

Article

Assessment of Water Quality Target Attainment and Influencing Factors Using the Multivariate Log-Linear Model in the Nakdong River Basin, Republic of Korea

Mi-Ah Kim ¹, Eunhye Na ², Suyoung Park ¹, Taegu Kang ¹ and Soyoung Lee ^{3,*}

¹ Water Quality Assessment Research Division, Water Environment Research Department, National Institute of Environmental Research, Hwangyeong-ro 42, Seo-gu, Incheon 22689, Korea; krmi@korea.kr (M.-A.K.); psuyoung@korea.kr (S.P.); taegu98@korea.kr (T.K.)

² Yeongsan River Research Center, National Institute of Environmental Research, Cheomdangwagi-ro 208, 5, Buk-gu, Gwangju 61011, Korea; eunye@korea.kr

³ Han River Research Center, National Institute of Environmental Research, 42 Dumulmeori-gil 68beon-gi, Yangseo-myeon, Yangpyeong-gun, Incheon 12585, Korea

* Correspondence: leeso1018@korea.kr; Tel.: +82-31-7707221; Fax: +82-31-7732268

Abstract: Because identifying the factors affecting water quality is challenging, water quality assessment of an individual component based on the arithmetic mean method cannot adequately support management policies. Therefore, in this study, we assessed the water quality target attainment at 24 sites in the Nakdong River Basin by applying multivariate log-linear models to identify factors influencing water quality, including flow and season. The temporal and seasonal water quality trend and flow were also analyzed using the calculated model coefficients. Specifically, weekly data on biological oxygen demand (BOD), total phosphorous (TP), and flow during 2013–2018 were used to investigate the 2018 water quality target attainment level for this river. The significance and suitability of the models were analyzed using the F-test, root mean squared error (RMSE), mean absolute percent error (MAPE), and adjusted R^2 values. All 24 models applied in this study showed statistical significance and suitability for the prediction of BOD and TP concentrations. Moreover, flow was identified as the main factor affecting water quality and had a predominant effect on BOD and TP concentrations in tributaries and the main stream, respectively. Furthermore, among the 24 sites, BOD and TP targets were evidently attained at 18 and 17 sites, respectively.

Keywords: multivariate log-linear model; water quality target; water quality influencing factors; Nakdong River



Citation: Kim, M.-A.; Na, E.; Park, S.; Kang, T.; Lee, S. Assessment of Water Quality Target Attainment and Influencing Factors Using the Multivariate Log-Linear Model in the Nakdong River Basin, Republic of Korea. *Sustainability* **2021**, *13*, 11094. <https://doi.org/10.3390/su131911094>

Academic Editors: Sung Min Cha and José Alberto Herrera-Melián

Received: 9 July 2021

Accepted: 30 September 2021

Published: 7 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Presently, to assess the quality of public waters, each year, the Ministry of Environment (MOE) in the Republic of Korea uses the arithmetic means of biological oxygen demand (BOD) and total phosphorous (TP) concentrations based on mean annual data to calculate water quality targets and determine the extent to which these targets are being attained. Additionally, the analysis of long-term data on water quality trends is used to identify the effects of national water quality management policies on the quality of public waters [1].

Water quality assessment based on the arithmetic mean method is widely used and easy to apply. However, it has limited applicability in assessing water quality due to seasonality, non-normality of data distribution, and ambiguous decision criteria for missing data and outliers. Further, during water quality assessment, we must account for the changes in water quality due to external environmental factors, such as abnormal climate phenomena (e.g., intense heat, extreme cold, droughts, and floods) and changes in flow owing to the presence of artificial structures, such as dams and weirs.

In particular, water quality is related to flow [2–4], which may be an important influencing factor for changes in water quality [4]. In the Republic of Korea, simple log-linear

models as well as multivariate log-linear models [5] have been applied to clarify the relationship between water flow and quality in the South Han River, North Han River, and Gyeongan Stream, which flow into Lake Paldang; the suitability of these models in this regard has been compared along with the analysis of the water quality trend [6]. Additionally, various studies have been conducted to calculate loads, estimate parameters, and predict suspended particle loads using the LOADEST multivariate regression mode, which is based on log-linear models [7–9]. Regarding studies on water quality assessment based on other methods, the attainment of water quality targets has been assessed using the arithmetic assessment method, which only uses total organic carbon (TOC) data [10]. Further, the exceedance of water quality targets has also been assessed by comparing the loads calculated for each section using the load duration curve [11–15]. In other countries, more detailed studies have also been conducted to assess water quality and loads using multivariate statistical models [5,16] that can explain the influence of different factors that contribute to water quality changes, including flow fluctuation, time, and season [17–19]. Specifically, Hirsch (2014) [20] investigated the negative effect of the presence of bias in large datasets on items, such as dissolved nitrate and TP contents using the five-parameter LOADEST model (L5), the seven-parameter LOADEST model (L7), and the weighted regression on time, discharge, and season (WRTDS), which are based on the multivariate regression model developed by Cohn et al. (1992) [5]. Furthermore, in 2019, the U.S. Geological Survey (USGS) analyzed the suspended sediment, total nitrogen, and TP contents as well as loads of the Kankakee River, which is located in Shelby (India), using regression analysis [21]. In Japan, the 75 percentile is applied when evaluating the degree of achievement of organic substances, and if it is satisfied, it is considered that the target standard has been achieved [22].

However, in the Republic of Korea, water quality assessment methods that consider environmental factors, such as flow and season, have not yet been actively applied, implying that new water quality assessment methods that consider such parameters are required. Therefore, in this study, the attainment of water quality targets was determined by applying multivariate log-linear models that consider water quality, flow, and season for sites with corresponding data on flow among representative sites in the sub-basin of the Nakdong River. The water quality trend as well as its dependence on flow as a function of time and season were also assessed using model coefficient values. The results of this study contribute to the development of specific water quality management policies that can be applied in the management of water quality influencing factors in future. It is also expected that the method proposed in this study can serve as an objective assessment method that can be employed to identify the effects of water quality management policies on the realization of water quality targets.

2. Materials and Methods

This study was conducted in three steps as shown in Figure 1. First, dam water quality, stream flow, and hydrologic data corresponding to the 2013–2018 period were collected at 24 sites in the Nakdong River sub-basin. Second, a multivariate log-linear model was applied to calculate the regression coefficients of seven parameters, and the F-test was used to validate the statistical significance of the model using a statistical analysis program, SPSS (IBM spss USA; version 18.0). Third, the water quality trend as well as the main water quality influencing factors were identified using the calculated multivariate log-linear model, and the accuracy of the calculation was verified by considering the difference between the predicted and observed values. Finally, the reference flow was set from 2013 to 2016, and the attainment of the 2018 water quality target was evaluated using the models that showed statistical significance.

