

Article

Assessing Land-Cover Effects on Stream Water Quality in Metropolitan Areas Using the Water Quality Index

TaeHo Kim , YoungWoo Kim, Jihoon Shin , ByeongGeon Go and YoonKyung Cha * 

School of Environment Engineering, University of Seoul, 163, Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Korea; willy1995@uos.ac.kr (T.K.); youngwoo0508@uos.ac.kr (Y.K.); sjh3473@uos.ac.kr (J.S.); rhqudrjs7@uos.ac.kr (B.G.)

* Correspondence: ykcha@uos.ac.kr

Received: 7 October 2020; Accepted: 21 November 2020; Published: 23 November 2020



Abstract: This study evaluated the influence of different land-cover types on the overall water quality of streams in urban areas. To ensure national applicability of the results, this study encompassed ten major metropolitan areas in South Korea. Using cluster analysis, watersheds were classified into three land-cover types: Urban-dominated (URB), agriculture-dominated (AGR), and forest-dominated (FOR). For each land-cover type, factor analysis (FA) was used to ensure simple and feasible parameter selection for developing the minimum water quality index (WQI_{min}). The chemical oxygen demand, fecal coliform (total coliform for FOR), and total nitrogen (nitrate-nitrogen for URB) were selected as key parameters for all land-cover types. Our results suggest that WQI_{min} can minimize bias in water quality assessment by reducing redundancy among correlated parameters, resulting in better differentiation of pollution levels. Furthermore, the dominant land-cover type of watersheds, not only affects the level and causes of pollution, but also influences temporal patterns, including the long-term trends and seasonality, of stream water quality in urban areas in South Korea.

Keywords: urban stream; factor analysis; land-cover type; metropolitan area; minimum water quality index; pollution

1. Introduction

Global urbanization is an ongoing trend, with 55% to 68% of the world's population projected to reside in urban areas by 2050 [1]. Urbanization induces multiple stressors, especially land-use/land-cover changes such as deforestation and the growth of industrial and residential areas, resulting in increased impervious surfaces [2–5]. Consequently, urbanization leads to a deterioration of water quality in streams through an increase in pollution sources and various hydromorphological changes [6–8]. Despite their at-risk status, streams in urban areas are crucial water resources with a number of designated uses, such as drinking water supply, recreation, and wildlife conservation [9–12].

Therefore, it is vital to establish management strategies for preventing or alleviating water quality problems; this requires efforts to monitor and assess stream water quality in urban areas. The water quality index (WQI), an approach that quantitatively integrates a number of chemical, physical, and biological water quality parameters, has been widely used to assess the water quality status of both surface and groundwater systems [13–17]. In recent years the advent of big data and the accumulation of monitored multivariate data has prompted a substantial increase in the application of WQI to environmental and ecological studies [18–20]. In many of these studies, the developed WQI has been used to capture long-term trends [21,22], seasonal fluctuations [23,24], or spatial variations [25,26] in

the overall stream water quality in urban areas. As well as determining the spatiotemporal patterns of stream water quality in urban areas, previous WQI-based research has also determined pollution sources and anthropogenic effects [27–29] and selected the key parameters that represent variations in water quality [30–33].

Recent assessments of urban stream water quality have increasingly employed parameter selection using a number of statistical methods, highlighting the advantages of this process for cost and time saving during assessment. For example, Wu et al. [33] used stepwise multiple regression, which assumes linearity between the WQI and each parameter, to select five parameters representing the water quality of streams in the highly developed area of Lake Taihu Basin, China. Tripathi and Singal [31] used both principal component analysis (PCA) and correlation analysis to select nine parameters to develop a WQI for the Ganga River, which flows through some highly polluted cities of India. Moreover, linear discriminant analysis was applied by Han et al. [34] to select parameters that most effectively differentiate temporal groups (wet versus dry period) and spatial groups (east vs. west parts of the lake) in the Fu River and Baiyangdian Lake, both of which are located in a highly populated region of northern China.

However, the spatial scales of previous parameter selection studies have been limited to single water bodies or single basins; thus, the parameters selected in these studies have limited applicability to other urban stream ecosystems. Furthermore, the effects of different types of anthropogenic activities (e.g., industry, cultivation, or forestation), on stream water quality in urban areas has rarely been considered [26,35]. To overcome these limitations, this study presents the first attempt, to our knowledge, to explicitly account for the effects of different land-cover types on the water quality response and key water quality parameters of urban streams. This study was conducted on a national scale, encompassing a wide range of hydromorphological and geographical characteristics and socioeconomic backgrounds, which are also key factors influencing water quality [36–40]. Therefore, this study aimed to provide parameter selection results that are both informative and applicable to other unexplored streams in urban areas of South Korea.

Streams across ten major metropolitan areas of South Korea were investigated. Cluster analysis was performed to classify stream watersheds based on their land-cover characteristics. Then, the objective WQI (WQI_{obj}) was calculated for each land-cover type using all available water quality parameters. The long-term trends of WQI_{obj} were evaluated using the seasonal Mann-Kendall (SMK) test, and only periods exhibiting temporal stability were used in further analyses. For each land-cover type, key parameters were selected using factor analysis (FA) to develop the minimum WQI (WQI_{min}). The objectives of this study were: (1) To assess the long-term trends and seasonality of the overall stream water quality in metropolitan areas in South Korea; (2) to analyze how different land-cover types affect stream water quality in urban areas and key water quality parameters; and (3) to evaluate the correlation between WQI_{obj} and WQI_{min} and relationships between WQI_{min} and land-covers.

2. Materials and Methods

2.1. Study Area and Data Description

Ten major metropolitan areas across South Korea, with populations of greater than one million, were included in this study (Seoul, Busan, Incheon, Daegu, Daejeon, Gwangju, Suwon, Ulsan, Changwon, and Goyang (Figure 1)) [41]. Within the study area, 81 water quality monitoring sites were selected at tributaries that directly or indirectly flow into either the Han, Geum, Nakdong, or Yeongsan Rivers, the four major rivers of South Korea. The selected monitoring sites covered 35 standard watersheds with the range of watershed area from 39 to 294.9 km², and a mean area of 103.29 km², the smallest unit of the drainage area division system in South Korea (<http://wamis.go.kr>). Water quality data were provided by the National Institute of Environmental Research of the Ministry of Environment (<http://water.nier.go.kr>). The data spanned the time period from 2007 to 2018, and the monitoring frequency varied by site from weekly to monthly. Among the 54 water quality parameters

initially included in the data, heavy metals and other toxic chemicals, such as mercury, cadmium, arsenic, and cyanide, were not included because at least 99.5% of the values for these parameters were either missing or below the detection limit. Furthermore, parameters without available reference values were not included in the analyses. The reference values (i.e., normalization factors and weights) required to develop the Bascaron WQI were provided by previous studies [27,42–44].

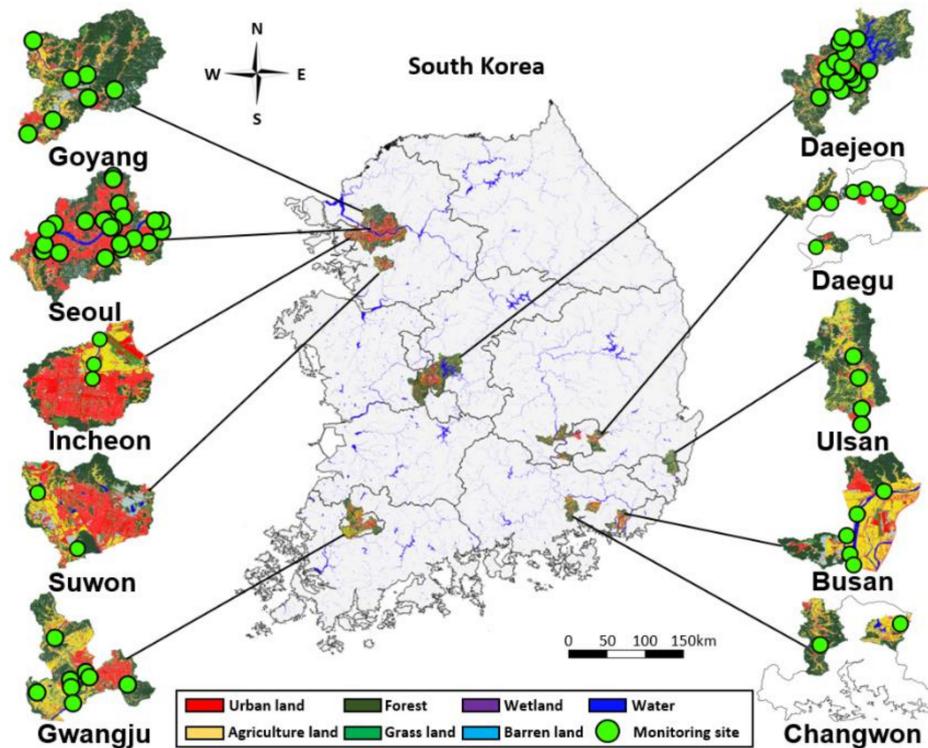


Figure 1. Location of monitoring sites in ten major metropolitan areas of South Korea.

