

Article

# Analysis of Causal Relationships for Nutrient Removal of Activated Sludge Process Based on Structural Equation Modeling Approaches

Yejin Kim <sup>1,\*</sup> , Seulah Lee <sup>2</sup>, Yeongdae Cho <sup>1</sup> and Minsoo Kim <sup>1</sup>

<sup>1</sup> Department of Environmental Engineering, Catholic University of Pusan, Busan 46252, Korea; nvirocho@gmail.com (Y.C.); elim.kim@gmail.com (M.K.)

<sup>2</sup> Department of Environmental Engineering, Pusan National University, Busan 46241, Korea; sseulaya@pusan.ac.kr

\* Correspondence: yjkim@cup.ac.kr or yjkim@pusan.ac.kr; Tel.: +82-51-510-0621

Received: 23 January 2019; Accepted: 28 March 2019; Published: 3 April 2019



**Abstract:** The removal process of activated sludge in sewage treatment plants is very nonlinear, and removal performance has a complex causal relationship depending on environmental factors, influent load, and operating factors. In this study, how causal relationships are expressed in collected data was identified by structural equation modeling. First, path modeling was attempted as a preliminary step in structural equation model (SEM) construction and, as a result, the nutrient-removal mechanism could not be sufficiently represented as a direct causal relationship between measured variables. However, as a result of the deduced SEMs for effluent total nitrogen (T-N) and total phosphorus (T-P) concentrations, accompanied by exploratory factor analysis to extract latent variables, a causal network was formed that describes the direct or indirect effect of the latent factors and measured variables. Hereby, this study suggests that it is possible to construct an SEM explaining the nutrient-removal mechanism of the activated-sludge process with latent variables. Moreover, nonlinear features embedded in the mechanism could be represented by SEM, which is a model based on linearity, by including causal relations and variables that were not derived by path analysis. This attempt to model the direct and indirect causalities of the process could enhance the understanding of the process, and help decision making such as changing the driving conditions that would be required.

**Keywords:** causal model; nutrient removal; path model; structural equation modeling (SEM); wastewater treatment plant

## 1. Introduction

The history of the activated-sludge process, which is the core process that determines treatment performance in the sewage-treatment process, has already been around for over 100 years [1]. Although the removal mechanism is well-described in textbooks, its characteristics are highly nonlinear and site-specific, so when problems with effluent quality arise, the operator's empirical knowledge plays a big role in maintaining process stability [2].

There are two kinds of approaches to understanding the inner mechanism of the activated-sludge process, one is a mathematical model-based approach to identify the theoretical structure, and the other is a data analysis-based approach to extract meaningful information or verify a hypothesis. As a first method, activated sludge models (ASMs) have been widely applied to field-scale and pilot-scale plants [3]. However, due to the complexity of Monod equation-based dynamics, and the number of parameters, there has also been an attempt to simplify the model to improve usability [4–7].

The efficacy of the second approach, the attempt to understand via statistical analysis of the obtained data from the activated-sludge process, mainly depends on the quality and quantity of the obtained data. Unlike the mathematical model-based approach, this method makes it possible to use various measurement variables that cannot be utilized for mathematical models such as pH and oxidation reduction potential (ORP). Besides the sensor variables, operational factors such as the F/M ratio and airflow rate can be analyzed with influent/effluent water-quality variables within a dataset with the same utility. The most popular methodology applied to data analysis is principal component analysis (PCA), which has been adapted for sewage-treatment-process monitoring and identification of the operational state [8–10]. For the other approaches to enhance the understanding of the processing mechanism, there have been attempts to use signal processing and pattern recognition to detect operational abnormality or sensor faults. Wimberger et al. (2008) [11] applied a signal filter to detect the faults of the sewage-treatment process of sequencing batch reactor types. Baklouti et al. (2018) [12] tried to detect sensor faults using an improved particle filter. Chow et al. (2018) [13] developed a fault-detection and alarm algorithm based on the correlation analysis of signals of ultraviolet (UV) spectroscopy.

A causal model can be considered to produce a more intuitive result for the purpose of supporting the judgment of the process operator and improving understanding the process. For the activated-sludge process, the causal-relationship model has mainly been attempted in the form of a rule base. As remarkable cases, Cortés et al. (2003) [14] and Comas et al. (2003) [15] established a rule-based decision-support system for the settleability of activated sludge, which is relatively difficult to implement using mathematical models. The interpretation of the process to diagnose and understand it through the construction of the rule base led to the establishment of the Bayesian network model, a type of probabilistic causal model. Chong et al. (1996) [2], who previously pointed out the difficulty of using mathematical models, attempted a rule-based probabilistic approach using a Bayesian network. Huang et al. (1999) [16] established a graphical model for effluent quality and enhanced it to a fuzzy causal-network model. Aulinal et al. (2011) [17] developed a causal model for wastewater-treatment-plant operation and made it implementable as a knowledge-based decision-support system via web programming equipped with a reasoning function. Li et al. (2013) [18] tried to use the Bayesian network model to predict the effluent quality of a sewage-treatment plant, and Garvajal et al. (2017) [19] tried Bayesian belief-network modeling on the disinfection process of treated effluent to monitoring residual chlorine.

The basis of causal modeling can be obtained from intuitionally collected operational knowledge and statistical analysis of an accumulated historical database. As an enhanced approach of multivariate statistical analysis, structural equation modeling (SEM) approaches can be suggested as a tool to investigate the causal relationships between variables of interest. Although they have mainly been applied to verify a hypothesis in the field of social science, recently applications to the natural sciences, including biology, have also been frequently reported [20–25]. For the biology, it is easy to find out its recent applications which include SEM for biological communities change in river ecosystem [26–29]. The first SEM trial on the water-quality field was by Zou et al. (1994) [30], who pointed out the drawback of the general regression model is that it cannot reflect measurement errors, and emphasized the usefulness of SEM considering both the composition of latent variables and the causal relation between them in one model. Earlier, Ariana et al. (2010) [31] analyzed the denitrification potential in wetland soil by constructing an SEM. As a recent application of SEM in the water-quality field, He et al. (2016) [32] discussed the multidimensional inter-relationships between various water-quality factors and intertwined unit process and established a causal model for meteorological, hydraulic, and water-quality factors based on probabilistic approach. Zhu et al. (2018) [33] tried to set up an SEM describing the effect of floodgate operation on nitrogen transformation in a river. For wastewater or the sewage-treatment process, Moreira et al. (2008) [34] conducted path analysis for *Escherichia coli* removal at a sewage-treating pond. As can be seen from these studies, SEM cannot be considered as limited

to the social sciences. It can be applied to numerical variables from biology and the natural sciences, and it is especially useful in identifying direct and indirect causal relationships structured together.

This paper suggests some SEM approaches to discover the stressors on effluent T-N and T-P concentrations from a sewage-treatment plant, and their direct and indirect affecting structure, although influence factors are, of course, theoretically well-established in textbooks. However, in spite of the highly nonlinear characteristics of the activated sludge system, operators tend to linearly perceive the causality. Therefore, verifying the linearity of the causal relationship would conversely be useful information to the operator.