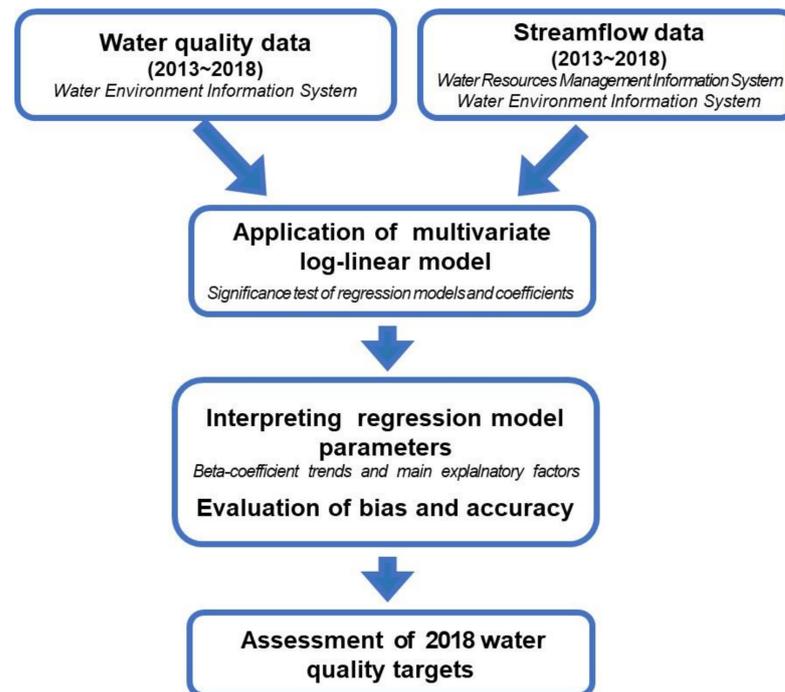


Figure 1. Flowchart representing the process used in the application of the multivariate log-linear models developed in this study.

2.1. Description of the Study Area

The Nakdong River (Figure 2), which is one of the four major rivers in the Republic of Korea, serves as an important water resource for major cities, such as Daegu, Ulsan, and Busan. It is 525 km long, and its watershed area is 23,817 km² (from latitude 35–37° N and longitude 124–131° E). Water from the Nakdong River basin supports a total population of over 13 million and is used for various purposes, including industrial and residential purposes (21.6%), agricultural purposes (51.0%), and flow maintenance (27.4%). Within the last few decades, rapid population growth coupled with industrial and urban development has resulted in the deterioration of the quality of water in this river. This is because contaminants, such as organic and inorganic materials, nitrates, and phosphates, are constantly introduced into this river [23]. Additionally, within the 2010–2011 period, eight weirs that are approximately 14–44 km apart were constructed along this river to manage water resources and control water flow [24]. The Sangju weir, which is the first, is located upstream, while the Changnyeong Haman weir is located downstream. The average annual precipitation in the central basin of this river (the Busan weather station) between 2004 and 2018 was 1050 mm, with more than half (57.5%) corresponding to the summer season, which is characterized by the highest temperatures, humidity levels, and evaporation rates [25]. Further, the Nakdong River basin occupies approximately 24% of the Korean land, and approximately 17.3% (5505 km²) of the total basin area is used for agricultural purposes. Among these, forest cover amounts to 68.6%, while wet paddy fields, dry paddy fields, urban areas, and others occupy 10.4%, 6.9%, 6.7%, and 7.4%, respectively [26]. The geological strata in the Nakdong River basin predominantly consists of sedimentary rocks, while metamorphic and igneous rocks are sparsely distributed. Of the 33 midwatershed representative sites along this river, 24 sites, with both water quality and quantity data, were selected for the assessment of the attainment of the 2018 water quality target using multivariate log-linear regression models.

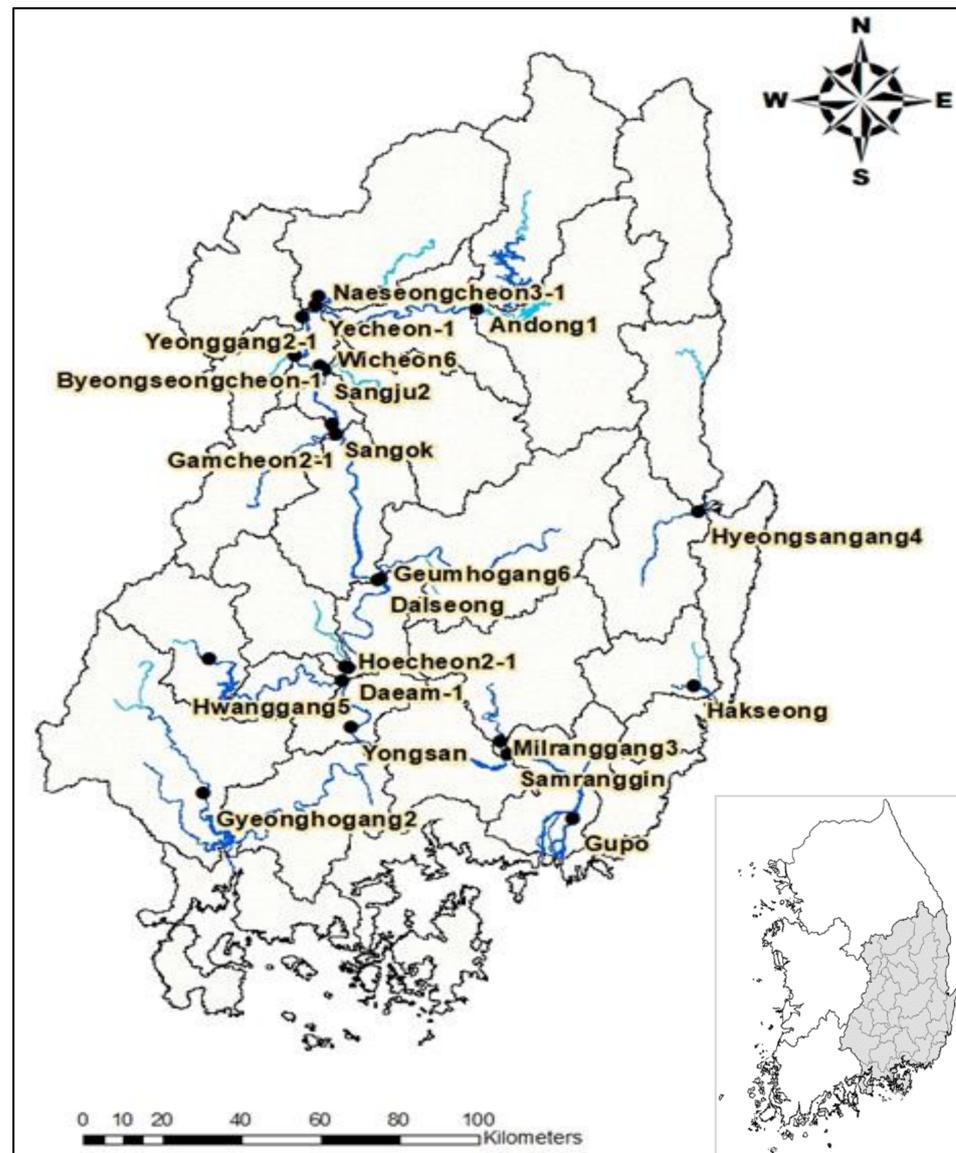


Figure 2. Map showing the location of the Nakdong River basin and study sites.

2.2. Approach Methods

2.2.1. Method of Data Collection

Three parameters, namely, BOD, TP, and stream flow, were selected for the analysis and assessment of the attainment of the 2018 water quality target. Daily data (2013–2018) on these parameters collected on a weekly basis from the 24 representative sites considered in this study were obtained from the Water Environment Information System (WEIS) [27], which is the largest national water quality database in the Republic of Korea. Further, daily dam hydrologic data were obtained from two databases, the national Water Resources Management Information System (WAMIS) [28] and My Water System [29].

Regarding flow data, K-water observation data were used for the Andong1, Sangju2, Sangok, Dalseong, Hwanggang1-1, and Samrangjin sites, while MOE flow observation data were used for the other sites. For analysis, the flow data were coupled with water quality data based on the water quality data collection day.