Fourteen water quality parameters were included in the analyses: Water surface temperature (Temp), electrical conductivity (EC), pH, dissolved oxygen (DO), five-day biochemical oxygen demand (BOD₅), chemical oxygen demand (COD), suspended solids (SS), total nitrogen (TN), ammonium nitrogen (NH₄⁺-N), nitrate nitrogen (NO₃⁻-N), total phosphorus (TP), orthophosphate phosphorus (PO₄³⁻-P), total coliform (TC), and fecal coliform (FC) (Table 1). Among the 81 monitoring sites initially selected for our study, 58 were included for the water quality assessment as they had measurements for all 14 water quality parameters. Land-cover data were provided by the Environmental Geographic Information System; the year of data collection varied from 2010 to 2018 depending on the region (<https://egis.me.go.kr>). The land-cover data involved seven categories: urban (or built-up) land, agricultural land, forested land, grassland, wetland, barren land, and water. For each of the 35 watersheds, the relative proportions of the seven land-cover categories were calculated using QGIS 2.18.16 [45] and ArcGIS 10.3 software [46].

Table 1. Reference values (i.e., normalization factors, C_i and weights, P_i) required to develop the Bascarón WQI and provided by previous studies [27,42–44].

Parameter	Unit	Relative Weight (P_i)	Normalization Factor (C_i)										
			100	90	80	70	60	50	40	30	20	10	0
Temp	°C	1	21/16	22/15	24/14	26/12	28/10	30/5	32/0	36/−2	40/−4	45/−6	>45/<−6
pH	-	1	7	7–8	7–8.5	7–9	6.5–7	6–9.5	5–10	4–11	3–12	2–13	1–14
EC	μS/cm	1	<750	<1000	<1250	<1500	<2000	<2500	<3000	<5000	<8000	≤12,000	>12,000
DO	mg/L	4	≥7.5	>7	>6.5	>6	>5	>4	>3.5	>3	>2	≤1	<1
BOD ₅	mg/L	3	<0.5	<2	<3	<4	<5	<6	<8	<10	<12	≤15	>15
COD	mg/L	3	<5	<10	<20	<30	<40	<50	<60	<80	<100	≤150	>150
SS	mg/L	4	<20	<40	<60	<80	<100	<120	<160	<240	<320	≤400	>400
TN	mg/L	2	<0.8	<3.8	<7.5	<13	<18	<27	<48	<85	<149	≤265	>265
NH ₄ ⁺ -N	mg/L	3	<0.01	<0.05	<0.1	<0.2	<0.3	<0.4	<0.5	<0.75	<1	≤1.25	>1.25
NO ₃ ⁻ -N	mg/L	2	<0.5	<2	<4	<6	<8	<10	<15	<20	<50	≤100	>100
TP	mg/L	1	<0.2	<1.6	<3.2	<6.4	<9.6	<16	<32	<64	<96	≤160	>160
PO ₄ ³⁻ -P	mg/L	1	<0.025	<0.05	<0.1	<0.2	<0.3	<0.5	<0.75	<1	<1.5	≤2	>2
TC	CFU/100 mL	3	<50	<500	<1000	<2000	<3000	<4000	<5000	<7000	<10,000	≤14,000	>14,000
FC	CFU/100 mL	3	<5	<50	<100	<200	<300	<400	<500	<700	<1000	≤1400	>1400

2.2. Statistical Analyses

2.2.1. Cluster Analysis (CA)

CA is an unsupervised pattern recognition technique, whereby individual objects are grouped into a number of clusters whose objects are more similar than those in other clusters. Among the available CA methods, hierarchical agglomerative CA (HACA) was used in this study. HACA is a successive process, in which two objects in the closest proximity form a cluster at the lowest hierarchy. In the next step, the newly generated two clusters in the closest proximity form a combined cluster. Merging continues until all objects are linked to form a single cluster at the highest hierarchy. The squared Euclidean distance was used as a measure for calculating the proximity between objects/clusters.

Furthermore, we employed the Ward's minimum variance linkage function, which uses distance information to merge objects into a hierarchical cluster tree and is visually represented by a dendrogram [47]. As HACA results in a single cluster, the dendrogram needs to be divided at a specific height to generate multiple clusters. The height in the dendrogram can be defined as $(D_{\text{link}}/D_{\text{max}}) \cdot 100$, where D_{link} is the linkage distance for a pair of objects/clusters and D_{max} is the maximum linkage distance. According to previous studies, the height for dendrogram partitioning was set to 60; that is, $(D_{\text{link}}/D_{\text{max}}) \cdot 100 > 60$ [48,49]. The CA was performed using 'dendrogram' function from the 'SciPy' library [50] in Python 3.6 [51]. To generate clusters based on land-cover type, HACA was performed using the relative proportions of the six land-cover types for each standard watershed (excluding water) as variables.

The differences in water quality parameters and WQI among different clusters were assessed using summary statistics and non-parametric tests, i.e., Kruskal Wallis H and Mann-Whitney U tests. Non-parametric tests were selected due to the non-normality of water quality parameters. The Kruskal Wallis H test examined the differences in distributions for the three clusters. When the significant differences occurred, as a post-hoc analysis, the Mann-Whitney U test was used to identify which cluster(s) revealed the significant difference in distribution from the other cluster(s). The Kruskal Wallis H and Mann-Whitney U tests were performed using 'kruskal' and 'mannwhitneyu' from the 'SciPy' library [50] Python 3.6 [51]. Statistical significance was indicated by p -value < 0.05 .

2.2.2. Water Quality Index (WQI) Development

The method for WQI development used in this study builds on the WQI_{obj} [43], a modification of the *Bascarón* WQI, also known as subjective WQI [13], which excludes the constant term multiplied to WQI_{obj} , which reflects the subjective judgment of overall water quality. The WQI_{obj} is calculated as follows,

$$WQI_{\text{obj}} = \frac{\sum_{i=1}^n C_i P_i}{\sum_{i=1}^n P_i} \quad (1)$$

where n is the number of available water quality parameters, C_i is a normalization factor that converts the value of a parameter into a common scale ranging from 0 to 100 with an interval of 10 (Table 1), P_i is the weight indicating the relative importance of parameters, which ranges from 1 to 4 (Table 1), and WQI_{min} is a simplification of WQI_{obj} indicating the minimum WQI [42,43] and is calculated as,

$$WQI_{\text{min}} = \frac{\sum_{i=1}^{n_{\text{min}}} C_i}{n_{\text{min}}} \quad (2)$$

Note that Equation (2) for WQI_{min} does not include the weight term, indicating that the parameters included in WQI_{min} assessment are considered equally important. Here, n_{min} is the number of key parameters, which is a subset of all n available parameters. The WQI_{obj} and WQI_{min} scores were graded into five classes to indicate the overall water quality status: excellent (91–100), good (71–90), medium (51–70), bad (26–50), and very bad (0–25) [42,43,52]. Also, when comparing WQI_{min} with

WQI_{obj} for evaluating whether they are well-correlated, linear regression ($WQI_{obj} = a \cdot WQI_{min} + b$) was performed using 'linregress' function from the 'SciPy' library [50] in Python 3.6 [51].

2.2.3. Seasonal Mann-Kendall (SMK) Test

The Mann-Kendall (MK) test is a non-parametric test that assesses if the temporal trend of a variable exhibits a monotonic increase or decrease [53,54]. The SMK test is an extension of the MK test that accounts for the effect of seasonality by performing the test separately for each pre-defined season [55]. In this study, the SMK was employed to identify the point in time at which WQI no longer shows a significant increasing or decreasing trend; this stabilized time period was divided into training and test sets, and further analyses were performed. CA results were used to assess the trend of monthly averaged WQI_{obj} for each land-cover cluster, initially using the data for the entire time period (2007–2018). When the trend showed a significant increase or decrease (p -values < 0.05), the SMK was performed excluding a 1-year period of data from the starting year. The test was iteratively performed until the trend appeared to be insignificant. The SMK test was performed using 'seasonal_test' function from the 'pyMannKendall' library [56] in Python 3.6 [51].

2.2.4. Factor Analysis (FA)

FA attempts to account for the structure (i.e., correlation and variation) of data, consisting of measured variables with a reduced number of factors, which are also termed latent variables. Exploratory FA (EFA), which does not assume a priori relationships among the factors and measured variables, was used to reveal the underlying factors behind the correlations among measured water quality parameters. Contrary to confirmatory FA, EFA does not posit any relationship between specific factors and measured variables. Therefore, it suits the purpose of this analysis.

Prior to analysis, the values of all water quality parameters except Temp and pH were log-transformed, and the values of all parameters were standardized to have a distribution with a mean of zero and standard deviation of one. To examine whether water quality data were suitable for FA, the Kaiser-Meyer-Olkin (KMO) test [57] and Bartlett's test [58] were performed. The FA was assumed to be valid when the KMO value exceeded 0.5 and the Bartlett's test result was significant (p -value < 0.05).

To determine the number of factors retained in the FA, Horn's parallel analysis (PA) was used [59]. PA compares the eigenvalues (which indicate the relative importance of a factor in explaining the variance of measured variables) from measured data with the eigenvalues from random data, which have the same sample size and number of variables as the measured data and are obtained using a Monte-Carlo simulation. The differences between the eigenvalues from measured data and the mean eigenvalues from random data were calculated. Factors with differences greater than zero were retained in the FA.