## 2. Materials and Methods

### 2.1. Operational Data Acquisition

The dataset used in this study was obtained from a field-scale A2/O sewage-treatment plant with a treatment capacity of 680,000 m<sup>3</sup>/day. Influent T-N and T-P concentration was 27 and 3 mg/L, respectively, and 8.8 and 1.1 mg/L for effluent. The daily operational records accumulated over 3 years consisted of 996 datasets for 42 variables. After excluding datasets that were missing records and outliers, 334 datasets were used to establish the path model and SEM. Because there was the opinion that the number of datasets should be greater than 150 [35], the amount of data prepared for this study was reasonable for trying SEM. The 42 variables, including meteorological variables, water-quality at each end of the unit process, and operational factors, are listed in Table 1. Then, the data were divided into two groups by random extraction using SPSS ver. 18; one group was used for model construction and the other for model validation. Meteorological data, such as air temperature (°C), rainfall (mm), and relative humidity (%), were included because they could act as a hidden factor influencing influent water quality. From the bioreactor, operational factors that affect nutrient-removal performance and state variables implicitly indicating them were included.

**Table 1.** Collected data variables.

| Items                   | Variables  |
|-------------------------|--|
| Weather conditions      | Air temperature (°C), rainfall (mm), relative humidity (%)   |
| Primary settling tank * | Water temperature (°C), pH, BOD (mg/L), COD (mg/L), SS (mg/L), T-N (mg/L), T-P (mg/L), alkalinity, S-BOD, HRT (h)  |
| Bioreactor **           | Flow (m <sup>3</sup> /d), water temperature (°C), pH, DO (mg/L), MLSS (mg/L), MLVSS (mg/L), SVI, sludge return ratio (m <sup>3</sup> /d), Sludge return ratio (%), F/M ratio, BOD loading (kg/m <sup>3</sup> ·day), SRT (day), A-SRT (day), Internal sludge-return ratio (%), ORP (mV), PO <sub>4</sub> -P (mg/L, at both anaerobic and anoxic tanks), NH <sub>4</sub> -N (mg/L, at both anoxic and oxic tanks), NO <sub>3</sub> -N (mg/L, at both anoxic and oxic tanks), air flow (m <sup>3</sup> /day), reactor volume (m <sup>3</sup> ), HRT (h) |
| Effluent                | Water temperature (°C), pH, BOD (mg/L), COD (mg/L), SS (mg/L), T-N (mg/L), T-P (mg/L), alkalinity, HRT (h)   |

\*: The items will be expressed with indication of "B\_in" (ex: B\_in\_BOD). \*\*: The items will be expressed with indication of "B\_" (ex: B\_SRT).

### 2.2. Structural Equation Model

#### 2.2.1. Path Model

Path analysis is an extension of multiple regression analysis [34], with a path model that depicts the direct and indirect effects of independent variables on one or more independent variables based on a hypothesis to be verified. Verification of the path model is performed by assigning data to the model and determining fitness. There are four types of direct causal relationships shown in Figure 1. Path analysis is the basis for constructing the basic structure of the SEM. Through path analysis, the causal relationship can be represented in more detail by confirming the inherent causal effects as

the direct, indirect, and total effect. The hypothesized relationship, illustrated at each path, is tested on its acceptance or not by the path direction and the significance of the path coefficient. The standardized path coefficient is key to explaining the strength of the causal path, which enables to compare it with other coefficients in one model. The reliability of the path coefficients should be confirmed based on test statistics such as critical ratio (C.R.) value and p-value. If the C.R. is above 1.96 and the p-value is smaller than 0.05, then the path is reliable. In this research, the p-value was used to judge reliability.



**Figure 1.** Path types used in a path model. (a) causal; (b) both-way causal; (c) correlate; (d) independent path.

To set up the path model for effluent T-N and T-P concentrations, the minimum variables considered to theoretically affect the targets were used to comprise the initial path models, respectively. Then, the initial model was examined for fitness using the fitness indices suggested in Table 2. Then, If the initial path model was judged to be appropriate, the model was extended by adding the variables and determining the fitness again. Through repetition of this process, the modified path model was derived. This method of establishing the initial model and then confirming the modified final path model through expansion and validation is a generalized process to deduce the rational model, which also has been proposed by Santibáñez-Andrade et al. (2015) [36].

**Table 2.** Evaluation indices about structure goodness.

| Index   | Acceptance Level | Classification          |
|---|------------------|-------------------------|
| Q value   | <3: Excellent    | Parsimony-fit indices   |
| Goodness of Fit Index (GFI)                     | >0.9: Excellent  | Absolute-fit indices    |
| Root Mean Square Error of Approximation (RMSEA) | <0.05: Excellent |                         |
|   | <0.08: Good      | Absolute-fit indices    |
|   | <0.1: Normal     |                         |
| Adjusted Goodness of Fit Index (AGFI)           | >0.8: Good       | Absolute-fit indices    |
| CFI   | >0.9: Excellent  | Incremental-fit indices |

### 2.2.2. Structural Equation Model

The structural equation model is a statistical multivariate model that can confirm the direct and indirect effect of independent variables on the dependent, and the degree of the causal relationship between them, related to a specific phenomenon [37–39]. Wright (1934) [40] introduced this methodology to the field of natural science to the biological population. Since then, the scope of applications has greatly expanded and has been applied to social sciences, psychology, chemistry, and biology. After setting up the initial model reflecting the hypothesis, regression analysis, correlation analysis, and factor analysis were used to confirm the causal relationship. Particularly, factor analysis should be used to investigate the causal structure of hidden factors and complete the structure. This is due to the difference from the path model, which is SEM, including latent variables representing the complex effects of various measured variables. Factor-analysis application in establishing SEM can be divided into exploratory factor analysis and confirmatory factor analysis. The confirmatory factor analysis-based SEM starts with the initial model constructed by the relationships that are already known, or a designed hypothesis. Then, the model is verified with the fitness indices, and the causal path in the model is tested for its significance. On the other hand, exploratory factor analysis performs factor analysis to find out and set the latent variables. This approach has the disadvantage of not being able to fully reflect the theoretical causal relation because it constitutes a causal relation depending on the measured data. The structural equation model established in this study was based on exploratory factor analysis because the latent variables were deduced by factor analysis and they formed the basic

structure of the SEM. When the theoretical causal relationships are not clear, this procedure is regarded as a rational approach in both natural-science [27,41] and social fields [42].

The structural equation model consists of factors and their causal relationships forming the structure model as described in Figure 2, and the measured variables ( $x_1, x_2, \dots$ ) related to latent variables with their observed error terms form the observed model. Here, the factor is also called a latent variable, which implies the combined effect of the observed variable. Each measurement variable contains an error, which means the extent to which the latent variable cannot be fully explained by measurement variables. This error comes from the measurement process and it is one of the main reasons for using latent variables in SEM. The advantage of the structural equation is that it can analyze the combined effects of a large number of factors. In addition, a measurement error can be considered and its size can be derived. However, since the assumption that data must follow a normal distribution is satisfied, the more data, the more preferable it is [35].

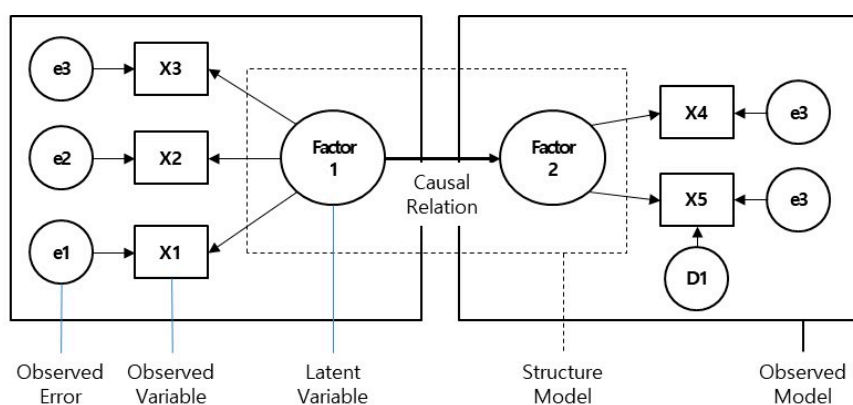


Figure 2. Basic concept of structural equation model.