2.2.2. Description of Multivariate Log-Linear Model

The multivariate statistical regression model proposed by Cohn et al. (1992) [5] was used to estimate concentrations and loads on the basis of trend, discharge, and seasonality.

The model included seven independent parameters: an intercept parameter, two parameters for quadratic fit to the logarithm of discharge, two parameters for quadratic fit to time, and two parameters for the sinusoidal function of seasonality. The model was defined as:

$$\ln(C) = \beta_0 + \beta_1 \ln\left(\frac{Q}{\tilde{Q}}\right) + \beta_2 \ln\left(\frac{Q}{\tilde{Q}}\right)^2 + \beta_3(T - \tilde{T}) + \beta_4(T - \tilde{T})^2 + \beta_5 \sin(2\pi T) + \beta_6 \cos(2\pi T) + \varepsilon \quad (1)$$

where C represents concentration (mg/L), Q represents discharge (m^3/s), T represents time in decimal years, \tilde{Q} and \tilde{T} represent centering variables, β_1 and β_2 represent regression coefficient corresponding to flow, β_3 and β_4 represent regression coefficients corresponding to time, β_5 and β_6 represent regression coefficients corresponding to seasonality, and ε represents the error, which was assumed to be independent and normally distributed, with a zero mean and a constant variance. In this study, Equation (1) was used to calculate the regression coefficients corresponding to the seven parameters, and finally these regression coefficients were used to assess the attainment of the water quality target.

\tilde{Q} and \tilde{T} were defined to reduce covariance among the independent parameters and enhance estimation precision. They were calculated via simplification, without affecting the prediction results according to Equations (2) and (3), respectively [5].

$$\tilde{T} = \bar{T} \frac{\sum_{i=1}^N (T_i - \bar{T})^3}{2 \sum_{i=1}^N (T_i - \bar{T})^2} \quad (2)$$

where \tilde{T} represents the center of the calibration data, \bar{T} represents the mean of the data, T_i represents the i th sampled data, and N represents the number of observations in the calibration dataset.

$$\bar{T} = \frac{1}{N} \sum_{i=1}^N T_i \quad (3)$$

2.2.3. Evaluation Method for the Multivariate Log-Linear Model

The significance of the regression models determined using the weekly 2013–2018 water quality and flow data was determined using the F-test. In this regard, the null hypothesis, H_0 was $\beta_i = 0$, while the alternative hypothesis, H_1 , was $\beta_i \neq 0$. After determining the significance of the models (rejection of the null hypothesis based on the results of the F-test), the significance ($\alpha = 0.1$ confidence intervals) and the values of each regression coefficient were estimated by conducting t-tests. Thus, the model was determined to be highly significant when the p -value was 0.01 (1% significance level), marginally significant when the p -value ranged between 0.01 and 0.1 (1–10% significance level), and nonsignificant when it was above 0.1 (10% significance level) [5,6].

Among the calculated regression coefficients, β_1 and β_3 represent flow- and time-related coefficients, respectively, and when β_1 is positive (+), it implies that the water quality concentrations have a tendency to increase as the flow increases, indicating that the watershed is highly affected by nonpoint pollution sources. Conversely, when it is negative (−), it implies that the water quality concentrations decrease as flow increases, indicating that the watershed is highly affected by point pollution sources. Further, when the time coefficient, β_3 , is positive (+), it implies that the water quality concentrations increase with time, and it is indicative of an increase in the influence of pollution sources or a change in land use, for example, urbanization and the construction of agricultural and livestock complexes and industrial facilities. However, when it is negative (−), it implies that the water quality concentrations decrease with time, indicating an improvement in water quality owing to the influence of corresponding water management policies, such as the introduction of environmental management facilities.

The more influential of the two explanatory factors, time and flow, was considered as the main explanatory factor by comparing the absolute values of the standardized flow

and time coefficients. Additionally, the regression coefficients, β_5 and β_6 , which are related to seasonal changes, offered the possibility to identify seasonal water quality trends [6].

To minimize the bias of the estimates calculated using the multivariate regression models and improve their accuracy, the suitability of the models was assessed using the root mean square error (RMSE), mean absolute percent error (MAPE), and adjusted R^2 values corresponding to the prediction error based on the comparison of the estimated and observed values corresponding to each site. The RMSE is suitable for the verification of the reliability of the absolute prediction error between predicted and observed values [30] calculated over a short period of time, whereas the MAPE, which is the percentage average of the absolute values of the errors between the actual and predicted values, can compensate for the shortcomings of size-dependent errors, such as units. Specifically, RMSE and MAPE can be expressed as shown in Equations (4) and (5), respectively; for both of them, a lower value can be interpreted as a higher suitability with lower bias [31].

$$\text{Root mean square error (RMSE)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (H_{io} - H_{ie})^2} \quad (4)$$

$$\text{Mean absolute percent error (MAPE)} = \frac{1}{n} \sum_{i=1}^n \left| \frac{H_{ie} - H_{io}}{H_{io}} \right| \times 100 (\%) \quad (5)$$

where n represents the total number of observed data, H_{io} represents the i -th observed value, and H_{ie} represents the i -th estimated value. Table 1 shows the accuracy interpretation criteria for MAPE values proposed by Lewis (1982) [32]. MAPE values below 10 are interpreted as highly accurate forecasting, while values above 50 are considered to be inaccurate forecasting. Further, based on the model suitability assessment results, 2018 BOD and TP concentrations were estimated, and the attainment of water quality targets for this year with respect to BOD and TP at the 24 sites considered in this study were assessed using the corresponding 2018 flow data.

Table 1. Interpretation of typical mean absolute percent error (MAPE) values [32].

MAPE	Interpretation
<10	Highly accurate forecasting
10–20	Good forecasting
20–50	Reasonable forecasting
>50	Inaccurate forecasting

3. Results and Discussions

3.1. Application of Multivariate Log-Linear Model

Table 2 shows the F-test results and the adjusted R^2 values obtained using the multivariate log-linear models for each site. Specifically, the F-test results indicated that all the regression models for the 24 sites showed a highly significant linear relationship between the explanatory and response factors ($p < 0.001$). The adjusted R^2 values, which represent the explanatory power of a regression equation, ranged from 0.100 to 0.573 for the BOD models, with the explanatory power corresponding to tributaries (31.4% on average) being higher than that corresponding to the main stream (24.7% on average). Further, for the TP models, the adjusted R^2 values ranged from 0.197 to 0.803, with the explanatory power corresponding to tributaries (57.6% on average) being higher than that corresponding to the main stream (37.3% on average), as was the case with the BOD models. However, the explanatory powers corresponding to the TP models were higher than those corresponding to the BOD models.

Table 2. Significance (F-tests) and adjusted coefficients of determination (R^2) of the multivariate regression models assessing biological oxygen demand (BOD) and total phosphorous (TP) in the Nakdong River Basin, from 2013 to 2018.