As a method for factor extraction, principal component analysis (PCA) was used [60,61]. The maximum likelihood method, another common method for factor extraction, was not selected because of its multivariate normality requirement, which is often not met for water quality parameters even after the transformation (e.g., log-transformation) of values. Squared factor loading, which reflects the proportion of variance in a measured variable explained by each factor, was calculated as a result of PCA implementation. The communality was calculated by summing the squared factor loadings of a given variable across all factors to indicate the proportion of variance in a measured variable explained by all factors. Moreover, the uniqueness was calculated by subtracting the communality from the total variance of a variable.

Factor rotation (i.e., the change in the axes of factors) was implemented to yield interpretable factors by attaining a simple structure for factor loadings. Without rotation, most variables load heavily onto the first and early factors, whereas rotation yields a simple structure in which each variable loads heavily onto only one factor, while loading lightly onto the other factors. Varimax rotation was used as a rotation method, which is a common type of orthogonal rotation. Orthogonal rotation assumes

that factors remain uncorrelated with one another. FA was performed using ‘principal’ function from ‘psych’ [62] packages in R 3.5.3 [63].

For each land-cover type, the water quality parameter that showed the highest loading factor associated with each retained factor was interpreted as a key water quality parameter. Accordingly, the number of retained factors corresponded to the number of key parameters representing the overall stream water quality in urban areas. The FA procedure was performed for each land-cover type determined by CA, and the key parameters for different land-cover types were used for the WQI_{\min} calculation.

3. Results

3.1. Land-Cover Characteristics of Metropolitan Areas in South Korea

Using the HACA, three clusters were generated based on the land-cover characteristics of 35 watersheds in ten major metropolitan areas (Figure 2a). Notably, each of the watersheds included in each of the three clusters had a single dominant land-cover: Urban, agriculture, and forest, respectively (Figure 2b, Table S1). The mean proportion of urban land for the 15 watersheds with urban-dominated land-cover (URB) was $0.50 (\pm \text{one standard deviation of } 0.12)$, which was higher than that of agricultural (0.06 ± 0.05) and forested (0.30 ± 0.10) land. In contrast, the five watersheds with agriculture-dominated land-cover (AGR) had a mean relative area of $0.44 (\pm 0.08)$ for agricultural land-cover, which was more dominant than urban (0.16 ± 0.07) and forested (0.24 ± 0.07) land-cover. The 15 watersheds with forest-dominated land-cover (FOR) were mainly composed of forested land, with a mean proportion of $0.60 (\pm 0.08)$, whereas the proportion of urban (0.12 ± 0.06) and agricultural (0.16 ± 0.04) land was relatively minor.

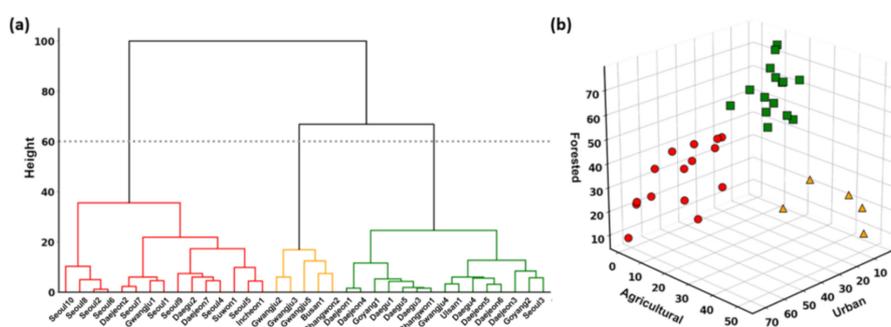


Figure 2. Clustering results of 35 watersheds, named metropolitan area with numbering, based on six land-cover categories. (a) Dendrogram exhibiting three clusters generated from hierarchical agglomerative cluster analysis. The horizontal dashed gray line represents the height for dendrogram partitioning, $(D_{\text{link}}/D_{\text{max}}) \cdot 100 > 60$. (b) Percentage (%) of the dominant land-cover type for each of the three clusters. The red circle, yellow triangle, and green square denote watersheds that are urban-dominated, agriculture-dominated, and forest-dominated, respectively.

The three land-cover types (URB, AGR, and FOR) were unevenly distributed across the metropolitan areas. Among the URB, 73.3% were concentrated in Seoul (nine watersheds) and its adjacent cities, Suwon (one watershed) and Incheon (one watershed). Three of the five AGR were located in Gwangju, whereas the other two were located in Busan and Changwon. The spatial distribution of FOR was also concentrated, with 33.3% in Daejeon and 26.7% in Daegu.

3.2. Land-Cover Effects on Stream Water Quality in Urban Areas

The long-term trends of overall water quality calculated using all available parameters (WQI_{obj}), based on the results of SMK tests, differed by land-cover type (Figure 3). For URB, WQI_{obj} values gradually improved until becoming stable in 2015 (Figure 3a). In comparison, WQI_{obj} values for AGR showed a greater improvement in early years before becoming stable in 2012 (Figure 3b). For FOR,

WQI_{obj} values did not show any significant trend during the entire period from 2007 to 2018 (Figure 3c). In more recent years (2015–2018), during which all land-cover types exhibited a stable trend, the overall water quality was worst for URB (p -values < 0.05 as a result of Kruskal Wallis H and Mann-Whitney U tests), as indicated by lower WQI_{obj} values (75.04 ± 9.90) than those for AGR (78.91 ± 8.31) and FOR (82.82 ± 7.97). Regardless of the land-cover type and time period, WQI_{obj} values tended to be lower during the wet season (July to September) than during the dry season (Figure 3).

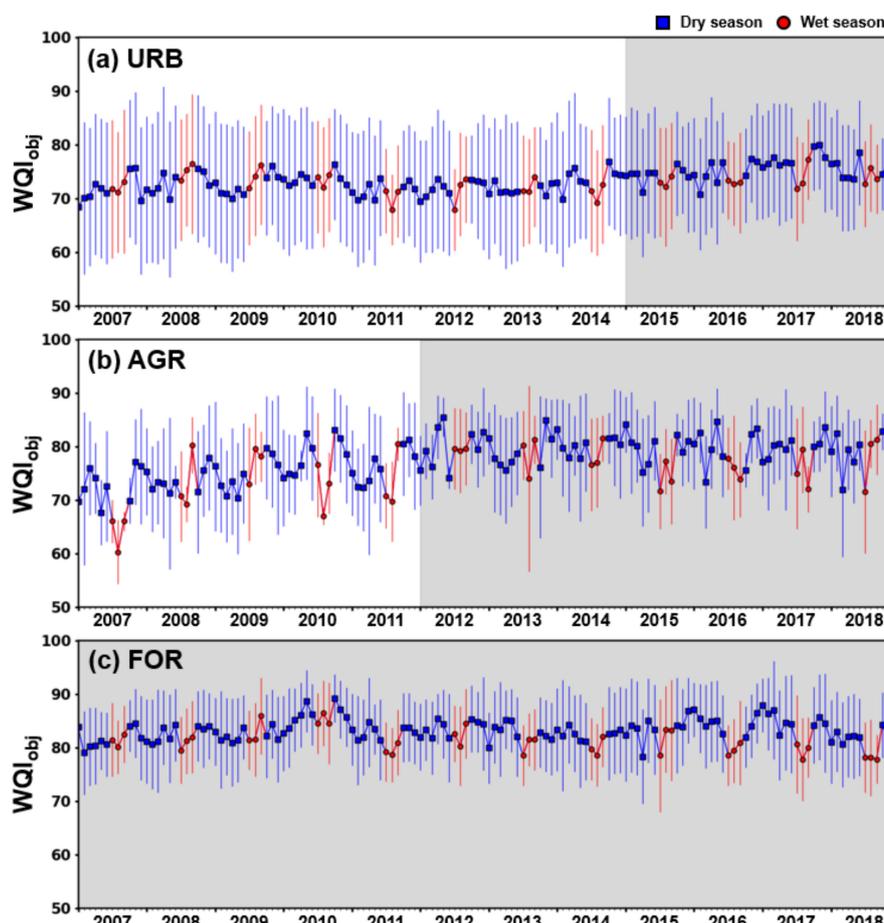


Figure 3. Long-term (2007–2018) trends of objective water quality index (WQI_{obj}) for watersheds with (a) urban-dominated, (b) agriculture-dominated, and (c) forest-dominated land-cover. Blue and red circles denote the mean monthly WQI_{obj} for dry and wet seasons, respectively, and vertical lines denote one standard deviation of monthly WQI_{obj} . The gray area represents a period exhibiting no significant increase or decrease in WQI_{obj} based on the results of seasonal Mann-Kendall tests.

The land-cover types of the watersheds influenced most water quality parameters in urban streams except for pH, EC, DO, and $PO_4^{3-}-P$, which were similar regardless of the dominant land-cover (Table 2). Compared with URB and AGR, FOR exhibited the lowest level of contamination for the majority of water quality parameters. The level of contamination between URB and AGR differed depending on the water quality parameter. In terms of nitrogen (TN and $NO_3^{-}-N$) and microbiological indicators (TC and FC), the streams in URB exhibited significantly worse conditions than those in AGR (Table 2). On the other hand, indicators for organic matter (BOD_5 and COD) and turbidity (SS) indicated significantly higher levels of water contamination in AGR than URB (Table 2).

