3.2574	0.2558
2012	84	0.1702	1.4718	2.3288	2.3880
2013	83	3.0846	0.5346	1.5333	1.0135
2014	81	1.7798	2.0273	2.7984	1.2407
2015	80	1.5083	1.0134	0.3545	1.3299
2016	83	4.9322	3.6671	4.6965	3.0539
2017	79	6.8813	2.7684	3.7508	2.0415
2018	82	5.4351	3.1975	2.9339	2.2420
AVERAGE		**3.4060**	**2.9853**	**2.6234**	**2.2028**

behaviour. However, this assumption may not hold true for time series data, which can have irregular patterns or complex behaviours that do not follow simple linear trends. When the data shows a noticeable shift from smoothness, BDF methods may generate imprecise trend estimations. Therefore, the limits of the BDF can be attributed to the difficulties in accurately capturing non-linearities in time series data. However, the application of BDF for WQI trend prediction makes this study unique as it tends to reveal its capabilities for environmental issues. For future research, employing massive data and higher-order method may lead to more accurate estimations. As for the authorities, it is recommended that more information is provided in the published report to offer better knowledge.

Acknowledgements

This work was supported by the Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References

Akrivis, G., & Li, B. (2016). *Maximum norm analysis of implicit-explicit backward difference formulas for nonlinear parabolic equations*. http://arxiv.org/abs/1606.01338

Baharuddin, A. (2017). Integrating social inclusion and sustainability science: Empowering local community through environmental ethics/education and capacity building initiatives. In *1st MOST Senior Official's Meeting (Asia -Pasific Region)*. http://sustainabledevelopment.un.org/

Cai, Y., Peng, X., Li, Q., Wang, K., Qin, X., & Guo, R. (2019). The numerical solution of space-dependent neutron kinetics equations in hexagonal-z geometry using backward differentiation formula with adaptive step size. *Annals of Nuclear Energy*, *128*, 203-208. https://doi.org/10.1016/j.anucene.2019.01.004

Chen, J., Zeng, X., & Peng, Y. (2017). Probabilistic analysis of wind-induced vibration mitigation of structures by fluid viscous dampers. *Journal of Sound and Vibration*, *409*, 287-305. https://doi.org/10.1016/j.jsv.2017.07.051

Chen, X., Mehl, C., Faney, T., & Di Meglio, F. (2023). Clustering-enhanced deep learning method for computation of full detailed thermochemical states via solver-based adaptive sampling. *Energy & Fuels*. https://doi.org/10.1021/acs.energyfuels.3c01955

Ibrahim, Z. B., & Zawawi, I. S. M. (2019). A stable fourth order block backward differentiation formulas (α) for solving stiff initial value problems. *ASM Science Journal*, *12*(Special Issue 6).

Inyinbor Adejumok, A., Adebesin Babatunde, O., Oluyori Abimbola, P., Adelani-Akande Tabitha, A., Dada Adewumi, O., & Oreofe Toyin, A. (2018). Water pollution: Effects, prevention, and climatic impact. In *Water Challenges of an Urbanising World*. InTech. https://doi.org/10.5772/intechopen.72018

Kadam, A. K., Wagh, V. M., Muley, A. A., Umrikar, B. N., & Sankhua, R. N. (2019). Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Modeling Earth Systems and Environment*, *5*(3), 951-962. https://doi.org/10.1007/s40808-019-00581-3

Katimon, A., Shahid, S., & Mohsenipour, M. (2018). Modeling water quality and hydrological variables using ARIMA: A case study of Johor River, Malaysia. *Sustainable Water Resources Management*, *4*(4), 991-998. https://doi.org/10.1007/s40899-017-0202-8

Kim, W. (2020). An improved implicit method with dissipation control capability: The simple generalised composite time integration algorithm. *Applied Mathematical Modelling*, *81*, 910-930. https://doi.org/10.1016/j.apm.2020.01.043

Kliestik, T. (2018). *The application of mathematical modeling to predict the financial health of bussinesses*.

Momoniat, E., Harley, C., & Berman, M. (2013). On the use of backward difference formulae to improve the prediction of direction in market related data. *Mathematical Problems in Engineering*, *2013*. https://doi.org/10.1155/2013/652653

Najah, A., Teo, F. Y., Chow, M. F., Huang, Y. F., Latif, S. D., Abdullah, S., Ismail, M., & El-Shafie, A. (2021). Surface water quality status and prediction during movement control operation order under COVID-19 pandemic: Case studies in Malaysia. *International Journal of Environmental Science and Technology*, *18*(4), 1009-1018. https://doi.org/10.1007/s13762-021-03139-y

Nodoushan, E. J. (2018). Monthly forecasting of water quality parameters within Bayesian Networks: A case study of Honolulu, Pacific Ocean. *Civil Engineering Journal*, *4*(1), 188. https://doi.org/10.28991/cej-030978

Salih, S. Q., Alakili, I., Beyaztas, U., Shahid, S., & Yaseen, Z. M. (2021). Prediction of dissolved oxygen, biochemical oxygen demand, and chemical oxygen demand using hydrometeorological variables: Case study of Selangor River, Malaysia. *Environment, Development and Sustainability*, *23*(5), 8027-8046. https://doi.org/10.1007/s10668-020-00927-3

Suárez-Carreño, F., & Rosales-Romero, L. (2021). Convergency and stability of explicit and implicit schemes in the simulation of the heat equation. *Applied Sciences (Switzerland)*, *11*(10). https://doi.org/10.3390/app11104468

Tiyasha, Tung, T. M., & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000-2020. *Journal of Hydrology*, *585*, 124670. https://doi.org/10.1016/j.jhydrol.2020.124670

Yan, J., Gao, Y., Yu, Y., Xu, H., & Xu, Z. (2020). A prediction model based on deep belief network and least squares SVR applied to cross-section water quality. *Water (Switzerland)*, *12*(7). https://doi.org/10.3390/w12071929

Zawawi, I. S. M., Haniffah, M. R. M., & Aris, H. (2022). Trend analysis on water quality index using the Least Squares Regression Models. *Environment and Ecology Research*, *10*(5), 561-571. https://doi.org/10.13189/eer.2022.100504

Zhang, Y., Guo, J., Qiu, B., & Li, W. (2019). Zhang neural dynamics approximated by backward difference rules in form of time-delay differential equation. *Neural Processing Letters*, *50*(2), 1735-1753. https://doi.org/10.1007/s11063-018-9956-8