Site	Biological Oxygen Demand (BOD) (mg·L ⁻¹)			Total Phosphorous (TP) (mg·L ⁻¹)		
	F	p-Value	Adjusted R ²	F	p-Value	Adjusted R ²
Andong1	27.806	<0.001	0.362	13.000	<0.001	0.200
Banbyeoncheon2-1	13.374	<0.001	0.210	38.986	<0.001	0.449
Yecheon-1	10.819	<0.001	0.222	12.503	<0.001	0.253
Naeseongcheon3-1	16.892	<0.001	0.249	191.481	<0.001	0.799
Yeonggang2-1	21.093	<0.001	0.339	26.586	<0.001	0.395
Byeongseongcheon-1	18.263	<0.001	0.306	61.599	<0.001	0.607
Sangju2	9.119	<0.001	0.214	27.109	<0.001	0.467
Wicheon6	9.394	<0.001	0.176	50.522	<0.001	0.558
Sangok	13.680	<0.001	0.274	27.223	<0.001	0.438
Gamcheon2-1	13.399	<0.001	0.240	134.039	<0.001	0.773
Dalseong	6.296	<0.001	0.100	21.031	<0.001	0.295
Geumhogang6	32.781	<0.001	0.399	101.119	<0.001	0.677
Hoecheon2-1	24.918	<0.001	0.379	79.149	<0.001	0.666
Daeam-1	8.424	<0.001	0.160	38.434	<0.001	0.491
Hwanggang1-1	12.837	<0.001	0.504	7.893	<0.001	0.375
Hwanggang5	6.162	<0.001	0.116	40.998	<0.001	0.505
Yongsan	5.951	<0.001	0.115	47.871	<0.001	0.551
Gyeonghogang2	27.419	<0.001	0.403	96.821	<0.001	0.710
Namgang4-1	24.520	<0.001	0.330	46.136	<0.001	0.486
Samrangjin	19.445	<0.001	0.278	27.334	<0.001	0.355
Milranggang3	29.914	<0.001	0.425	18.705	<0.001	0.311
Gupo	10.883	<0.001	0.171	35.621	<0.001	0.420
Hyeongsangang4	16.382	<0.001	0.244	11.476	<0.001	0.180
Hakseong	14.588	<0.001	0.573	19.902	<0.001	0.285

Furthermore, the explanatory powers of the BOD and TP models at the Geumhogang6 (40.0%) and Hoecheon2-1 (67.0%) sites in the midstream section were found to be higher than those in the upstream and downstream sections. Regarding BOD concentration, the Hakseong site in the Taehwa River watershed exhibited the highest explanatory power (57.0%), and regarding TP concentration, the Naeseongcheon3-1 site, a tributary in the midstream section of the Nakdong River, showed the highest explanatory power (79.9%).

Generally, it is known that multivariate log-linear models can explain approximately 10–50% of the variability (explanatory power) of continuously observed concentrations [5]. The multivariate log-linear models calculated in this study explained approximately 10–60 and 10–80% of BOD and TP concentrations, respectively, indicating that they are suitable for predicting parameter concentrations and identifying the water quality trend taking flow into consideration. Further, the explanatory powers of these models were found to be higher in tributaries, which are sensitive to the influence of flow, than in the main stream.

3.2. Interpreting of Model Parameters

Table 3 shows the results of the examination of the coefficients of the regression models for the Nakdong River basin and their statistical significance. Further, Table 4 shows the

results of the analysis of the water quality trend for each site within the 2013–2018 period. The flow-related regression coefficients, β_1 and β_2 , were highly significant or significant for all the sites, except for the Yongsan site, while the time-related regression coefficients, β_3 and β_4 , were highly significant or significant for all the sites, except the Yeonggang2-1, Sangju2, Daeam-1, Hwanggang5, and Yongsan sites. Furthermore, the coefficients β_5 and β_6 , which are regression coefficients corresponding to seasonality, were significant for all the sites. Specifically, the flow-related regression coefficients (β_1) of BOD and TP concentrations were more significant than their time-related regression coefficients (β_3). Further, the flow-related coefficients of TP concentration, which ranged from -0.219 to 0.501 , showed statistical significance at all sites, except for the Hwanggang1-1 and Milyanggang3 sites, while that of BOD concentration as shown in Table 3, which ranged from -0.216 to 0.256 , was statistically significant mainly at the upstream and downstream sites. Furthermore, organic matter concentration showed a tendency to decrease owing to an increase in flow rather than time, and a clear tendency of improved water quality was observed, particularly in the downstream section. Regarding BOD concentration, the regression coefficient corresponding to time (β_3) was significant at 13 out of the 24 sites. For TP concentration, it ranged from -0.266 to 0.082 and showed statistical significance for all sites except the Sangju2, Sangok, Hakseong, Hyeongsangang4, and Hakseong sites. This implies that time had an effect on nutrient concentration, which, however, was more dependent on flow than time. Unlike BOD concentration, TP concentration showed more vulnerability to the inflow of nonpoint pollutants than to the pollutant dilution effect owing to an increase in flow during rainfall. The regression coefficients for seasonality (β_5 and β_6) were also found to be statistically significant at all the sites. In Table 4, the change in BOD concentration at the main stream sites varied depending on the site. Additionally, flow and time were identified as the main explanatory factors for four and seven sites, respectively, and regarding the change in BOD concentration in the tributaries, flow was identified as the main explanatory factor for 10 out of the 13 tributary sites. Further, regarding the change in TP concentration, flow was identified as the main explanatory factor for 9 of the 11 main stream sites. It was also observed that TP concentration showed a tendency to increase across the basin but tended to decrease over time. This is considered to be due to the water quality improvement effect resulting from an increase in investments aimed at improving the TP treatment capacities of water treatment facilities for the removal of point pollutants in the main stream sections of the four major rivers in the Republic of Korea from 2009 [33,34]. Considering the 13 tributary sites, flow and time were identified as the main explanatory factors for six sites and seven sites, respectively. For the main stream sites, the main explanatory factor for TP concentration was flow rather than time, and it was different for each site in the tributaries.

Table 3. Beta-coefficient values of statistically significant BOD and TP regression models within the Nakdong River Basin from 2013–2018.

Site	Variable	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Andong1	BOD	-0.016	-0.148	-0.063	0.054	-0.005	-0.079	-0.205
	TP	-3.843	-0.219	-0.193	-0.060	0.018	-0.215	-0.358
Banbyeoncheon2-1	BOD	0.150	-0.099	0.026	-0.057	-0.019	-0.052	-0.208
	TP	-4.167	0.035	0.066	-0.115	-0.002	-0.134	-0.214
Yecheon-1	BOD	0.139	0.058	-0.005	0.097	-0.012	-0.008	0.074
	TP	-4.221	0.275	0.053	0.082	0.026	0.103	0.026
Naeseongcheon3-1	BOD	-0.148	0.256	0.049	-0.028	-0.048	0.208	-0.203
	TP	-3.070	0.501	0.030	-0.166	0.024	-0.203	-0.013
Yeonggang2-1	BOD	0.229	-0.079	0.005	-0.021	-0.032	0.237	-0.311
	TP	-4.012	0.046	0.055	-0.094	-0.008	-0.011	-0.375

Table 3. Cont.