Table 2. Summary statistics (mean \pm one standard deviation) of 14 water quality parameters from 2015 to 2018 for watersheds with urban-dominated (URB), agricultural-dominated (AGR), and forest-dominated (FOR) land-cover. Asterisks (*) denote parameters whose mean value for either URB or AGR is significantly higher (or lower in the case of DO) than the other (p -value < 0.05 based on Kruskal Wallis H and Mann-Whitney U tests).

Parameter	Unit	Watershed Type		
		URB	AGR	FOR
Temp	°C	16.33 \pm 1.56	16.99 \pm 0.69	15.56 \pm 1.52
pH	-	7.81 \pm 0.32	7.76 \pm 0.29	7.79 \pm 0.34
EC	μ S/cm	455.51 \pm 175.19	497.27 \pm 257.41	384.76 \pm 209.78
DO	mg/L	10.52 \pm 1.46	10.49 \pm 0.88	11.03 \pm 1.04
* BOD ₅	mg/L	3.05 \pm 2.37	4.00 \pm 0.49	1.69 \pm 1.01
* COD	mg/L	5.79 \pm 3.04	8.17 \pm 0.9	4.38 \pm 2.16
* SS	mg/L	7.37 \pm 5.82	17.29 \pm 3.06	6.51 \pm 5.27
* TN	mg/L	5.92 \pm 3.18	3.49 \pm 1.42	3.24 \pm 1.43
NH ₄ ⁺ -N	mg/L	0.87 \pm 1.35	0.52 \pm 0.44	0.22 \pm 0.32
* NO ₃ ⁻ -N	mg/L	3.86 \pm 1.83	2.12 \pm 0.78	2.26 \pm 0.75
TP	mg/L	0.11 \pm 0.10	0.10 \pm 0.02	0.06 \pm 0.03
PO ₄ ³⁻ -P	mg/L	0.05 \pm 0.07	0.03 \pm 0.01	0.03 \pm 0.02
* TC	CFU/100 mL	49.20 $\times 10^3 \pm 69.12 \times 10^3$	18.09 $\times 10^3 \pm 24.97 \times 10^3$	11.73 $\times 10^3 \pm 10.49 \times 10^3$
* FC	CFU/100 mL	14.01 $\times 10^3 \pm 36.73 \times 10^3$	22.14 $\times 10^2 \pm 27.98 \times 10^2$	16.44 $\times 10^2 \pm 22.51 \times 10^2$

3.3. Key Water Quality Parameters for Different Land-Cover Types

The water quality data were suitable for the application of FA, as indicated by the results of the KMO test (0.82 for URB, 0.67 for AGR, and 0.73 for FOR) and Barlett's test (p -value < 0.05 for all land-cover types). To perform the FA, the data measured during the more recent years (2015–2018), when the WQI_{obj} values stabilized for all land-cover types, were divided into training (2015–2016) and testing (2017–2018) data sets. The results of FA using the training data indicated that three factors apiece should be retained for URB, AGR, and FOR (Table S2). For each land-cover type, the water quality parameters with the highest factor loading, associated with each of the retained factors, were selected as the key parameters for the WQI_{min} calculation (Table 3). Frequently, for a given factor, more than one water quality parameter had a factor loading greater than 0.75 [64], which is indicative of a strong correlation between the factor and the parameter (Table 3). In such cases, the parameters were generally highly correlated to each other, with a Pearson's correlation coefficient ranging from 0.49 to 0.88 (Figure S1). Consequently, the three key parameters selected for URB were COD, FC, and NO₃⁻-N, in order of corresponding factors (Table 3). Three parameters were selected for AGR were FC, COD, and TN (Table 3). The three parameters selected for FOR were COD, TN, and TC (Table 3).

3.4. Comparison between WQI_{obj} and WQI_{min}

Using the test data, the relationships between monthly WQI_{min} and WQI_{obj} values were assessed; WQI_{min} and WQI_{obj} generally exhibited moderate to strong, linear relationships with R^2 values of 0.66 for URB, 0.78 for AGR, and 0.73 for FOR (Figure 4). For both WQI_{obj} and WQI_{min} , URB was generally associated with the poorest overall water quality, with mean WQI values of 75.79 and 67.20, respectively. Further, based on both WQI_{obj} and WQI_{min} , the overall water quality for AGR (mean WQI values of 78.86 and 73.39) was generally poorer than that for FOR (mean WQI values of 82.41 and 77.41). The location of intersection, where the regression line and one-to-one line cross, differed by land-cover type: 87.48 for URB, 81.98 for AGR, and 88.14 for FOR (Figure 4). Below the intersection, WQI_{obj} values tended to be higher than WQI_{min} scores, whereas the opposite was true above the intersection (Figure 4). As the proportion of values below the intersection was greatest for URB, the positive difference between the mean WQI_{obj} and WQI_{min} values for URB (8.59) was greater than that for AGR (5.47) and FOR (5.00). Within each land-cover type, the variation of WQI_{min} values, with one standard deviation of 13.61 for URB, 13.05 for AGR, and 12.09 for FOR, was greater than the variation of WQI_{obj} values, with one standard deviation of 9.27 for URB, 8.62 for AGR, and 7.97 for

FOR. Note that the degree of variation in WQI values, in descending order, was URB, AGR, and FOR for both WQI_{obj} and WQI_{min} .

Table 3. Factor loadings for 14 water quality parameters for watersheds with urban-dominated (URB), agricultural-dominated (AGR), and forest-dominated (FOR) land-cover. Asterisks (*) indicate a factor loading greater than 0.75 or the highest factor loading in the factor. Var (%) represents the explained variance of total variance for each factor.

Parameter	Watershed Type								
	URB			AGR			FOR		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Temp	0.261	0.241	−0.639	0.243	0.643	−0.578	0.220	0.473	−0.580
pH	0.205	−0.661	−0.213	−0.666	0.264	−0.434	−0.567	0.021	0.040
EC	0.553	0.108	0.568	−0.340	0.295	0.612	0.342	0.021	0.605
DO	−0.164	−0.677	0.379	−0.634	−0.383	−0.071	−0.441	−0.578	0.340
BOD ₅	* 0.867	0.148	0.062	−0.192	* 0.797	0.031	* 0.751	0.004	0.079
COD	* 0.905	0.155	0.015	−0.185	* 0.879	0.018	* 0.897	0.081	−0.027
SS	* 0.860	0.050	−0.129	0.028	* 0.791	0.073	* 0.825	0.144	−0.033
TN	0.416	0.330	0.742	0.074	0.015	* 0.930	0.108	0.127	* 0.930
NH ₄ ⁺ -N	0.449	0.503	0.353	0.280	0.279	0.685	0.657	0.144	0.236
NO ₃ ⁻ -N	−0.006	0.145	* 0.806	0.040	−0.359	* 0.819	−0.123	0.065	* 0.910
TP	0.696	0.560	0.102	0.609	0.557	−0.098	0.651	0.411	−0.030
PO ₄ ³⁻ -P	0.340	* 0.751	0.125	0.704	−0.150	−0.095	0.141	0.667	0.028
TC	0.158	* 0.810	0.048	* 0.862	−0.157	−0.033	0.062	* 0.825	0.077
FC	0.226	* 0.831	0.025	* 0.865	−0.024	0.084	0.099	* 0.825	0.091
Var (%)	27.2	25.9	16.4	25.3	23.7	21.0	26.0	18.7	18.5

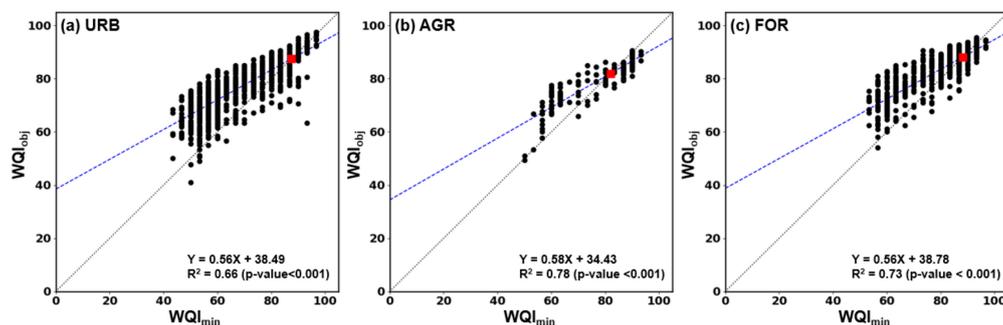


Figure 4. Relationships between objective water quality index (WQI_{obj}) and minimum WQI (WQI_{min}) for watersheds with, (a) urban-dominated, (b) agriculture-dominated, and (c) forest-dominated land use. Black circles denote WQI values calculated using the testing data set (2017–2018). Black dotted and blue dashed lines represent one-to-one, and regression lines, respectively. Red square represents the point of intersection between the one-to-one line and regression line.