The fitness index to examine the suitability of the constructed SEM can be divided into three categories, absolute-fit indices, parsimony-fit indices, and incremental-fit indices (Table 2). The absolute fit index is an index of how well the research model reflects input data. It can be said that the developed model itself is evaluated without comparing with other models. The chi-square statistic is often used as a representative. However, chi-square statistic has a disadvantage of underestimating the fit as the size of the sample increases. As an alternative to chi-square, RMSEA (root mean square error of approximation) is a value adjusted by the chi-square statistic to the degree of freedom and sample size and has been used in many studies recently. The RMSEA of 0.05 or less means very good fitness and below 0.8 means good fitness. If it is less than 0.1, it can be said to be normal. GFI (goodness of fit index), an index used in this study as an absolute fit index, is the most widely used fit index in the structural equation model. GFI is also indirectly affected by the size of the samples as well as the chi-square statistic. In order to compensate for this, AGFI (adjusted goodness of fit index) is used as an indicator that the GFI value is modified by using the degree of freedom of the model. Both GFI and AGFI have values between 0 and 1, and 0.9 and 0.8, respectively, indicate a good fit of the model. The Incremental fit index is an index indicating how well the research model reflects the input data than the null model, unlike the absolute fit index that evaluates the model itself. The most representative index is the NFI (normed fit index), but NFI has the disadvantage of less sensitivity to the complexity of the model. Therefore, CFI (comparative fit index) can be used as modified NFI. Parsimony fit index means the degree to which the model reaches the maximum degree of fit required fitness for each estimated coefficient. As a representative parsimony fit index, there is Q-value used in this study, which is a value obtained by dividing the chi-square value by degree of freedom to compensate for the disadvantage of the chi-square statistic.

The absolute fit index, the incremental fit index, and the parsimony fit index evaluate the model by different criteria. When evaluating one structural equation model, it is not desirable to evaluate by

only one kind of index. In most studies, two or more indices from a least two categories are applied and evaluated. In this study, five indexes were evaluated in three categories to evaluate the fitness of the structural equation model to be constructed more carefully using data with high uncertainty.

This research adapted the Q value, goodness of fit (GFI), and root mean square error of approximation (RMSEA) among absolute-fit indices. The adjusted GFI (AGFI) of parsimony-fit indices and the comparative fit index (CFI) as an incremental-fit index were also adapted for checking compliance with various criteria. GFI and AGFI were suggested by Jöreskog and Sörbom [37], and they can be regarded as the most popular indices to test SEM fitness. If the GFI value is greater than 0.9, it can be generally judged as properly constructed [33,42–44]. An AGFI value larger than 0.9 is seen as “excellent fitness”, but over 0.8 can be “acceptable” [45,46]. The smaller the value of RMSEA, the most popular index, the better [47], but no larger than 0.06 [42,44,48] or 0.08 [20,43,49,50]. The CFI, the most frequently used incremental index, guarantees the model’s fitness when greater than 0.8 [49] or 0.9 [36,50], and best close to the value of 1.0 [20,50–52]. In this research, AMOS v.20.0 (IBM Statistics, Inc.) was used for path modeling and the SEM.

### 3. Results

#### 3.1. Structural Equation Modeling for Effluent T–N

##### 3.1.1. Path Model of Effluent T–N

##### Initial Path Model for Effluent T–N

The initial path model for effluent T–N concentration (Figure 3a) consists of seven variables, including the target: airflow rate (B\_Airflow), return-sludge ratio (B\_Sludge return ratio), DO concentration in the aeration tank (B\_DO), MLSS concentrations in the aeration tank (B\_MLSS), ammonia and nitrate concentration in the aeration tank (Oxic\_NH<sub>4</sub>-N and Oxic\_NO<sub>3</sub>-N), and effluent T–N concentration (Effluent T–N). These variable configurations fully explain the theoretical knowledge that nitrate and ammonia in the aeration tank affect effluent T–N. After an initial trial of model setting with minimum variables as Oxic\_NH<sub>4</sub>-N and Oxic\_NO<sub>3</sub>-N, B\_DO and Effluent T–N, other variables were added or removed by repetitively checking the fitness to obtain the best model showing acceptable index values. In the confirmed initial path model, the indices showed that its fitness was acceptable except for an RMSEA of 0.089 (Table 3).

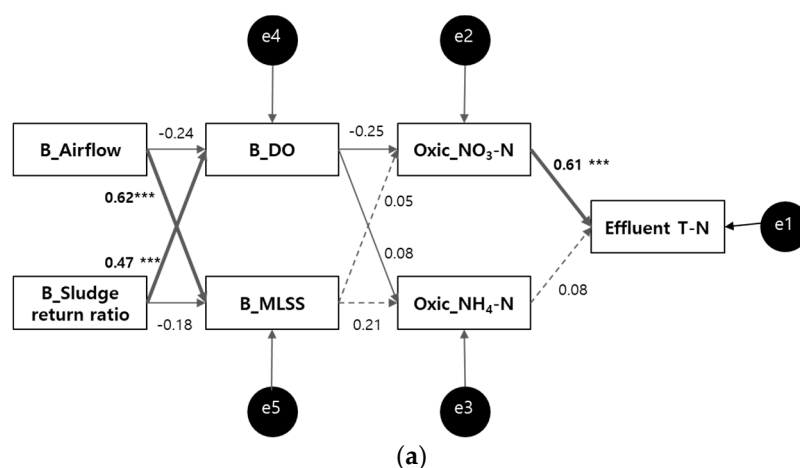
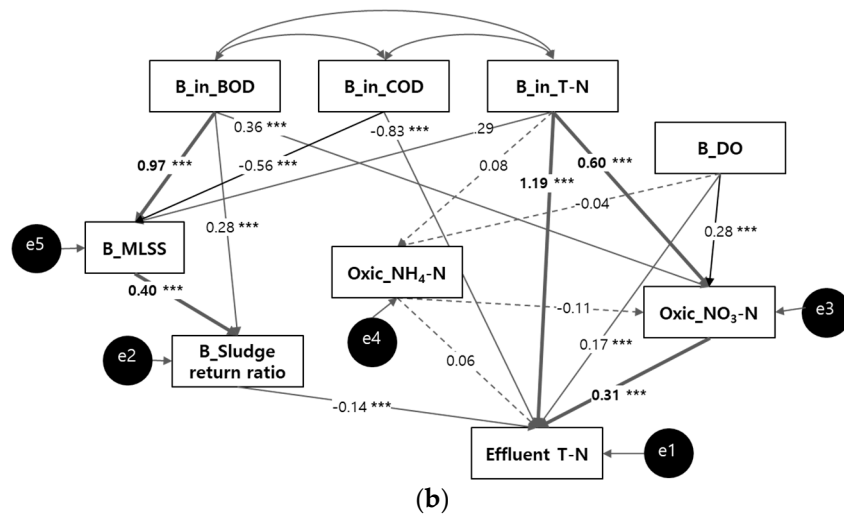


Figure 3. Cont.