Site	Variable	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Byeongseongcheon-1	BOD	0.344	0.102	0.064	-0.043	-0.017	0.257	-0.312
	TP	-2.965	0.304	0.090	-0.139	0.031	-0.015	-0.224
Sangju2	BOD	0.536	0.021	-0.013	0.001	0.004	0.196	0.100
	TP	-3.707	0.297	0.101	0.020	0.035	0.253	-0.064
Wicheon6	BOD	0.730	-0.017	-0.019	-0.012	0.006	-0.046	-0.304
	TP	-3.607	0.134	0.058	-0.101	0.013	-0.142	-0.332
Sangok	BOD	0.628	0.014	-0.025	0.072	0.003	0.096	0.108
	TP	-3.868	0.241	0.069	-0.005	0.022	0.324	-0.073
Gamcheon2-1	BOD	0.167	0.066	0.085	-0.105	-0.026	0.274	-0.032
	TP	-2.945	0.351	0.061	-0.266	0.039	-0.117	-0.113
Dalseong	BOD	0.827	0.003	-0.009	0.033	-0.003	0.000	-0.102
	TP	-3.601	0.144	0.039	-0.063	0.016	-0.329	-0.129
Geumhogang6	BOD	1.147	-0.140	0.094	-0.036	-0.033	0.194	-0.455
	TP	-2.478	0.178	0.084	-0.151	-0.005	-0.097	-0.337
Hoecheon2-1	BOD	0.131	-0.038	0.060	-0.039	-0.003	0.168	-0.314
	TP	-3.592	0.245	0.069	-0.092	0.023	-0.035	-0.403
Daeam-1	BOD	0.806	0.005	0.013	0.005	0.002	0.227	-0.069
	TP	-3.122	0.115	0.026	-0.119	0.006	-0.271	-0.149
Hwanggang1-1	BOD	-0.218	-0.182	0.144	-0.036	-0.013	0.305	-0.328
	TP	-3.363	0.042	0.089	-0.126	0.012	-0.087	-0.382
Hwanggang5	BOD	-0.258	0.112	0.064	-0.001	-0.054	0.028	-0.085
	TP	-3.766	0.350	0.026	-0.086	0.001	-0.191	-0.142
Yongsan	BOD	0.772	-0.019	0.002	-0.015	0.003	0.049	-0.201
	TP	-3.212	0.183	0.062	-0.142	0.015	-0.369	-0.198
Gyeonghogang2	BOD	0.063	-0.012	0.100	-0.082	-0.013	0.185	-0.234
	TP	-3.653	0.252	0.094	-0.125	-0.005	-0.030	-0.638
Namgang4-1	BOD	0.789	-0.167	-0.017	-0.003	0.008	0.015	-0.338
	TP	-3.243	0.079	0.038	-0.116	0.011	-0.134	-0.310
Samrangjin	BOD	0.611	-0.088	0.005	-0.030	0.021	0.074	-0.258
	TP	-3.165	0.246	0.071	-0.039	0.025	-0.204	-0.093
Milyanggang3	BOD	0.537	-0.216	0.003	0.011	-0.011	0.131	-0.448
	TP	-3.343	0.002	0.070	-0.087	-0.010	0.018	-0.181
Gupo	BOD	0.589	-0.041	0.011	-0.040	0.003	0.127	-0.188
	TP	-3.297	0.210	0.027	-0.038	0.050	-0.159	-0.138
Hyeongsangang4	BOD	0.669	-0.153	0.015	0.022	-0.021	-0.034	-0.339
	TP	-2.640	-0.134	0.044	-0.024	-0.012	-0.133	-0.021
Hakseong	BOD	0.522	-0.046	0.010	-0.108	0.013	0.166	-0.325
	TP	-2.664	0.088	0.017	-0.010	-0.015	-0.246	-0.205

Notes: Highly significant ($p < 0.01$): ■ Significant ($p < 0.05$): ■

Table 4. Water quality trend description and identification of main explanatory factors.

Site	Stream	Variable	Flow (Q)	Time (T)	Seasonality	Main Explanatory Factor
Andong1	Main	BOD	↓	↑	O	Q
		TP	↓	↓	O	Q
Banbyeoncheon2-1	Tributary	BOD	↓	↓	O	Q
		TP	↑	↓	O	T
Yecheon-1	Main	BOD	↑	↑	O	T
		TP	↑	↑	-	Q
Naeseongcheon3-1	Tributary	BOD	↑	-	O	Q
		TP	↑	↓	O	Q
Yeonggang2-1	Tributary	BOD	↓	-	O	Q
		TP	↑	↓	O	T
Byeongseongcheon-1	Tributary	BOD	↑	↓	O	Q
		TP	↑	↓	O	Q
Sangju2	Main	BOD	-	-	O	Q
		TP	↑	-	O	Q
Wicheon6	Tributary	BOD	-	-	O	Q
		TP	↑	↓	O	Q
Sangok	Main	BOD	-	↑	O	T
		TP	↑	-	O	Q
Gamcheon2-1	Tributary	BOD	-	↓	O	T
		TP	↑	↓	O	T
Dalseong	Main	BOD	-	↑	O	T
		TP	↑	↓	O	Q
Geumhogang6	Tributary	BOD	↓	↓	O	Q
		TP	↑	↓	O	T
Hoecheon2-1	Tributary	BOD	-	↓	O	T
		TP	↑	↓	O	Q
Daeam-1	Main	BOD	-	-	O	T
		TP	↑	↓	O	T
Hwanggang1-1	Tributary	BOD	↓	-	O	Q
		TP	-	↓	O	T
Hwanggang5	Tributary	BOD	↑	-	O	Q
		TP	↑	↓	O	Q
Yongsan	Main	BOD	-	-	O	T
		TP	↑	↓	O	T
Gyeonghogang2	Tributary	BOD	-	↓	O	T
		TP	↑	↓	O	Q
Namgang4-1	Tributary	BOD	↓	-	O	Q
		TP	↑	↓	O	T
Samrangjin	Main	BOD	↓	↓	O	Q
		TP	↑	↓	O	Q
Milyanggang3	Tributary	BOD	↓	-	O	Q
		TP	-	↓	O	T
Gupo	Main	BOD	↓	↓	O	T
		TP	↑	↓	O	Q
Hyeongsangang4	Main	BOD	↓	-	O	Q
		TP	↓	-	O	Q
Hakseong	Main	BOD	-	↓	O	T
		TP	↑	-	O	Q

Notes: -, Nonsignificant; ↓, Decrease; ↑, Increase; Q, Flow; T, time.

3.3. Evaluation of Bias and Accuracy Estimations

Table 5 shows the results of the analysis of the accuracy of the multivariate log-linear models in predicting 2018 water quality parameters. Overall, the predicted TP concentration values reflected the observed values to a greater extent than the predicted BOD concentration values. Further, the observed values were better reflected at the mid-upstream sites than at the downstream sites, and MAPE was below 50% for all the sites; this is indicative reasonable forecasting. Furthermore, the RMSE values obtained for BOD and TP concentrations were in the ranges 0.3–1.5 and 0.012–0.064 mg/L, respectively, and among the 24 sites, Andong1 and Banbyeoncheon2-1 sites exhibited the highest prediction accuracies for BOD and TP concentrations, respectively, i.e., RMSE values of 0.3 and 0.012 mg/L, respectively (Figure 3).

Table 5. Evaluation of bias and accuracy estimate methods for BOD and TP concentrations from 2013 to 2018 across the 24 main stream and tributary sites of the Nakdong River Basin.