3.5. Spatial Distribution of Overall Stream Water Quality in Urban Areas

WQI values by site, calculated for 2015–2018, indicated that WQI_{obj} and WQI_{min} values were highly linearly correlated, with an R^2 value of 0.84 (Figure 5b). However, there was a clear tendency for WQI_{obj} values to be higher than WQI_{min} values (Figure 5a,b). The difference in values between WQI_{obj} and WQI_{min} led to differences in WQI classification in 25.9% of the 58 monitoring sites (Figure 5a). In Seoul, the change in calculation method from WQI_{obj} to WQI_{min} yielded a change of classification from good to medium in 33.3% of 18 monitoring sites. In the other five metropolitan areas (i.e., Daejeon, Gwangju, Daegu, Busan, and Ulsan), a change in classification occurred in one or two sites, accounting for 7.7–40.0% of the sites in each area (Figure 5a). In the remaining five metropolitan areas (i.e., Goyang, Suwon, Incheon, and Changwon), no change in WQI classification occurred in response to application of the WQI_{min} (Figure 5a).

3.6. Seasonality of Overall Stream Water Quality in Urban Areas

From 2015 to 2018, the monthly patterns of overall water quality calculated using WQI_{min} differed by land-cover type (Figure 6). For URB, which exhibited the worst overall water quality, the proportion of WQI_{min} values corresponding to equal to or worse than medium status increased during the wet season (July to September), whereas the proportion of good to excellent status sites increased during the dry season (all other months) (Figure 6a). For FOR, the WQI_{min} status was consistently better than or equal to medium, and the proportion of medium status sites increased during the wet season (Figure 6c). For AGR, the WQI_{min} status tended to worsen during the wet season, with an increase in the proportion of medium status sites; however, this seasonality was less consistent compared with other land-cover types (Figure 6b).

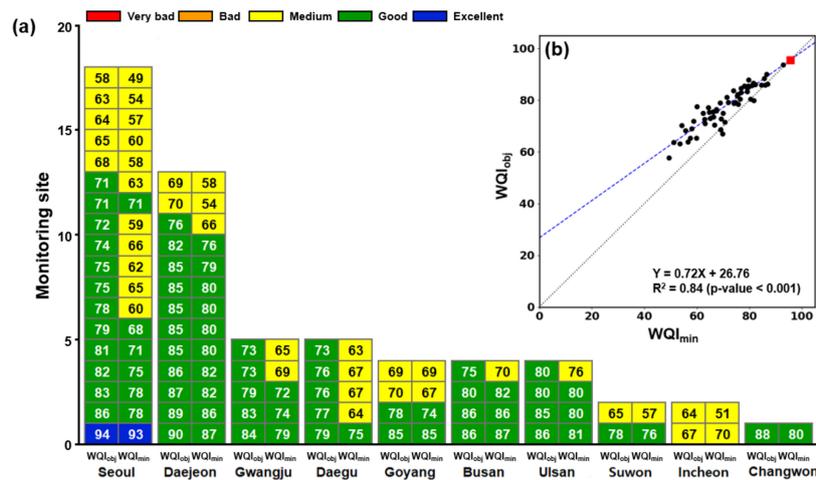


Figure 5. Spatial distribution of the water quality index (WQI) in ten major metropolitan areas of South Korea. (a) Mean objective WQI (WQI_{obj}) and minimum WQI (WQI_{min}) values and grades from 2015 to 2018 for each of the 58 monitoring sites. (b) Relationship between mean WQI_{obj} and WQI_{min} values.

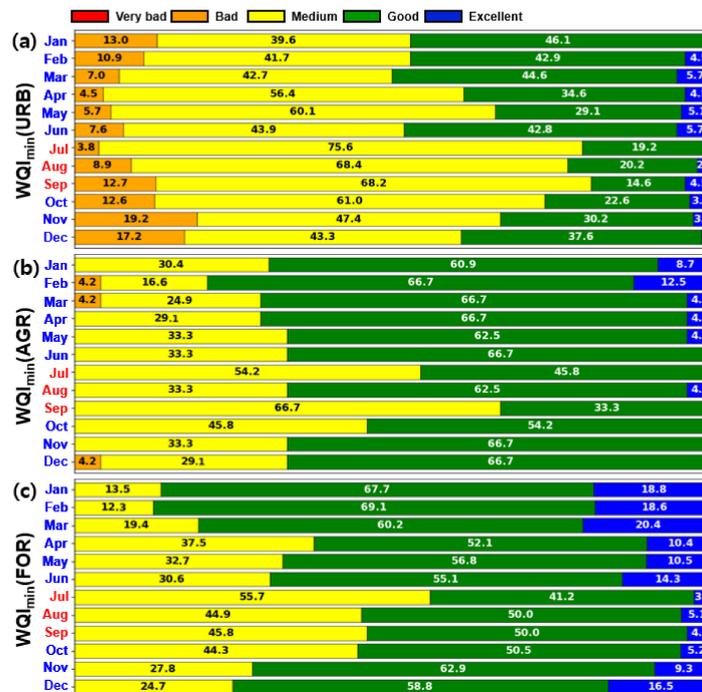


Figure 6. Monthly distribution (%) of minimum water quality index (WQI_{min}) grades from 2015 to 2018 for watersheds with, (a) urban-dominated, (b) agriculture-dominated, and (c) forest-dominated land-cover. Month names for dry and wet seasons are colored blue and red, respectively.

4. Discussion

4.1. Suitability of FA as a Parameter Selection Method

In this study, FA, which involves factor extraction and rotation processes, was used to reduce multiple intercorrelated physical, chemical, and biological water quality parameters into a smaller number of latent factors, and to select key water quality parameters that had the strongest correlation with a given latent factor. In previous studies, along with subjective judgments [27,33,65–69], multivariate statistical techniques were employed to select parameters on an objective basis. For example, stepwise multiple regression has been used [33,69,70] to determine the set of parameters that could best explain the variance of WQI_{obj} . Compared with unsupervised learning (e.g., FA), regression is a supervised method that requires reference values; in this case, WQI_{obj} values for training data. However, because of the multi-collinearity and the resulting bias, WQI_{obj} is not often a suitable reference.

Furthermore, previous studies have used PCA at the first step followed by Pearson's correlation analysis to extract water quality parameters that showed high contributions to selected components and low correlations with other parameters [31,66]. Post-hoc correlation analysis was required, since few first factors derived from PCA are strongly associated with most of the correlated parameters. Therefore, the application of PCA alone is not sufficient to attain key parameters that represent extracted factors. To address this limitation, in this study PCA was conducted in conjunction with factor rotation, which yields a simple structure for the factor loading matrix, in which only a small number of variables have high loadings onto a given factor and do not overlap among the factors. As a result, parameters with high loadings on a given factor appear to be more distinct and homogeneous. Therefore, a set of parameters with high loadings across all factors are expected to represent multifaceted aspects of water quality. Furthermore, the use of varimax rotation as a factor rotation method ensures the extracted factors are uncorrelated with one another, facilitating the selection of key parameters, the relationships among which can be assumed to be independent. Therefore, factor rotation used in conjunction with PCA does not require subsequent correlation analysis, which simplifies parameter selection to a single-step process.

4.2. Key water Quality Parameters

Selected key water quality parameters were similar among different land-cover types (COD, FC, and NO_3^- -N for URB; COD, FC, and TN for AGR; COD, TC, and TN for FOR), indicating that the relationships among parameters were consistent regardless of land-cover type. For example, across all land-cover types, COD, BOD_5 , and SS were closely correlated (Figure S1) and had high loadings with the same factor (Table 3). The high correlations were shown, since the three parameters commonly account for biodegradable organic matter. In addition, for one being the subset of the other, FC and TC, and NO_3^- -N and TN, were closely related to each other (Figure S1) and had the highest loadings onto the same factor for all land-cover types (Table 3). Note that phosphorus parameters showed moderate to strong associations with TC and FC within the same factor for all land-cover types (Table 3). Therefore, rather than phosphorus parameters, either TC or FC, which showed higher loadings with the factor than the phosphorus parameters, was selected as the key parameter. A possible speculation over this co-occurrence tendency is that phosphorus and fecal indicator bacteria may originate from the same pollution source (e.g., domestic sewage and agricultural runoff) or the same mechanism (e.g., sediment release), but future research will be necessary for interpreting the causal relationships.

The presence of multiple parameters with almost equally high loadings onto a given factor necessitated comparisons between WQI_{min} and modified WQI_{min} , in which a key parameter (e.g., COD) is replaced by its surrogate parameter (e.g., BOD_5) that was strongly related to the key parameter within the same factor. The results illustrated that modified WQI_{min} was generally in close agreement with WQI_{min} (Figure S2), suggesting that a set of parameters that shows high loadings within the same factor can be used interchangeably. Note that, compared with other sets of parameters, linear relationships

between WQI_{min} and modified WQI_{min} for fecal indicator bacteria were weaker because of the large variability inherent in FC and TC concentrations. Nonetheless, given the marginal differences in factor loading between TC and FC regardless of land-cover type (Table 3), between the two parameters, the key parameter should be selected depending on management focus or data availability.