**Figure 3.** Developed path model for effluent T-N concentration. (dashed line: insignificant path coefficient, \*\*\*:  $p$ -value < 0.001). (a) Initial path model; (b) modified path model.

**Table 3.** Fitness results of developed path models for effluent T-N concentration.

| Classification | Goodness-of-Fit Criterion | Initial Model      |                    | Modified Model     |                    |
|----------------|---------------------------|--------------------|--------------------|--------------------|--------------------|
|                |                           | Result             | Validation         | Result             | Validation         |
| Q value        | below 3                   | 2.303<br>(Fitness) | 1.373<br>(Fitness) | 1.652<br>(Fitness) | 2.111<br>(Fitness) |
| GFI            | above 0.9                 | 0.991<br>(Fitness) | 0.976<br>(Fitness) | 0.967<br>(Fitness) | 0.958<br>(Fitness) |
| AGFI           | above 0.8                 | 0.945<br>(Fitness) | 0.938<br>(Fitness) | 0.907<br>(Fitness) | 0.881<br>(Fitness) |
| RMSEA          | below 0.05<br>(below 0.1) | 0.089<br>(Fitness) | 0.047<br>(Fitness) | 0.063<br>(Fitness) | 0.082<br>(Fitness) |
| CFI            | above 0.9                 | 0.983<br>(Fitness) | 0.979<br>(Fitness) | 0.989<br>(Fitness) | 0.980<br>(Fitness) |

The direction of each path and significant results are shown in the initial path model (Figure 3a). The path results of statistical significance showed that the two paths from B\_MLSS to both Oxic\_NO<sub>3</sub>-N and Oxic\_NH<sub>4</sub>-N did not satisfy the 95% level of significance ( $p < 0.05$ ). Comparing the significant path coefficients except for these paths, there were three paths that had relatively high importance. First, airflow rate (B\_Airflow) had a relatively high effect on MLSS concentration in the bioreactor. The second, the strength of the path from the NO<sub>3</sub>-N concentration in the oxic tank (Oxic\_NO<sub>3</sub>-N) to the effluent T-N (Effluent T-N) concentration was high, a natural result in theory. Consistent with the theory is the return-sludge ratio (B\_Sludge return ratio) having a large influence on the DO (dissolved oxygen) concentration in the bioreactor (B\_DO), and B\_DO having a significant effect on both NH<sub>4</sub>-N and NO<sub>3</sub>-N concentration in the aeration tank. The strength of the path from Oxic\_NO<sub>3</sub>-N to effluent T-N in the aeration tank was high, as its path coefficient was 0.61, and significant ( $p < 0.05$ ), whereas the path from Oxic\_NH<sub>4</sub>-N to effluent T-N was not significant. This result explains the mechanism of nitrogen removal in the activated-sludge process. When nitrification efficiency is high, ammonia concentration in the aeration tank would be low. The standardized path coefficient means the level of influence, and it corresponds to the regression coefficient in the standardized model. In this respect, it is reasonable that the path coefficient from B\_Airflow to B\_MLSS is high, and the path coefficient to B\_DO is low (−0.24). The DO concentration measured in the reactor indicates the remaining DO after being used for carbon oxidation and nitrification. When the airflow rate is excessive, then it is possible to have a linear correlation with DO concentration. However, the effect of MLSS concentration on nitrate and ammonia nitrogen was evaluated as statistically insignificant. This is due to the fact that

MLSS concentration is a composite result of microbial growth, the amount of returned sludge and its degree of concentration, and inflow rate.

The deduced initial path model was verified with the prepared data group for validation and not used in the model setup. Validation fitness was acceptable, with a Q value of 1.373, GFI of 0.976, AGFI of 0.938, RMSEA of 0.047, and CFI of 0.979 (Table 3).

#### Modified path model for effluent T-N

The initial path model for effluent T-N was enhanced to describe the causal relationships of a larger number of variables (Figure 3b). As the initial path model could not explain the effect of influent variation, the measured variables of bioreactor inflow concentrations were added and tested with fitness indices. Through the iterative process of finding a suitable model, airflow rate (B\_Airflow) was omitted from the causal network, and BOD, COD, and T-N inflow concentration (B\_in\_BOD, B\_in\_COD, B\_in\_T-N) were introduced into the model. The fitness of this modified path model for effluent T-N was acceptable for all indices.

There are four paths interpreted as meaningless by having p-values greater than 0.05, and they are represented by dashed lines in Figure 3b. For the significant paths, there are three paths that are from influent BOD concentration to MLSS concentration in the bioreactor, from influent T-N concentration to effluent T-N showing effluent T-N variation depending on influent loading, and the path from influent T-N to the aeration-tank nitrate. These two paths related to influent and effluent T-N concentration means that, even with stable nitrogen-removal performance, fluctuation of effluent T-N is due to influent fluctuation. It can also be interpreted as showing the limitation of nitrogen removal using inner nitrate return from the aeration tank to the anoxic. This is in agreement with the well-known theory and the processing mechanism, and it is meaningful to find these relationships embedded in the measured variable numbers.

MLSS concentration remains in the model because influent BOD is included in it as an influencing factor on effluent T-N. There was an attempt to build a path model that includes influent BOD as an essential internal carbon source for denitrification and excludes MLSS concentration, but the model could not meet the fitness standard. The insignificant path from B\_DO to Oxid\_NH<sub>4</sub>-N can be estimated as having the same theory as the remaining DO concentration mentioned above.

As the validation results of the modified path model for effluent T-N using the datasets prepared by random extraction for model verification, the model fitness was acceptable with the index values of 2.111 for Q, 0.958 for GFI, 0.881 for AGFI, 0.082 for RMSEA, and 0.980 for CFI.

#### 3.1.2. SEM for Effluent T-N

As mentioned previously, factor analysis was performed to extract latent variables that would be the core of the causal relationship. With eigenvalues larger than 1 and the absolute value of factor loading higher than 0.5 (Table 4), four latent factors were extracted (Table 5). All the factor loadings are according to the nature of the variables under the influence of each latent factor, the name of each latent variable was assigned. To set up the first SEM trial for effluent T-N, the latent variables were structured upon the causal concept obtained from the results of path modeling. Then, through the repeated process of tentatively adding the measured variables to the latent variables, the initial SEM was extended. For each trial, model fitness was determined and a model was established that includes the most and most reasonable variables (Figure 4). There are three meaningful paths with statistical significance. First, the return flow-related factor to effluent T-N shows the developed SEM having high correspondence with the well-known nitrogen-removal mechanism in process types of MLE or A2/O adapting internal nitrate recycling. It should be noted that the inner sludge-recycling ratio and the SRT of the aeration tank could not be included in the path model, whereas the latent variables under them could form an irreplaceable part of the SEM. The second, the path from the inflow-related factor to the operational factor shows natural causality between operational actions and influent variation. The third path is from the operational factor to effluent T-N. This path can be combined with the