Site	BOD		TP	
	RMSE (mg·L ⁻¹)	MAPE (%)	RMSE (mg·L ⁻¹)	MAPE (%)
Andong1	0.3	22.74	0.028	47.66
Banbyeoncheon2-1	0.7	34.18	0.012	29.31
Yecheon-1	0.4	21.33	0.064	42.75
Naeseongcheon3-1	0.8	48.24	0.025	24.28
Yeonggang2-1	0.7	35.74	0.024	33.98
Byeongseongcheon-1	1.3	40.73	0.040	26.34
Sangju2	0.6	25.83	0.015	38.08
Wicheon6	0.9	33.56	0.016	28.73
Sangok	0.5	20.32	0.014	39.85
Gamcheon2-1	0.9	45.19	0.030	23.30
Dalseong	0.6	19.82	0.023	46.84
Geumhogang6	1.5	33.31	0.030	22.51
Hoecheon2-1	0.6	33.36	0.016	26.41
Daeam-1	0.9	29.84	0.020	28.95
Hwanggang1-1	0.5	31.98	0.023	35.47
Hwanggang5	0.4	41.30	0.016	29.81
Yongsan	0.9	30.65	0.019	31.33
Gyeonghogang2	0.6	29.53	0.020	30.69
Namgang4-1	0.8	28.23	0.020	28.45
Samrangjin	0.7	24.28	0.025	33.94
Milranggang3	0.9	36.13	0.014	26.55
Gupo	0.8	29.68	0.023	33.65
Hyeongsangang4	1.1	37.40	0.040	36.28
Hakseong	1.5	48.95	0.048	33.79

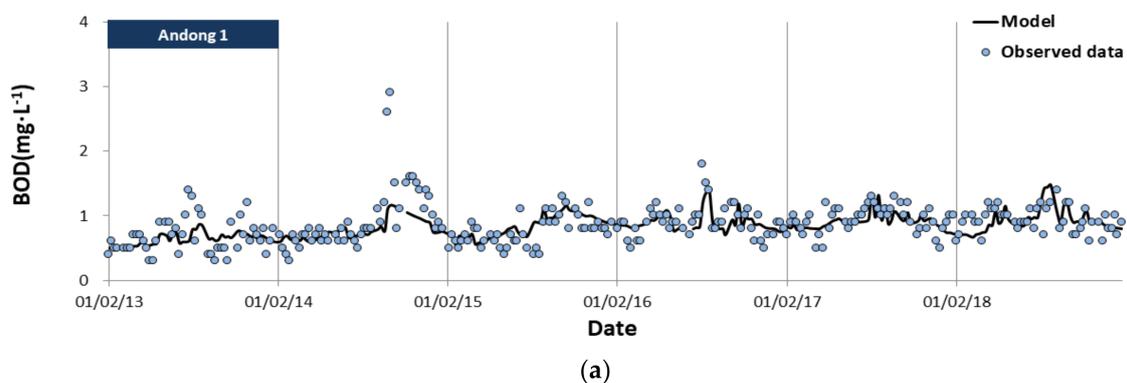


Figure 3. Cont.

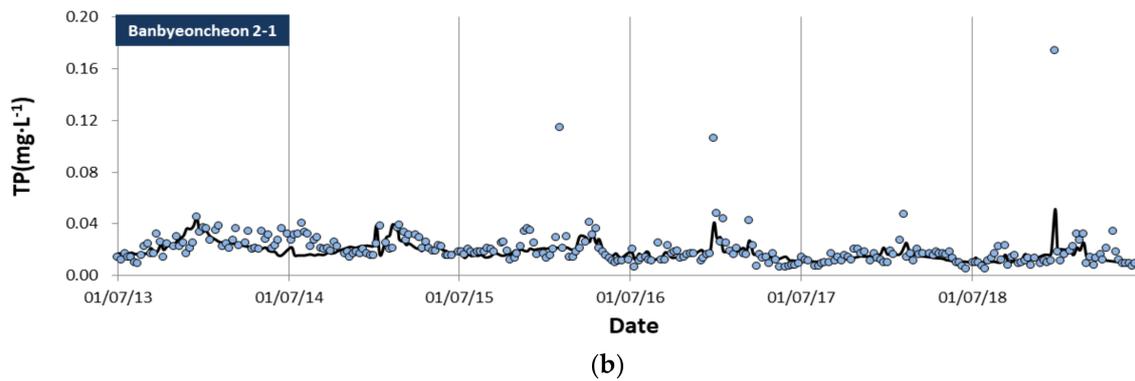


Figure 3. Comparison of observed and modeled values for the most accurate sites of (a) BOD and (b) TP from 2013–2018.

The observed and predicted 2018 data for the 24 sites were compared using the verified models as shown in Figure 4, and the comparison showed similar tendencies throughout the basin. When the accuracy of the predicted 2018 data was evaluated, the RMSE and MAPE values for BOD concentration were 0.2 mg/L and 10.6%, respectively, and for TP concentration, they were 0.007 mg/L and 13.7%, respectively, indicating the excellent suitability of the model results.

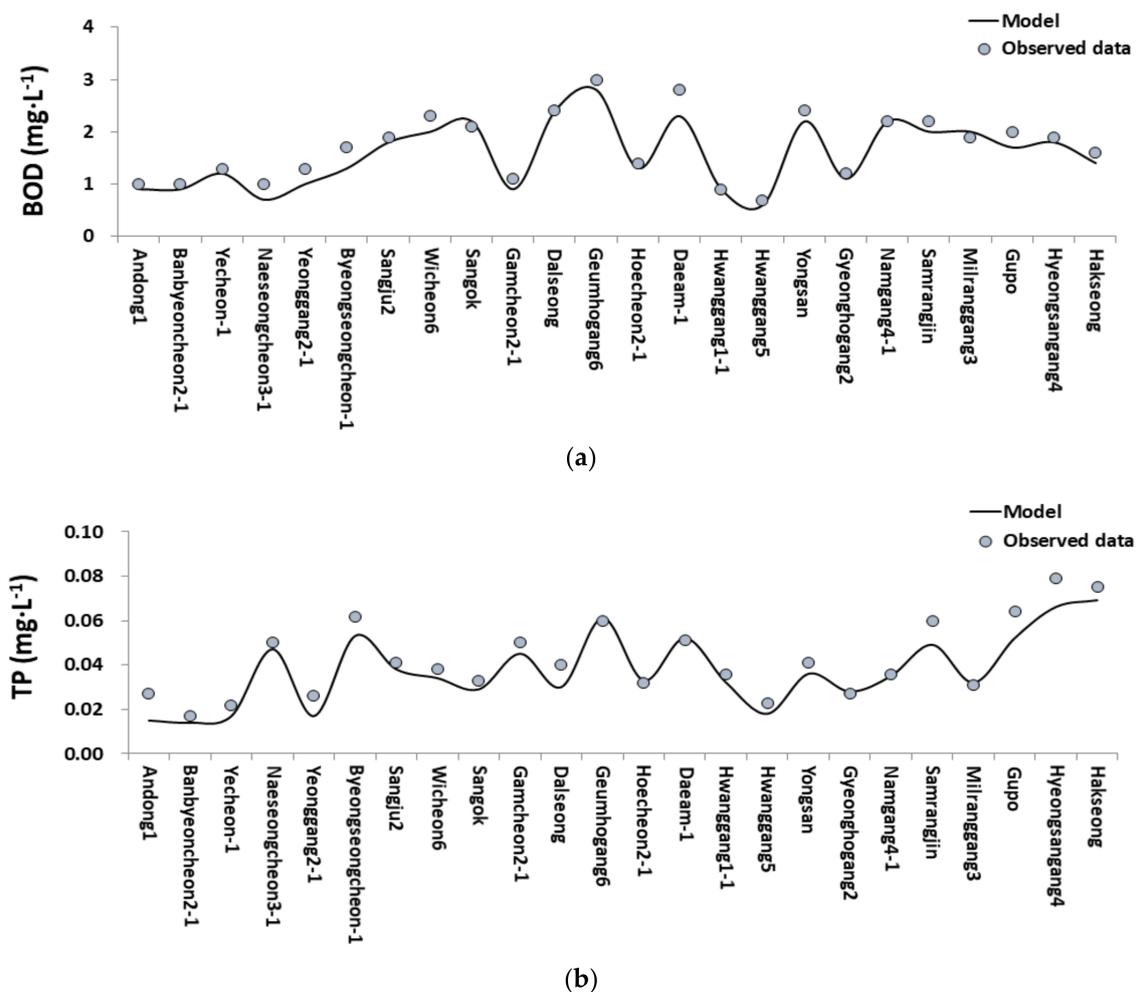


Figure 4. Observed and modeled (a) BOD and (b) TP concentrations during 2018 across the 24 analysis sites across the Nakdong River Basin.