The results of FA need to be interpreted and applied with care. The factor extraction process of FA determines the factors worth retaining, and the subsequent factor rotation, whereby the factors become least correlated with each other, yields the proportion of variance explained by a given factor to be distributed more evenly among the factors. Therefore, it is not particularly valid to prioritize the factors and the consequent key parameters. Instead, the selected key parameters should be considered independently of each other and as equally important. In this regard, assigning different weights to key water quality parameters with equal importance should not be included as a step for WQI_{min} development. Previous studies reported that using weights improved the linearity between WQI_{min} and WQI_{obj} [33,70]. In contrast to these findings, we found that the use of weights, which were estimated based on two methods, the relative weight [33,70] and the percentage of variance explained by the given factor (Table 3), yielded only slight differences in the WQI_{min} - WQI_{obj} relationships (Figure S3).

It should be acknowledged that the water quality data, used in this study, did not include several widely measured parameters, such as parameters for minerals, salts, metals and flow rate. If such parameters were added to the data, FA may include additional factors and key parameters. Moreover, the results of parameter selection did not contain the basic water quality parameters of Temp and pH in the key parameter list for any land-cover type. In addition, despite being frequently included as a key parameter [42,43,68–70] in previous studies, DO was not selected for any land-cover type in this study (Table 3). Variations in Temp, pH, and DO may be influenced by anthropogenic activities but are also attributable to natural variability. That is, they exhibit diurnal fluctuations and are strongly influenced by meteorological conditions [33,65]. Our results suggest that Temp, pH, and DO, whose patterns are substantially influenced by natural variations, may not successfully capture the total variance of stream water quality in urban areas, and may not be suitable for being included as key parameters.

4.3. Comparison between WQI_{min} and WQI_{obj}

Our results of test data showed that WQI_{min} and WQI_{obj} have close linear relationships across all land-cover types (Figure 4), suggesting that WQI_{min} can be used to predict WQI_{obj} using the established regression model. However, WQI_{min} values tended to be higher than WQI_{obj} above a certain threshold and lower than WQI_{obj} below this threshold. This tendency indicates that the use of WQI_{min} eliminates the “eclipse effect” [71], which arises from the redundancy inherent in WQI_{obj} ; accordingly, WQI_{obj} is subject to overestimating bad water quality status and underestimating good water quality status. The removal of redundancy was also evidenced by the larger variance of WQI_{min} compared with that of WQI_{obj} for all land-cover types (Figure 4). Therefore, the development and use of WQI_{min} is expected to improve the identification of the overall water quality status and the level of water pollution in streams across urban areas. Our results demonstrate that the method selection for WQI assessment has important resource and management implications. Changing the method from WQI_{obj} to WQI_{min} altered the spatial distribution of the overall water quality status; this status change occurred in a minor to substantial portion of monitoring sites, depending on the metropolitan area (Figure 5). This change suggests that the use of WQI_{min} instead of WQI_{obj} , which may involve a status change from “good” to “medium” or vice versa, may affect priority setting and resource allocation among individual watersheds or groups of watersheds.

4.4. Land-Cover Effects on Stream Water Quality in Urban Areas

Our results indicate that the dominant land-cover affected the overall stream water quality in urban areas, with mean values of both WQI_{obj} and WQI_{min} decreasing in the order: FOR > AGR > URB (Figure 4). The dominant land-cover type also contributed to the deterioration of differing water quality parameters (i.e., nitrogen and microbiological indicators for URB, but organic matter and turbidity for AGR) (Table 2). The long-term trends of overall water quality differed by land-cover type (Figure 3). Over the last decade, WQI_{obj} trends for URB and AGR exhibited early improvement before becoming stable, whereas the trend for FOR did not change significantly (Figure 3). These patterns support that, across the country, management programs implemented to control point or non-point sources for URB and AGR were effective in improving overall stream water quality [72–75]. Moreover, the implementation of conservation measures against continuing development pressures in metropolitan areas played a role maintaining the water quality in FOR. Furthermore, the land-cover type exerted an influence on the seasonality of overall water quality (Figure 6). In recent years (2015–2018), the seasonal patterns of WQI_{min} have differed for URB and FOR, whereas AGR exhibited less obvious seasonality. The less consistent seasonality for AGR may be partly attributable to the small sample size ($n = 287$, compared with n for URB = 1881 and n for FOR = 1162) corresponding to AGR. During the wet season, both URB and FOR exhibited a negative change in overall water quality with an increase in the proportion of “medium” and “good” status sites relative to “excellent” status sites (Figure 6). For URB with typically high proportions of impervious surfaces, stormwater runoff may play a significant role in decreasing overall water quality during the wet season [76–78]. Moreover, an increase in sediment discharge as well as sediment perturbation with rainfall events may facilitate the release of pollutants into surface water [79–82], resulting in a decrease in overall water quality during the wet season in both URB and FOR. In contrast, subsequent to the wet season, when dilution effects can occur [83–85], URB alone exhibited an increase in the proportion of “bad” status sites relative to “medium” and “good” status sites (Figure 6). This indicates that, not only non-point sources, but also point sources, such as wastewater treatment plant effluent, are significant forms of pollution for URB.

5. Conclusions

This study provided a statistical framework for implementing parameter selection in order to develop an objective WQI_{min} in a single-step process. Comparisons between WQI_{obj} and WQI_{min} suggested that WQI_{min} calculated with the key parameters yielded comparable results to WQI_{obj} . Furthermore, WQI_{min} reduced the eclipse effects arising from the use of correlated parameters for water quality assessment to result in a better differentiation between good and bad water quality statuses. These results have implications for management authorities, especially those motivated to launch their own monitoring network system but who have limited available resources. In this context, our results can be used to reduce monitoring demands by prioritizing the monitoring importance of a minimal number of water quality parameters. The results of WQI_{min} confirmed that the dominant land-cover type of watersheds influence multidimensional aspects of urban stream water quality; namely, the overall degree and level of pollution as well as long-term and seasonal patterns. To confirm our results, future studies should expand the number of water quality parameters exhibiting various characteristics.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4441/12/11/3294/s1>. Figure S1: Matrices of the Pearson's correlation coefficient for the period 2015–2016 among 14 water quality parameters for (a) urban-dominated (URB), (b) agricultural-dominated (AGR), and (c) forest-dominated (FOR) land-cover. Water quality parameters with high factor loadings (>0.75) on the same factor are outlined in the same color, Figure S2: Relationships between the minimum water quality index (WQI_{min}) and modified WQI_{min} from 2015 to 2018. To develop the modified WQI_{min} , key parameter values were predicted using the established linear relationship between a key parameter and a surrogate parameter. Then, predicted values were converted into normalization factors for WQI_{min} calculation. In the x-axis label, WQI_{min} (COD \rightarrow BOD₅) indicates that biochemical oxygen demand (BOD₅) was used as the surrogate for the key parameter of chemical oxygen demand (COD). Black dotted lines indicate 1:1 lines. Figure S3: Relationships between objective and minimum water quality indices (WQI_{obj} and WQI_{min}) from 2017 to 2018. Weights were determined using two methods; for a-c, a relative weight was assigned to each key parameter and for d-f, the percent variance explained by a given extracted factor was assigned to each key parameter. Black dotted lines and blue dashed lines indicate 1:1 lines and regression lines, respectively. Table S1: Proportions of three land-cover categories (urban, agricultural, and forested land) for urban-dominated watersheds (URB), agricultural-dominated watersheds (AGR), and forest-dominated watersheds (FOR). Table S2: Parallel analysis results comparing eigenvalues and simulated mean eigenvalues for urban-dominated (URB), agriculture-dominated (AGR), and forest-dominated (FOR) land-cover. The simulated mean eigenvalue indicates the mean eigenvalue calculated from randomly generated simulation data. Asterisks (*) indicate that the eigenvalue is higher than the corresponding simulated mean eigenvalue.

Author Contributions: Conceptualization, T.K. and Y.C.; data curation, T.K., Y.K., J.S. and B.G.; formal analysis, T.K.; funding acquisition, T.K., Y.K., J.S., B.G. and Y.C.; investigation, T.K., Y.K., J.S. and B.G.; methodology, T.K. and Y.C.; project administration, T.K. and Y.C.; resources, T.K., Y.K., J.S. and B.G.; software, T.K. and Y.C.; supervision, T.K. and Y.C.; validation, T.K. and Y.C.; visualization, T.K. and Y.C.; writing—original draft, T.K. and Y.C.; writing—review and editing, T.K. and Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C1009961).

Acknowledgments: The authors would like to thank the editors and reviewers for useful comments which are helpful in improving the manuscript quality.