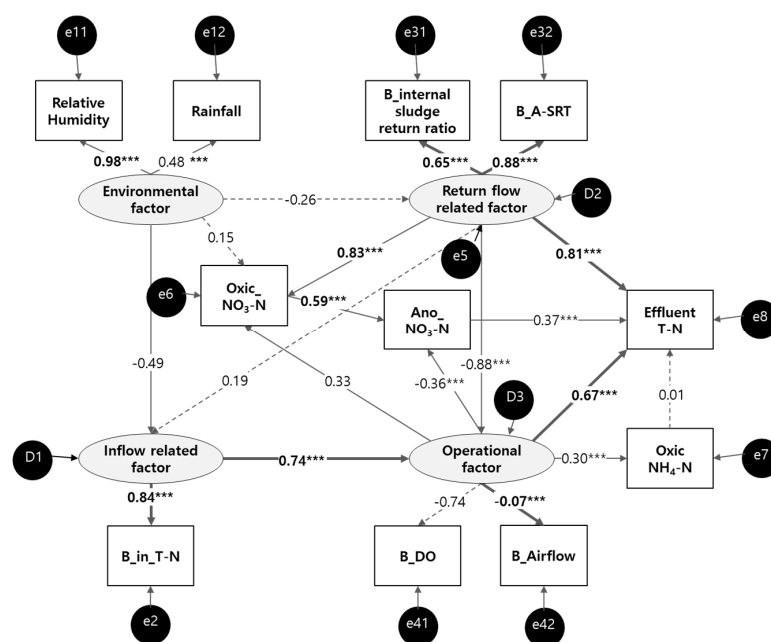
path from the influent-related factor to the operational factor to perfectly explain the dynamics of the nitrogen-removal process. In detail, the combined path describes the procedure of nitrogen removal, and influent variation induces a change in the operational factor, which, in turn, affects effluent quality along with the influence of the return flow-related factor. The SEM constructed for effluent T-N is different from the modified path model, as it was possible to extensively model the variable effects that were difficult to be utilized in the path model.

**Table 4.** Factor loadings for the variables related to effluent T-N concentration.

| Variable                       | Component No. |              |              |              |
|--------------------------------|---------------|--------------|--------------|--------------|
|                                | 1             | 2            | 3            | 4            |
| Rainfall                       | 0.038         | −0.652       | −0.232       | 0.253        |
| Relative humidity              | −0.114        | −0.796       | −0.046       | −0.137       |
| B_pH                           | −0.154        | −0.758       | −0.136       | −0.126       |
| B_in_BOD                       | <b>0.858</b>  | 0.278        | 0.006        | 0.103        |
| B_in_COD                       | <b>0.933</b>  | 0.214        | 0.118        | 0.047        |
| B_in_SS                        | <b>0.874</b>  | −0.192       | −0.103       | −0.065       |
| B_in_TN                        | <b>0.725</b>  | 0.490        | 0.177        | 0.254        |
| B_in_TP                        | <b>0.869</b>  | 0.401        | 0.116        | 0.021        |
| B_Sludge return ratio          | 0.293         | <b>0.790</b> | 0.011        | 0.026        |
| B_internal sludge return ratio | 0.365         | <b>0.650</b> | 0.449        | 0.134        |
| B_A-SRT                        | 0.229         | <b>0.794</b> | 0.241        | 0.255        |
| B_SRT                          | 0.163         | <b>0.880</b> | 0.127        | 0.078        |
| B_Air flow                     | 0.069         | 0.102        | −0.704       | −0.044       |
| B_DO                           | −0.190        | −0.139       | <b>0.639</b> | 0.027        |
| B_MLSS                         | 0.325         | −0.456       | <b>0.565</b> | 0.226        |
| B_SVI                          | 0.262         | −0.450       | <b>0.551</b> | 0.370        |
| B_F/M ratio                    | 0.000         | −0.085       | −0.690       | −0.064       |
| Effluent_T-N                   | 0.333         | 0.268        | 0.266        | <b>0.610</b> |

**Table 5.** Latent variables deduced by factor analysis for effluent T-N concentration.

| Factor              | Variables   |
|---------------------|---|
| Environmental       | Rainfall, relative humidity, B_pH                                     |
| Inflow-related      | B_in_BOD, B_in_COD, B_in_SS, B_in_T-N, B_in_T-P                       |
| Return flow-related | B_Sludge return ratio, B_Internal sludge-return ratio, B_A-SRT, B_SRT |
| Operational         | B_Air flow, B_DO, B_MLSS, B_SVI, B_F/M ratio                          |



**Figure 4.** Structural equation model for effluent T-N concentration.

Sanches Fernandes et al. (2018) [29] described the type of connection between a latent variable and the measured variables to which it is linked as two types of model, the reflective model and formative model. In this research, the formative model was implemented. The formative model is the case where the effect of latent variables is expressed as linked measured variables. On the other hand, measured variables are viewed as latent-variable causes. Therefore, “B\_Internal sludge return ratio” and “B\_A-SRT” represent the effect of the “Return flow-related factor” latent variable. In the cases of “Inflow-related factor” and “Environmental factor”, the relationship can be regarded as reflective. However, when the reflective type was applied to the factors, model fitness deteriorated. Therefore, it should be interpreted as rainfall and humidity included in the dataset, expressing the effect of environmental impact. The suggested SEM for effluent T–N shows the importance of the causal structure of the latent variables, and that the extent to which a latent variable is exposed depends on the choice and usage of the measurement variables in the target domain. The fitness of the model was acceptable for all indices as shown in Table 6.

**Table 6.** Fitness of structural equation model for effluent T–N.

| Factor  | Criterion    | Result (Test) | Result (Validation) | Fitness/Not |
|---------|--------------|---------------|---------------------|-------------|
| Q value | <3           | 2.455         | 2.604               | Fitness     |
| GFI     | >0.9         | 0.924         | 0.919               | Fitness     |
| AGFI    | >0.8         | 0.847         | 0.838               | Fitness     |
| RMSEA   | <0.05 (<0.1) | 0.094         | 0.098               | Fitness     |
| CFI     | > 0.9        | 0.917         | 0.912               | Fitness     |

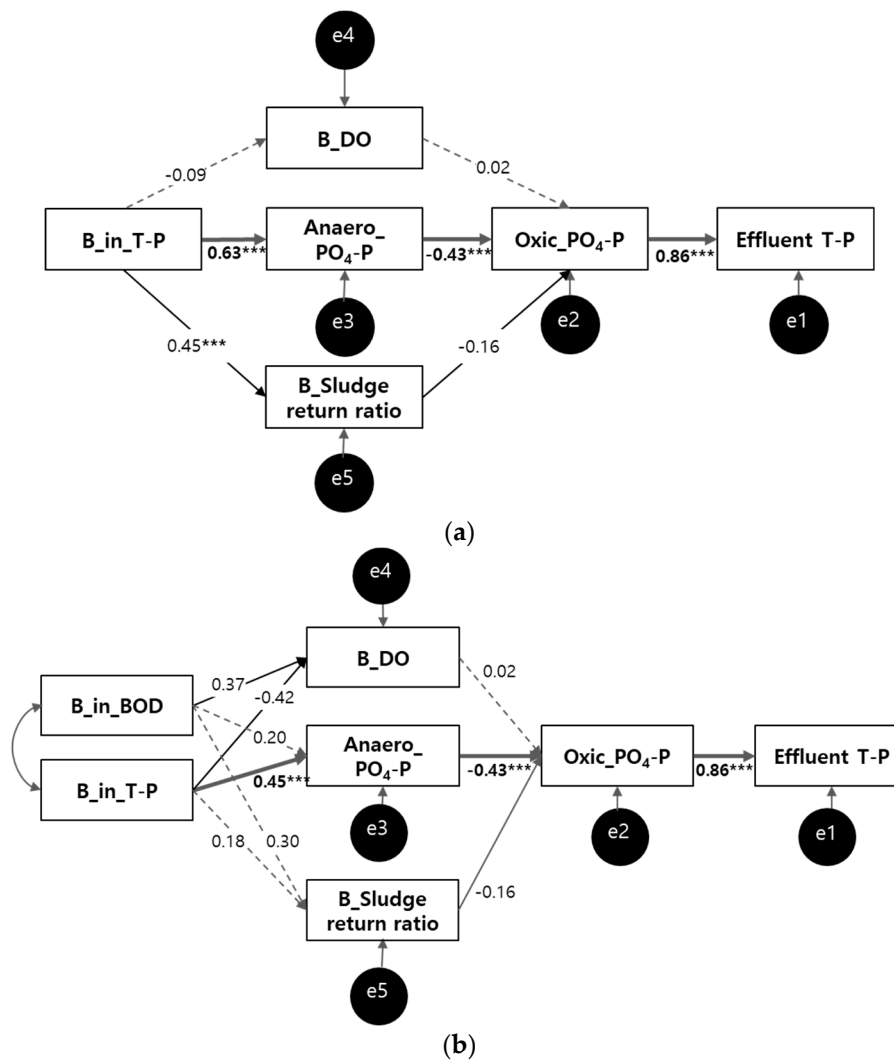
### 3.2. Structural Equation Modeling for Effluent T–P