3.4. Assessment of the Attainment of the 2018 Water Quality Target

When the attainment of the 2018 water quality targets at the 24 sites was assessed using the multivariate log-linear models that considered changes in flow within the 2013–2018 period (Table 6), it was observed that the water quality targets with respect to BOD and TP concentrations were attained at 75.8% and 70.8% of the sites, respectively. These values are higher than those obtained using the annual average observed concentration method (58.3% and 50.0% for BOD and TP, respectively) [35]. Regarding BOD concentration, the nonattainment of water quality targets changed to attainment for the large tributaries, which are affected to a greater extent by flow, compared with the main stream sites, and for TP concentration, the attainment rate was improved for both the main stream and tributary sites. This appears to be due to the organic pollutant dilution effect resulting from an increase in flow and the influence of continuous efforts to improve water quality, such as the reinforcement of water quality standards for the water discharged from public sewage treatment facilities and the expansion of TP treatment capacity of sewage and wastewater treatment facilities [32,36].

Table 6. Comparison of observed and estimated BOD and TP attainment values for water quality targets across the Nakdong River Basin, 2018.

Site	BOD (mg·L ⁻¹)			TP (mg·L ⁻¹)		
	Water Quality Target	Attainment		Water Quality Target	Attainment	
		Observed	Estimated		Observed	Estimated
Andong1	≤1	○	○	≤0.02	X	○
Banbyeoncheon2-1	≤2	○	○	≤0.04	○	○
Yecheon-1	≤1	X	X	≤0.02	X	○
Naeseongcheon3-1	≤1	○	○	≤0.02	X	X
Yeonggang2-1	≤1	X	○	≤0.02	X	○
Byeongseongcheon-1	≤2	○	○	≤0.04	X	X
Sangju2	≤1	X	X	≤0.02	X	X
Wicheon6	≤2	X	○	≤0.04	○	○
Sangok	≤1	X	X	≤0.02	X	X
Gamcheon2-1	≤1	X	○	≤0.02	X	X
Dalseong	≤2	X	X	≤0.04	○	○
Geumhogang6	≤3	○	○	≤0.1	○	○
Hoecheon2-1	≤2	○	○	≤0.04	○	○
Daeam-1	≤3	○	○	≤0.1	○	○
Hwanggang1-1	≤2	○	○	≤0.04	○	○
Hwanggang5	≤1	○	○	≤0.02	X	○
Yongsan	≤2	X	X	≤0.04	X	○
Gyeonghogang2	≤2	○	○	≤0.04	○	○
Namgang4-1	≤2	X	X	≤0.04	○	○
Samrangjin	≤2	X	○	≤0.04	X	X
Milranggang3	≤2	○	○	≤0.04	○	○
Gupo	≤2	○	○	≤0.04	X	X
Hyeongsangang4	≤3	○	○	≤0.1	○	○
Hakseong	≤3	○	○	≤0.1	○	○

4. Conclusions

In this study, among 32 representative sites in the sub-basin of the Nakdong River, multivariate log-linear models were applied to 24 sites where flow data were available to analyze the water quality trend and assess the water quality targets. The significance and suitability of the models were evaluated using the F-test, RMSE, MAPE, and adjusted R^2 values. Thus, it was observed that all the 24 models showed statistical significance. Specifically, the explanatory power for TP concentration was higher than that for BOD concentration, and the explanatory power corresponding to tributaries was higher than that corresponding to the main stream.

Further, based on the multivariate linear models, flow was identified as the main factor that affects water quality with respect to BOD and TP. In particular, flow had a dominant effect on TP and BOD concentrations at main stream and tributary sites, respectively. This observation is highly related to organic pollutant concentration dilution resulting from an increase in flow and the increase in water quality owing to the inflow of nutrients as a result of the discharge of nonpoint pollutants during rainfall. For tributaries with low stream flow, it was observed that the change in water quality was sensitive to the inflow of pollutants. Considering each water quality parameter, BOD concentration tended to decrease, while TP concentration tended to increase as flow increased. However, both BOD and TP concentrations showed a tendency to improve over time. This is considered to be the effect of investments related to environmental improvement, such as the expansion of treatment facilities for biodegradable organic matter and TP treatment over the last 30 years.

Furthermore, based on the RMSE and MAPE values obtained, the multivariate log-linear models showed suitability in the prediction of water quality parameters. The evaluation of the 2018 water quality target attainment rate on this basis showed that the BOD concentration target was attained to a greater extent than the TP concentration target. This could be attributed to the positive effects resulting in significant improvement in BOD concentration in severely polluted main rivers owing to the dilution effect brought about by the increase in flow as well as the implementation of the drastic water quality management policies in the 1980s and 1990s.

Based on their effectiveness in accurately explaining the water quality parameters, the multivariate regression models applied in this study were determined to be more suitable for the prediction of nutrient concentrations instead of organic matter concentrations and the prediction of the water quality of tributaries rather than that of the main stream. This water quality assessment method allowed the identification of water quality influencing factors by linking water quality, flow, and time (seasonality). It is expected that the results derived in this study will be used as basic data for the preparation of water quality-flow integrated management plans as well as objective assessment methods for the identification of the effects of water quality management policies on water quality. In the future, it will be necessary to conduct further research with higher accuracy by expanding the number of survey sites for both water quality and flow and by securing long-term observation data.