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

References

1. United Nations Department of Economic and Social Affairs. 68% of the World Population Projected to Live in Urban Areas by 2050, Says UN. Available online: <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html> (accessed on 20 February 2020).
2. Arnold, C.L., Jr.; Gibbons, C.J. Impervious surface coverage: The emergence of a key environmental indicator. *J. Am. Plan. Assoc.* **1996**, *62*, 243–258. [[CrossRef](#)]
3. Defries, R.S.; Rudel, T.; Uriarte, M.; Hansen, M. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nat. Geosci.* **2010**, *3*, 178–181. [[CrossRef](#)]
4. Dewan, A.M.; Yamaguchi, Y. Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization. *Appl. Geogr.* **2009**, *29*, 390–401. [[CrossRef](#)]
5. Grimm, N.B.; Faeth, S.H.; Golubiewski, N.E.; Redman, C.L.; Wu, J.; Bai, X.; Briggs, J.M. Global change and the ecology of cities. *Science* **2008**, *319*, 756–760. [[CrossRef](#)] [[PubMed](#)]
6. Barnett, T.P.; Pierce, D.W.; Hidalgo, H.G.; Bonfils, C.; Santer, B.D.; Das, T.; Bala, G.; Wood, A.W.; Nozawa, T.; Mirin, A.A.; et al. Human-induced changes in the hydrology of the western United States. *Science* **2008**, *319*, 1080–1083. [[CrossRef](#)] [[PubMed](#)]
7. Carpenter, S.R.; Caraco, N.F.; Correll, D.L.; Howarth, R.W.; Sharpley, A.N.; Smith, V.H. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecol. Appl.* **1998**, *8*, 559–568. [[CrossRef](#)]
8. Meybeck, M.; Helmer, R. The quality of rivers: From pristine stage to global pollution. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **1989**, *75*, 283–309. [[CrossRef](#)]
9. Everard, M.; Moggridge, H.L. Rediscovering the value of urban rivers. *Urban Ecosyst.* **2012**, *15*, 293–314. [[CrossRef](#)]
10. Findlay, S.J.; Taylor, M.P. Why rehabilitate urban river systems? *Area* **2016**, *38*, 312–325. [[CrossRef](#)]
11. Francis, R.A. Positioning urban rivers within urban ecology. *Urban Ecosyst.* **2012**, *15*, 285–291. [[CrossRef](#)]
12. Paul, M.J.; Meyer, J.L. Streams in the urban landscape. *Annu. Rev. Ecol. Syst.* **2001**, *32*, 333–365. [[CrossRef](#)]

13. Bascaron, M. Establishment of a methodology for the determination of water quality. *Bol. Inf. Medio Ambient.* **1979**, *9*, 30–51.
14. Brown, R.M.; McClelland, N.I.; Deininger, R.A.; Tozer, R.G. A water quality index: Do we dare? *Water Sew. Works* **1970**, *117*, 339–343.
15. Canadian Council of Ministers of the Environment. Canadian Water Quality Guidelines for the Protection of Aquatic Life: CCME Water Quality Index 1.0, User's Manual. In *Canadian Environmental Quality Guidelines*; Canadian Council of Ministers of the Environment: Edmonton, AB, Canada, 2001.
16. Cude, C.G. Oregon water quality index: A tool for evaluating water quality management effectiveness. *J. Am. Water. Resour. Assoc.* **2001**, *37*, 125–137. [[CrossRef](#)]
17. Ramakrishnaiah, C.R.; Sadashiyah, C.; Ranganna, G. Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. *J. Chem.* **2009**, *6*, 523–530. [[CrossRef](#)]
18. Abbasi, T.; Abbasi, S.A. *Water Quality Indices*; Elsevier: Amsterdam, The Netherlands, 2012.
19. Lumb, A.; Sharma, T.C.; Bibeault, J.F. A review of genesis and evolution of water quality index (WQI) and some future directions. *Water Qual. Expo. Health* **2011**, *3*, 11–24. [[CrossRef](#)]
20. Sutadian, A.D.; Muttill, N.; Yilmaz, A.G.; Perera, B.J.C. Development of river water quality indices—A review. *Environ. Monit. Assess.* **2016**, *188*, 58. [[CrossRef](#)]
21. Huang, J.; Zhang, Y.; Arhonditsis, G.B.; Gao, J.; Chen, Q.; Wu, N.; Dong, F.; Shi, W. How successful are the restoration efforts of China's lakes and reservoirs? *Environ. Int.* **2019**, *123*, 96–103. [[CrossRef](#)]
22. Vatanpour, N.; Malvandi, A.M.; Talouki, H.H.; Gattinoni, P.; Scesi, L. Impact of rapid urbanization on the surface water's quality: A long-term environmental and physicochemical investigation of Tajan river, Iran (2007–2017). *Environ. Sci. Pollut. Res.* **2020**, *27*, 8439–8450. [[CrossRef](#)]
23. Srivastava, P.K.; Mukherjee, S.; Gupta, M.; Singh, S.K. Characterizing monsoonal variation on water quality index of River Mahi in India using geographical information system. *Water Qual. Expo. Health* **2011**, *2*, 193–203. [[CrossRef](#)]
24. Verma, R.K.; Murthy, S.; Tiwary, R.K.; Verma, S. Development of simplified WQIs for assessment of spatial and temporal variations of surface water quality in upper Damodar river basin, eastern India. *Appl. Water Sci.* **2019**, *9*, 21. [[CrossRef](#)]
25. Şener, Ş.; Şener, E.; Davraz, A. Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). *Sci. Total Environ.* **2017**, *584*, 131–144. [[CrossRef](#)] [[PubMed](#)]
26. Tian, Y.; Jiang, Y.; Liu, Q.; Dong, M.; Xu, D.; Liu, Y.; Xu, X. Using a water quality index to assess the water quality of the upper and middle streams of the Luanhe River, northern China. *Sci. Total Environ.* **2019**, *667*, 142–151. [[CrossRef](#)]
27. Koçer, M.A.T.; Sevgili, H. Parameters selection for water quality index in the assessment of the environmental impacts of land-based trout farms. *Ecol. Indic.* **2014**, *36*, 672–681. [[CrossRef](#)]
28. Ma, Z.; Song, X.; Wan, R.; Gao, L. A modified water quality index for intensive shrimp ponds of *Litopenaeus vannamei*. *Ecol. Indic.* **2013**, *24*, 287–293. [[CrossRef](#)]
29. Wu, Y.; Chen, J. Investigating the effects of point source and nonpoint source pollution on the water quality of the East River (Dongjiang) in South China. *Ecol. Indic.* **2013**, *32*, 294–304. [[CrossRef](#)]
30. de Souza Pereira, M.A.; Cavalheri, P.S.; de Oliveira, M.Â.C.; Magalhães Filho, F.J.C. A multivariate statistical approach to the integration of different land-uses, seasons, and water quality as water resources management tool. *Environ. Monit. Assess.* **2019**, *191*, 549. [[CrossRef](#)]
31. Tripathi, M.; Singal, S.K. Use of Principal Component Analysis for parameter selection for development of a novel Water Quality Index: A case study of river Ganga India. *Ecol. Indic.* **2019**, *96*, 430–436. [[CrossRef](#)]
32. Tripathi, M.; Singal, S.K. Allocation of weights using factor analysis for development of a novel water quality index. *Ecotoxicol. Environ. Saf.* **2019**, *183*, 109510. [[CrossRef](#)]
33. Wu, Z.; Wang, X.; Chen, Y.; Cai, Y.; Deng, J. Assessing river water quality using water quality index in Lake Taihu Basin, China. *Sci. Total Environ.* **2018**, *612*, 914–922. [[CrossRef](#)]
34. Han, Q.; Tong, R.; Sun, W.; Zhao, Y.; Yu, J.; Wang, G.; Shrestha, S.; Jin, Y. Anthropogenic influences on the water quality of the Baiyangdian Lake in North China over the last decade. *Sci. Total Environ.* **2020**, *701*, 134929. [[CrossRef](#)]
35. Rodríguez-Romero, A.J.; Rico-Sánchez, A.E.; Mendoza-Martínez, E.; Gómez-Ruiz, A.; Sedeño-Díaz, J.E.; López-López, E. Impact of changes of land use on water quality, from tropical forest to anthropogenic occupation: A multivariate approach. *Water* **2018**, *10*, 1518. [[CrossRef](#)]