#### 3.2.1. Path Model of Effluent T–P

##### Initial Path Model for Effluent T–P

The initial path derived for effluent T–P concentration consisted of five related variables affecting the target: influent T–P concentration in the bioreactor (B\_in\_T–P), DO concentration of the aeration tank (B\_DO), PO<sub>4</sub>–P concentration in the anaerobic tank (Anaero\_PO<sub>4</sub>–P), sludge-recycling ratio from the settling tank to the bioreactor (B\_Sludge return ratio), and PO<sub>4</sub>–P concentration of the aeration tank (Oxic\_PO<sub>4</sub>–P) (Figure 5a). The fitness of this initial path model was acceptable for all indices and is listed in Table 7. There were two paths with an insignificant *p*-value, from influent T–P concentration (B\_in\_T–P) to PO<sub>4</sub>–P concentration of the aeration tank (Oxic\_PO<sub>4</sub>–P), and from B\_DO to Oxic\_PO<sub>4</sub>–P. This is due to the fact that PO<sub>4</sub>–P level was kept low in spite of the variation of influent T–P loading in the bioreactor. For the relationship between B\_DO and Oxic\_PO<sub>4</sub>–P, DO concentration is a variable that can largely be affected by the amount of carbon oxidation and nitrification, not only by excessive phosphorus accumulation.

There are two paths that have strong causality, from PO<sub>4</sub>–P concentration in the aeration tank (Oxic\_PO<sub>4</sub>–P) to effluent T–N, and from influent T–P concentration and PO<sub>4</sub>–P concentration in the anaerobic tank. The former path is natural because PO<sub>4</sub>–P in the aerobic tank would pass the settling tank almost without any change. The latter would also be the result of adding the effect of phosphorus release in the anaerobic tank to the influent loading effect. DO concentration of the aeration tank (B\_DO) was included to complete the fitness of path model but did not have significant path-correlation values in either the path to or from that variable. The two paths explain the effect of the sludge amount for phosphorus release and uptake. The higher the number of micro-organisms for phosphorus uptake, the lower the measured phosphorus concentration after uptake, so the path coefficient is interpreted as having a weak negative value. The fitness of the initial path model for effluent T–N was acceptable for all indices, with 1.659 for Q, 0.974 for GFI, 0.931 of AGFI, 0.063 of RMSEA, and 0.981 for CFI.



**Figure 5.** Developed path model for effluent T-N concentration. (a) Initial path model; (b) modified path model.

**Table 7.** Fitness results of the developed path models for effluent T-P concentration.

| Classification | Goodness-of-Fit Criterion | Initial Model      |                    | Modified Model     |                    |
|----------------|---------------------------|--------------------|--------------------|--------------------|--------------------|
|                |                           | Result             | Validation         | Result             | Validation         |
| Q value        | below 3                   | 1.263<br>(Fitness) | 1.659<br>(Fitness) | 1.902<br>(Fitness) | 1.949<br>(Fitness) |
| GFI            | above 0.9                 | 0.981<br>(Fitness) | 0.974<br>(Fitness) | 0.969<br>(Fitness) | 0.967<br>(Fitness) |
| AGFI           | above 0.8                 | 0.949<br>(Fitness) | 0.931<br>(Fitness) | 0.913<br>(Fitness) | 0.908<br>(Fitness) |
| RMSEA          | below 0.05(below 0.1)     | 0.040<br>(Fitness) | 0.063<br>(Fitness) | 0.074<br>(Fitness) | 0.076<br>(Fitness) |
| CFI            | above 0.9                 | 0.995<br>(Fitness) | 0.981<br>(Fitness) | 0.987<br>(Fitness) | 0.982<br>(Fitness) |

Modified Path Model for Effluent T-P

The final confirmed path model for effluent T-P (Figure 5b) with the acceptable fitness results listed in Table 7, as influent BOD concentration (B\_in\_BOD) was added to reflect the influent effect

on the initial model. The significance of the path coefficient results indicated four unreliable paths, “B\_in\_BOD→Anaero\_PO<sub>4</sub>-P”, “B\_in\_BOD→B\_Sludge return ratio”, “B\_in\_T-P→B\_Sludge return ratio”, and “B\_DO→Oxic PO<sub>4</sub>-P”. Except for these, all path coefficients were significant, especially the two paths of “B\_in\_T-P→Anaero\_PO<sub>4</sub>-P” and “Oxic\_PO<sub>4</sub>-P→Effluent T-P”, which coincided with the results of the initial model. Fitness was all acceptable with a Q value of 1.949, GFI of 0.967, 0.908 of AGFI, 0.076 of RMSEA, and 0.982 of CFI.

### 3.2.2. SEM for Effluent T-P

Like the model-derivation process for T-N, the latent variables extracted by factor analysis and their relationships were added to the initially obtained paths from path modeling, from PO<sub>4</sub>-P in the anaerobic tank to PO<sub>4</sub>-P of the aerobic tank, followed by effluent T-P. Through the deduced factor loadings from the factor analysis (Table 8), the latent variables were extracted as the four groups listed in Table 9. In contrast to T-N, MLSS concentration, sludge-return rate, and SRT-related variables were all tied to one latent variable, defined as operating factors. Other measured variables, such as DO, SVI, and F/M ratios, were grouped into reactor-related latent factors. After setting up the causal network between the latent variables, through the iterative trials of adding or removing the measured variables by checking the fitness of the model, the final SEM was deduced as shown in Figure 6.