Author Contributions: Conceptualization, M.-A.K., S.L., E.N. and S.P.; methodology, E.N. and S.P.; formal analysis, M.-A.K. and S.L.; data curation, M.-A.K. and S.L.; writing—original draft preparation, M.-A.K.; writing—review and editing, M.-A.K. and S.L.; supervision, S.L.; project administration, T.K., S.L. and E.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute of Environmental Research (NIER), grant number NIER-2020-04-02-129.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research was funded by the National Institute of Environmental Research (NIER), under the Ministry of Environment (MOE) of the Republic of Korea, grant number NIER-2020-04-02-129.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ministry of Environment (MOE). *Republic of Korea, The Evaluation Regulations of Water Quality Target*; Ministry of Environment (MOE): Sejong, Korea, 2019; p. 1.
2. Johnson, N.M.; Likens, G.E.; Bormann, F.H.; Fisher, D.W.; Pierce, R.S. A working model for the variation in stream water chemistry at the Hubbard Brook Experimental Forest, New Hampshire. *Water Resour. Res.* **1969**, *5*, 1353–1363. [CrossRef]
3. Hirsch, R.M.; Slack, J.R.; Smith, R.A. Techniques of trend analysis for monthly water quality data. *Water Resour. Res.* **1982**, *18*, 107–121. [CrossRef]
4. Smith, R.A.; Hirsch, R.M.; Slack, J.R. *A Study of Trends in Total Phosphorous Measurements at Stations in the NASQAN Network*; Technical Report Water Supply Paper 2190; U.S. Geological Survey: Reston, VA, USA, 1982; pp. 1–8.
5. Cohn, T.A.; Caulder, D.L.; Gilroy, E.J.; Zynjuk, L.D.; Summers, R.M. The validity of a simple log-linear model for estimation fluvial constituent loads—An empirical study involving nutrient loads entering Chesapeake Bay. *Water Resour. Res.* **1992**, *28*, 2353–2364. [CrossRef]
6. Na, E.H.; Park, S.S. Continuity simulation and trend analysis of water qualities in incoming flows to Lake Paldang by log linear models. *Korean J. Limnol.* **2003**, *36*, 336–343.
7. Shin, M.H.; Seo, J.Y.; Choi, Y.H.; Kim, J.; Shin, D.; Lee, Y.J.; Jung, M.S.; Lim, K.J.; Choi, J. Evaluation of LOADEST model applicability for NPS pollutant loads estimation from agricultural watershed. *J. Korean Soc. Water Qual.* **2009**, *25*, 212–220.
8. Park, Y.S.; Lee, J.M.; Shin, Y.; Shin, M.H.; Park, J.H.; Hwang, H.; Ryu, J.; Park, J. Evaluation of regression models in LOADEST to estimate suspended solid load in Hangang Waterbody. *J. Korean Soc. Agric. Eng.* **2015**, *57*, 37–45.
9. Kim, K.; Kang, M.S.; Song, J.H.; Park, J. Estimation of LOADEST coefficients according to watershed characteristics. *J. Korean Water Resour.* **2018**, *51*, 151–163.
10. Lee, J.H.; Lee, S.H.; Lee, S.H.; Lee, J.K. Evaluation of attainment ratio on water quality goal of the mid-watershed representative station. *J. Korean Soc. Water Environ.* **2017**, *33*, 525–530.
11. U.S. Environmental Protection Agency (EPA). *An Approach for Using Load Duration Curves in the Development of TMDLs*; U.S. Environmental Protection Agency: Washington, DC, USA, 2007.
12. Hwang, H.S.; Park, B.K.; Kim, Y.S.; Park, K.J.; Cheon, S.U.; Lee, S.J. Research on the applicability of the load duration curve to evaluate the achievement of target water quality in the unit watershed for a TMDL. *J. Korean Soc. Water Environ.* **2011**, *27*, 885–895.
13. Park, J.H.; Kim, K.S.; Hwang, K.S.; Lee, Y.W.; Lim, B.G. Application of load duration curve and estimation of delivery ratio by flow durations using discharge load rating curve at Jiseok Stream Watershed. *J. Korean Soc. Water Environ.* **2013**, *29*, 523–530.
14. Kim, E.K.; Ryu, J.C.; Kim, H.T.; Kim, Y.S.; Shin, D.S. Application of the load duration curve (LDC) to evaluate the achievement rate of target water quality in the Han-River Watersheds. *J. Korean Soc. Water Environ.* **2015**, *31*, 732–738. [CrossRef]
15. Ha, D.W.; Kim, S.G.; Song, C.S.; Ko, G.Y. Effect analysis on target water quality change by measurement data in Yochon a point using load duration curve. *J. Korean Soc. Environ. Technol.* **2016**, *17*, 105–113.
16. Hesel, D.R.; Hirsch, R.M. *Statistical Methods in Water Resources: U.S. Geological Survey, Techniques of Water-Resources Investigations, 04-A3*; U.S. Geological Survey: Reston, VA, USA, 2002; p. 522.
17. Lystrom, D.J.; Rinella, F.A.; Rickert, D., A.; Zimmermann, L. *Multiple Regression Modeling Approach for Regional Water Quality Management*; Environmental Research Laboratory, EPA-600/778-198; U.S. Environmental Protection Agency, U.S. Department of the Interior: Portland, OR, USA, 1978; p. 60.
18. Lizárraga, J.S. *Nutrient and Sediment Concentrations, Trends, and Loads from Five Subwatersheds in the Patuxent River Basin, Maryland, 1986–1996*; U.S. Geological Survey Water-Resources Investigations Report 98-4221; U.S. Geological Survey: Reston, VA, USA, 1999; p. 31.
19. Runkel, R.L.; Crawford, C.G.; Cohn, T.A. *Load Estimator (LOADEST)—A Fortran Program for Estimating Constituent Loads in Streams and Rivers*; U.S. Geological Survey Techniques and Methods, 4-A5; U.S. Geological Survey: Reston, VA, USA, 2004; p. 69.
20. Hirsch, R.M. Large biases in regression-based constituent flux estimates: Causes and diagnostic tools. *J. Am. Water Resour. Assoc.* **2014**, *50*, 1401–1424. [CrossRef]
21. Lathrop, T.R.; Bunch, A.R.; Downhour, M.S. *Regression Models for Estimating Sediment and Nutrient Concentrations and Loads at the Kankakee River, Shelby, Indiana, December 2015 through May 2018*; U.S. Geological Survey Scientific Investigation Report 2019-5087; U.S. Geological Survey: Reston, VA, USA, 2019; p. 13.
22. Choi, J.Y.; Han, D.H. Development of Water Quality Standard for TOC as Organic Matter Index. *Seoul Stud.* **2001**, *12*, 173–184.
23. Nakdong River Environment Research Center (NIER). *Changes of Water Environment and Phytoplankton Community Structures in the Nakdong River*; 11-1480523-001712-01; National Institute of Environmental Research: Incheon, Korea, 2013; pp. 3–5.
24. Park, J.S.; Wang, D.; Lee, W.H. Evaluation of weir construction on water quality related to algal blooms in the Nakdong River. *Environ. Earth Sci.* **2018**, *77*, 408. [CrossRef]
25. Korea Meteorological Administration (KMA). Available online: <http://data.kma.go.kr> (accessed on 17 December 2019).

26. National Institute of Environmental Research (NIER). *Report of the National Pollution Source Survey*; Based on the 2017, 11-1480523-000429-10; National Institute of Environmental Research: Incheon, Korea, 2019; p. 143.
27. Water Environment Information System (WEIS). Available online: <http://water.nier.go.kr> (accessed on 15 December 2019).
28. Water Resources Management Information System (WAMIS). Available online: <http://www.wamis.go.kr> (accessed on 15 December 2019).
29. My Water System (K-Water). Available online: <http://www.water.or.kr> (accessed on 17 December 2019).
30. Rim, C.S. Comparison of incoming solar radiation equations for evaporation estimation. *CNU J. Agric. Sci.* **2011**, *38*, 129–143.
31. Park, H.S.; Park, S.J.; Lim, J.G. Forecasting advertising communication effects using TV advertising exposure data. *Broadcast. Commun.* **2018**, *19*, 165–197.
32. Lewis, C.D. *Industrial and Business Forecasting Methods*; Butterworth Publishing: London, UK, 1982; p. 40.
33. Ministry of Environment (MOE). *The 2nd Basic Plan for Water Environment Management*; 11-1480000-001483-14; Ministry of Environment (MOE): Sejong, Korea, 2018; p. 28.
34. Ahn, J.H.; Han, D.H.; Cho, J.H.; Kim, S.B.; Lee, S.M. Assessing the Feasibility of Strengthening Point-Source Nitrogen Management Policy. Korea Environment Institute: Sejong, Korea, 2018; p. 52.
35. National Institute of Environmental Research (NIER). *Report of the National Water Quality Assessment*; 11-1480523-002395-10; National Institute of Environmental Research: Incheon, Korea, 2018; p. 29.
36. Lee, B.K.; Kim, H.J.; Hyun, Y.J.; Jung, S.H.; Kim, E.Y. *A Study for Integrated Management of Water Quantity and Quality to Improve the Water Environment of 4 Rivers Restoration Project*; Research Report 2012-04; Korean Environment Institute (KEI): Sejong, Korea, 2012; pp. 525–530.