36. Cha, Y.; Cho, K.H.; Lee, H.; Kang, T.; Kim, J.H. The relative importance of water temperature and residence time in predicting cyanobacteria abundance in regulated rivers. *Water Res.* **2017**, *124*, 11–19. [[CrossRef](#)] [[PubMed](#)]
37. Gebler, D.; Wiegler, G.; Szoszkiewicz, K. Integrating river hydromorphology and water quality into ecological status modelling by artificial neural networks. *Water Res.* **2018**, *139*, 395–405. [[CrossRef](#)] [[PubMed](#)]
38. Tong, S.T.; Chen, W. Modeling the relationship between land use and surface water quality. *J. Environ. Manag.* **2002**, *66*, 377–393. [[CrossRef](#)]
39. Whitehead, P.G.; Jin, L.; Bussi, G.; Voepel, H.E.; Darby, S.E.; Vasilopoulos, G.; Manley, R.; Rodda, C.; Hutton, C.; Hackney, C.; et al. Water quality modelling of the Mekong River basin: Climate change and socioeconomics drive flow and nutrient flux changes to the Mekong Delta. *Sci. Total Environ.* **2019**, *673*, 218–229. [[CrossRef](#)]
40. You, Q.; Fang, N.; Liu, L.; Yang, W.; Zhang, L.; Wang, Y. Effects of land use, topography, climate and socio-economic factors on geographical variation pattern of inland surface water quality in China. *PLoS ONE* **2019**, *14*, e0217840. [[CrossRef](#)]
41. Statistics Korea, 2017 Population and Housing Census of Korea. Available online: <http://kostat.go.kr/> (accessed on 26 February 2020).
42. Kannel, P.R.; Lee, S.; Lee, Y.S.; Kanel, S.R.; Khan, S.P. Application of water quality indices and dissolved oxygen as indicators for river water classification and urban impact assessment. *Environ. Monit. Assess.* **2007**, *132*, 93–110. [[CrossRef](#)]
43. Pesce, S.F.; Wunderlin, D.A. Use of water quality indices to verify the impact of Córdoba City (Argentina) on Suquia River. *Water Res.* **2000**, *34*, 2915–2926. [[CrossRef](#)]
44. Sánchez, E.; Colmenarejo, M.F.; Vicente, J.; Rubio, A.; García, M.G.; Travieso, L.; Borja, R. Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecol. Indic.* **2007**, *7*, 315–328. [[CrossRef](#)]
45. QGIS Development Team. QGIS geographic information system. Open Source Geospatial Foundation Project; QGIS Development Team. 2018. Available online: <http://qgis.osgeo.org> (accessed on 4 January 2019).
46. Environmental Systems Research Institute. *ArcGIS Desktop: Release 10.3*; Environmental Systems Research Institute: Redlands, CA, USA, 2014.
47. Forina, M.; Armanino, C.; Raggio, V. Clustering with dendrograms on interpretation variables. *Anal. Chim. Acta* **2002**, *454*, 13–19. [[CrossRef](#)]
48. Shrestha, S.; Kazama, F. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin. *Jpn. Environ. Model. Softw.* **2007**, *22*, 464–475. [[CrossRef](#)]
49. Singh, K.P.; Malik, A.; Mohan, D.; Sinha, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Res.* **2004**, *38*, 3980–3992. [[CrossRef](#)] [[PubMed](#)]
50. Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open Source Scientific Tools for Python. 2001. Available online: <https://www.scipy.org> (accessed on 25 December 2019).
51. Van Rossum, G.; Drake, F.L., Jr. *Python Reference Manual*; Centrum voor Wiskunde en Informatica: Amsterdam, The Netherlands, 1995.
52. Dojlido, J.; Raniszewski, J.; Woyciechowska, J. Water quality index applied to rivers in the Vistula river basin in Poland. *Environ. Monit. Assess.* **1994**, *33*, 33–42. [[CrossRef](#)] [[PubMed](#)]
53. Kendall, M.G. *Rank Correlation Methods*; Griffin: Oxford, UK, 1948.
54. Mann, H.B. Nonparametric tests against trend. *Econometrica* **1945**, *13*, 245–259. [[CrossRef](#)]
55. 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](#)]
56. Hussain, M.; Mahmud, I. pyMannKendall: A python package for non parametric Mann Kendall family of trend tests. *J. Open Source Softw.* **2019**, *4*, 1556. [[CrossRef](#)]
57. Kaiser, H.F. An index of factorial simplicity. *Psychometrika* **1974**, *39*, 31–36. [[CrossRef](#)]
58. Barlett, M.S. Properties of sufficiency and statistical tests. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **1937**, *160*, 268–282.
59. Horn, J.L. A rationale and test for the number of factors in factor analysis. *Psychometrika* **1965**, *30*, 179–185. [[CrossRef](#)]
60. Thompson, B. *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*; American Psychological Association: Washington, DC, USA, 2004.

61. Williams, B.; Onsmann, A.; Brown, T. Exploratory factor analysis: A five-step guide for novices. *Australas. J. Paramed.* **2010**, *8*, 1–13. [[CrossRef](#)]
62. Revelle, W.R. Psych: Procedures for Personality and Psychological Research. 2017. Available online: <https://CRAN.R-project.org/package=psych> (accessed on 16 February 2019).
63. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
64. Liu, C.W.; Lin, K.H.; Kuo, Y.M. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci. Total Environ.* **2003**, *313*, 77–89. [[CrossRef](#)]
65. Debels, P.; Figueroa, R.; Urrutia, R.; Barra, R.; Niell, X. Evaluation of water quality in the Chillán River (Central Chile) using physicochemical parameters and a modified water quality index. *Environ. Monit. Assess.* **2005**, *110*, 301–322. [[CrossRef](#)] [[PubMed](#)]
66. El Najjar, P.; Kassouf, A.; Probst, A.; Probst, J.L.; Ouaini, N.; Daou, C.; El Azzi, D. High-frequency monitoring of surface water quality at the outlet of the Ibrahim River (Lebanon): A multivariate assessment. *Ecol. Indic.* **2019**, *104*, 13–23. [[CrossRef](#)]
67. Liou, S.M.; Lo, S.L.; Wang, S.H. A generalized water quality index for Taiwan. *Environ. Monit. Assess.* **2004**, *96*, 35–52. [[CrossRef](#)] [[PubMed](#)]
68. Sun, W.; Xia, C.; Xu, M.; Guo, J.; Sun, G. Application of modified water quality indices as indicators to assess the spatial and temporal trends of water quality in the Dongjiang River. *Ecol. Indic.* **2016**, *66*, 306–312. [[CrossRef](#)]
69. Wang, J.; Fu, Z.; Qiao, H.; Liu, F. Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China. *Sci. Total Environ.* **2019**, *650*, 1392–1402. [[CrossRef](#)]
70. Nong, X.; Shao, D.; Zhong, H.; Liang, J. Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method. *Water Res.* **2020**, *178*, 115781. [[CrossRef](#)]
71. Landwehr, J.M.; Deininger, R.A. A comparison of several water quality indexes. *J. Water Pollut. Control Fed.* **1976**, *48*, 954–958.
72. Korea Environment Institute. *River Management and Ecological Restoration in Response to Climate Change*; Korea Environment Institute: Sejong, Korea, 2012.
73. National Institute of Environmental Research. *A Study on the Improvement for TMDL System Enforcement—Analysis of the Pollutant Load Contribution and the Establishment of Monitoring Standards for Decentralized Wastewater Treatment System*; National Institute of Environmental Research: Incheon, Korea, 2018.
74. National Institute of Environmental Research. *Customized Policy Support for Nonpoint Pollution Management and Water Circulation Improvement (III)*; National Institute of Environmental Research: Incheon, Korea, 2018.
75. National Institute of Environmental Research. *Hydraulic and Hydrologic Scenario Modelling for Prevention and Outbreak Response of Algal Bloom (II)—Focused on Monitoring-Based Contaminant Transport*; National Institute of Environmental Research: Incheon, Korea, 2018.
76. Goonetilleke, A.; Thomas, E.; Ginn, S.; Gilbert, D. Understanding the role of land use in urban stormwater quality management. *J. Environ. Manag.* **2005**, *74*, 31–42. [[CrossRef](#)]
77. Lee, J.H.; Bang, K.W. Characterization of urban stormwater runoff. *Water Res.* **2000**, *34*, 1773–1780. [[CrossRef](#)]
78. Yang, Y.Y.; Toor, G.S. Sources and mechanisms of nitrate and orthophosphate transport in urban stormwater runoff from residential catchments. *Water Res.* **2017**, *112*, 176–184. [[CrossRef](#)] [[PubMed](#)]
79. Belabed, B.E.; Meddour, A.; Samraoui, B.; Chenchouni, H. Modeling seasonal and spatial contamination of surface waters and upper sediments with trace metal elements across industrialized urban areas of the Seybouse watershed in North Africa. *Environ. Monit. Assess.* **2017**, *189*, 265. [[CrossRef](#)]
80. Hasan, H.H.; Jamil, N.R.; Aini, N. Water quality index and sediment loading analysis in Pelus River, Perak, Malaysia. *Procedia Environ. Sci.* **2015**, *30*, 133–138. [[CrossRef](#)]
81. Lane, P.N.; Sheridan, G.J. Impact of an unsealed forest road stream crossing: Water quality and sediment sources. *Hydrol. Process.* **2002**, *16*, 2599–2612. [[CrossRef](#)]
82. O'Mullan, G.D.; Juhl, A.R.; Reichert, R.; Schneider, E.; Martinez, N. Patterns of sediment-associated fecal indicator bacteria in an urban estuary: Benthic–pelagic coupling and implications for shoreline water quality. *Sci. Total Environ.* **2019**, *656*, 1168–1177. [[CrossRef](#)]

83. Fairbairn, D.J.; Karpuzcu, M.E.; Arnold, W.A.; Barber, B.L.; Kaufenberg, E.F.; Koskinen, W.C.; Novak, P.J.; Rice, P.J.; Swackhamer, D.L. Sources and transport of contaminants of emerging concern: A two-year study of occurrence and spatiotemporal variation in a mixed land use watershed. *Sci. Total Environ.* **2016**, *551*, 605–613. [[CrossRef](#)]
84. Whitehead, P.G.; Wilby, R.L.; Rattarbee, R.W.; Kernan, M.; Wade, A.J. A review of the potential impacts of climate change on surface water quality. *Hydrol. Sci. J.* **2009**, *54*, 101–123. [[CrossRef](#)]
85. Xu, G.; Li, P.; Lu, K.; Tantai, Z.; Zhang, J.; Ren, Z.; Wang, X.; Yu, K.; Shi, P.; Cheng, Y. Seasonal changes in water quality and its main influencing factors in the Dan River basin. *Catena* **2019**, *173*, 131–140. [[CrossRef](#)]

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



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).