**Table 8.** Factor loadings for the variables related to effluent T-N concentration.

| Variable                       | Component No. |              |                |              |                |
|--------------------------------|---------------|--------------|----------------|--------------|----------------|
|                                | 1             | 2            | 3              | 4            | 5              |
| Rainfall                       | 0.083         | −0.508       | −0.176         | −0.111       | 0.548          |
| Relative humidity              | −0.093        | −0.797       | −0.107         | −0.055       | 0.079          |
| B_pH                           | −0.106        | −0.680       | −0.009         | −0.252       | −0.370         |
| B_in_BOD                       | <b>0.845</b>  | 0.313        | 0.006          | 0.118        | −0.023         |
| B_in_COD                       | <b>0.920</b>  | 0.226        | 0.105          | 0.133        | −0.077         |
| B_in_SS                        | <b>0.893</b>  | −0.134       | −0.070         | −0.116       | 0.082          |
| B_in_TN                        | <b>0.685</b>  | 0.489        | 0.119          | 0.365        | −0.111         |
| B_in_TP                        | <b>0.852</b>  | 0.401        | 0.118          | 0.110        | −0.138         |
| B_Sludge return ratio          | 0.292         | <b>0.827</b> | 0.143          | −0.174       | −0.021         |
| B_internal sludge return ratio | 0.346         | <b>0.628</b> | 0.479          | 0.179        | −0.101         |
| B_A-SRT                        | 0.191         | 0.216        | <b>0.696</b>   | 0.494        | −0.052         |
| B_SRT                          | 0.132         | 0.059        | <b>0.779</b>   | 0.409        | −0.167         |
| B_Air flow                     | 0.031         | 0.078        | − <b>0.807</b> | 0.131        | −0.211         |
| B_DO                           | −0.156        | 0.058        | 0.131          | 0.115        | <b>0.791</b>   |
| B_MLSS                         | 0.296         | <b>0.643</b> | −0.442         | 0.094        | 0.060          |
| B_SVI                          | 0.253         | 0.272        | −0.367         | 0.026        | <b>0.677</b>   |
| B_F/M ratio                    | −0.010        | −0.043       | −0.006         | −0.107       | − <b>0.715</b> |
| Effluent_T-P                   | 0.087         | −0.220       | 0.291          | <b>0.751</b> | −0.093         |

**Table 9.** Latent variables deduced by factor analysis for effluent T-P concentration.

| Factor           | Variables   |
|------------------|---|
| Environmental    | Rainfall, relative humidity, B_pH   |
| Inflow-related   | B_in_BOD, B_in_COD, B_in_SS, B_in_T-N, B_in_T-P   |
| Operational      | B_Air flow, B_MLSS, B_Sludge return ratio, B_Internal sludge return ratio, B_A-SRT, B_SRT |
| Reactor-related. | B_DO, B_SVI, B_F/M ratio  |

Fitness (Table 10) showed that all indices except for RMSEA satisfied the fitness-standard values for training model. RMSEA is decided by degrees of freedom and the model error. The larger the model error and the smaller the degree of freedom are, the larger the RMSEA value. In this case, the degree of freedom was 48, so the RMSEA value over 0.1 could be caused by the model error. Since

all other indices except RMSEA met the criteria, the deduced model can be regarded as having a normal level of fitness.

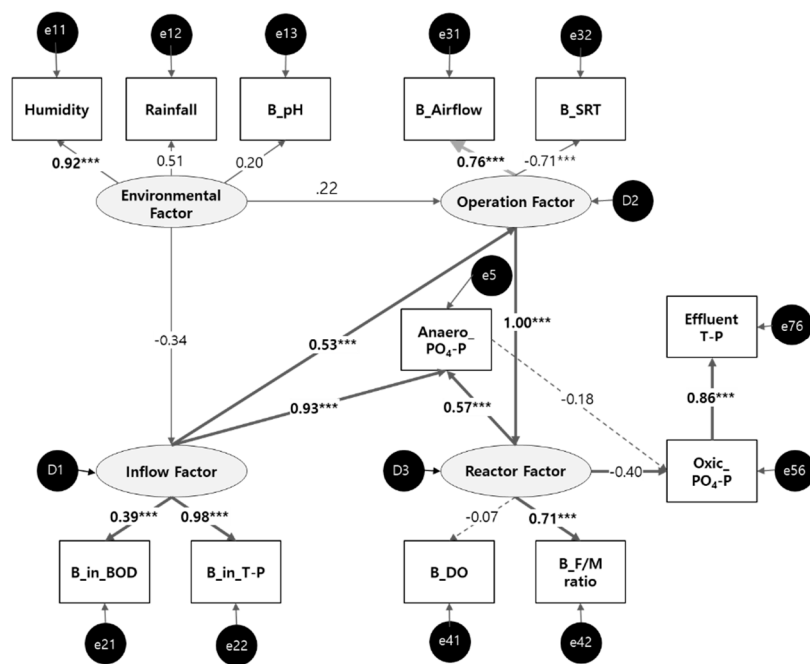


Figure 6. Structural equation model for effluent T-P concentration.

Table 10. Fitness of the structural equation model for effluent T-P.

| Factor  | Criterion    | Result (Test)                  | Result (Validation) |
|---------|--------------|--------------------------------|---------------------|
| Q value | <3           | 2.892 (Fitness)                | 3.439 (Not fitness) |
| GFI     | >0.9         | 0.883 (Not Fitness)            | 0.861 (Not fitness) |
| AGFI    | >0.8         | 0.810 (Fitness)                | 0.775 (Not fitness) |
| RMSEA   | <0.05 (<0.1) | 0.107 (Not Fitness, but close) | 0.121 (Not fitness) |
| CFI     | >0.9         | 0.911 (Fitness)                | 0.854 (Not fitness) |

As the results of the path coefficients and their statistical significance, some important paths could be extracted. As in the case of T-N, the influent-related factor strongly affects the operational factor. The influent factor was expressed as influent T-P and BOD concentration, and the operational factor as the amount of airflow (B\_Airflow). These causalities between latent factors and the measurement variables expressed by them can be interpreted as presupposing the phosphorus-removal mechanism. The strong causality from the operational factor to the reactor factor proves that this model was constructed on well-measured and -managed reliable data. The strong path coefficient from the reactor factor to PO<sub>4</sub>-P in the aeration tank can be interpreted as the effect of the F/M ratio on phosphorus release and uptake. From the fitness validation results, no index value met the fitness criteria (Table 10). However, since index values are distributed near the reference value, and the fitness results of the test SEM were good, it is considered that better fitness can be obtained if a larger number of datasets is applied.

#### 4. Discussion

In this study, we investigated the linearity of the dynamics of nutrient removal during sewage treatment by constructing a path model and a structural equation model. As already known, the mechanism of nutrient removal is nonlinear, so complex linear causal relationships cannot be identified through path modeling. This is due to limitations of the path model, which can only structure linear relationships between measured variables. On the other hand, the SEM can reflect the

influence between multiple latent variable factors, thus representing the direct and indirect effect of more variables. In addition, the path model does not take into account the error term of each variable, but structural equation modeling can take into account the error term of each measurement variable and the structural error of latent variables so that a more accurate causal model can be constructed.

In this sense, it should be noticed that the causal relationship, which was not derived by the path diagram, could be implemented as a structural equation model where the causal relationship of the latent variable acts as the main subject. It can be assumed that the nonlinearity of nutrient removal using activated sludge can be reflected to some extent as a causal relationship of latent variables.

All of the path models derived from this study satisfied the fitness criterion. However, the SEM of T-P did not satisfy the fitness standard in the model-verification process. This can be due to a variety of causes, but the most likely is the number of data points used. Anderson and Gerbing [35] stated that the number of data points should be larger than at least 150. In this study, the number of data points used for model establishment and validation was 167, respectively, and it would have been possible to have better fitness if a larger number of data points were used. Another possible cause is uncertainty from data organization. Generally, it takes time to affect the biological characteristics of activated sludge when operating conditions are changed. Therefore, direct causal trends are not well-reflected in the measured data.

The path model is known as the most basic form of the structural equation model. Path analysis is performed on the premise of the following assumptions. The relationship between independent variables and dependent variables is linear and summative. In addition, there is no problem in the measurement of the variable itself, and there is no measurement error. These assumptions make a difference from the structural equation model. The structural equation model admits the existence of measurement error. The effect of independent variables on dependent variables can be expressed in a complex way through latent variables. However, path modeling should not be regarded as having any value in comparison with the structural equation model. Path modeling can be tried as a previous step in the structural equation modeling in that it can identify the structure of the combined effects of the linearity and the significance of that existing between one variable and one variable. Preferably, the derived path model for the same data can be the backbone of the structural equation model. However, due to the characteristic of this study, the implementation of nonlinearity through latent variables, the ideal model transformation process has not been realized. Since the derived path model reflects the nutrient removal mechanism, the attempt of the path model is not meaningless. However, it can be said that the structural equation modeling is more suitable to express causality of wastewater treatment plant data with high measurement uncertainty and complex cause-effect relationships.

In addition, it is noteworthy that theoretically known knowledge is somewhat reflected in the SEM proposed in this study. So far, SEM has been mainly used to verify an explored, so there is no case of applying it to the removal mechanism of the sewage-treatment process. This is because the type of process and its mechanism have reliably been recognized for certainty. However, the purpose of this study was to ascertain whether such mechanisms are inherent to the data, and how their causality is expressed with some degree of linearity. As a result, there are not many variables that are linear causative factors of effluent quality. The most obvious factor of linearity was the path from the concentration of nutrients in the aeration tank, the last compartment of the bioreactor, to effluent quality. The complex and linear causal paths between the variables, on the other hand, were found to be rare, but the SEM was constructed to sufficiently reflect the causal relationship between the potential variables.

A further consideration in future studies is that more data should be collected. The water quality items used in this research were limited. If the water quality data obtained from the field included various carbon source concentration in the influent, the path model and the developed SEM would have been able to express more mechanisms. The data used in this research included only BOD and COD for carbonaceous material. Therefore, the detailed mechanism could not be implemented in the developed models. In addition, because the T-N removal and T-P removal mechanisms are closely

related, modeling of the combination of these two dependent variables could be possible. As already theoretically known, T-N and T-P cannot be removed without a carbon source. The nitrate produced in the oxic tank flows into the anaerobic tank through the settling tank and works as an electron acceptor for anoxic carbon oxidation inhibiting phosphorus release. For this reason, in the developed SEM for T-N, Anaero\_PO<sub>4</sub>-P, a variable indicating the phosphorus release is described to be affected by two latent variables such as “Inflow factor” and “React factor” comprising F/M ratio. In addition, the SEM described the combined effect of the latent variable “Operational factor” including aeration flow and the “Reactor factor” including F/M ratio to the PO<sub>4</sub>-P concentration in the oxic tank. This reflects that there is an indirect combined effect of the amount of the carbon source and the aeration flow in the oxic tank to the PO<sub>4</sub>-P concentration in the oxic tank.

In addition, if excessive aeration is given to the oxic tank to obtain maximum nitrification, oxygen will flow into the anoxic tank by the nitrate return flow, thereby inhibiting denitrification. Therefore, it is hard to simply conclude the relationships between nitrogen and DO as proportional or inverse, as shown in Figure 3. There is an inverse proportional relationship between B\_DO and Oxic\_NH<sub>4</sub>-N whereas weak positive correlation with oxic\_NO<sub>3</sub>-N. This is expressed at the SEM for T-N as that the latent variable “Operational factor” has a direct effect to the effluent T-N with high correlation factor, 0.67. For the case of Oxic\_NO<sub>3</sub>-N, the fact that it is affected by the latent variable “Return flow related factor” with the high correlation value 0.83 can be regarded as reflecting the theoretical fact for the level of Oxic\_NO<sub>3</sub>-N decided by the nitrate return ratio.

Therefore, this study is meaningful in improving the understanding of the activated sludge mechanism, and as a new SEM attempt to remove nutrients in the activated sludge process. If this kind of approach is applied to the data obtained other wastewater treatment plants with the same process and gets the same results, it will add confidence in the results obtained in this research. The SEM approach can be easily implemented using various tools such as AMOS, R, and LISREL. The linear and nonlinear characteristics found from SEM based approaches can be used to set up a new type of model such as a Bayesian network model or to simplify the complex nonlinearity of Activated Sludge Models.

## 5. Conclusions

From a traditional point of view, the performance of a sewage-treatment plant is determined by the control of operational factors against changing influent conditions, but uncertainty is high because the main removal dynamics are based on a biological mechanism. Despite these uncertainties, theories are well-established and the causal relationship of biological nutrient removal mechanisms are explained in detail in textbooks as a result of many studies on process dynamics. Many researchers have tried mathematical modeling to improve models and theories with simulations but have rarely confirmed such theories from the accumulated data.

In this study, we tried to find out whether the causal relationship of the nutrient-removal mechanism is actually embedded in a historical database using a structural equation model. From the path-modeling results, which is a process of the structural equation model, we concluded that the relationship between variables cannot be represented as linear. However, the results of the construction of the structural equation model were interesting because the SEM could well describe the nutrient-removal mechanism in the sewage-treatment process using latent variables. This study implies that the causal-relation model of the sewage-treatment plant could be constructed by expressing nonlinear causal relations as the combined effect of latent variables. As previously discussed in the literature review, this study is the first to attempt the SEM approach to the operation data of the sewage treatment plant. Regarding T-P, the fitness in the validation task was not perfect, but the SEM derived for the T-P as training model and for T-N could confirm what kind of associations exist between the numerical data.

**Author Contributions:** As the corresponding author of this paper, Y.K. led the whole process of applying methodologies, writing manuscript, and submission procedures. S.L. and Y.C. took the part related with SPSS work to get the actual results. M.K. contributed to the literature review.

**Funding:** This work was supported by the Korean Ministry of Environment as “Program for promoting commercialization of promising environmental technologies (Project no. 201700193001).”

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. Jenkins, D.; Wanner, J. *Activated Sludge—100 Years and Counting*; IWA Publishing: London, UK, 2014.
2. Chong, H.G. Rule-based versus probabilistic approaches to the diagnosis of faults in wastewater treatment processes. *Artif. Intell. Eng.* **1996**, *1*, 265–273. [[CrossRef](#)]
3. Rieger, L.; Gillot, S.; Langergraber, G.; Ohtsuki, T.; Shaw, A.; Takács, I.; Winkler, S. Chapter 5.4 Calibration and validation. In *Guidelines for Using Activated Sludge Models*; Scientific and Technical Report No. 22; IWA Publishing: London, UK, 2013.
4. Cadet, C. Simplifications of Activated Sludge Model with preservation of its dynamic accuracy. *IFAC Proc. Vol.* **2014**, *47*, 7134–7139. [[CrossRef](#)]
5. Müller, T.G.; Noykova, N.; Gyllenberg, M.; Timmer, J. Parameter identification in dynamical models of anaerobic waste water treatment. *Math. Biosci.* **2002**, *177–178*, 147–160. [[CrossRef](#)]
6. Kim, H.W.; Lim, H.; Wie, J.; Lee, I.; Colosimo, M.F. Optimization of modified ABA<sup>2</sup> process using linearized ASM2 for saving aeration energy. *Chem. Eng. J.* **2014**, *251*, 337–342. [[CrossRef](#)]
7. Santa Cruz, J.A.; Mussati, S.F.; Scenna, N.J.; Gernaey, K.V.; Mussati, M.C. Reaction invariant-based of reduction of the activated sludge model ASM1 for batch applications. *J. Environ. Chem. Eng.* **2016**, *4*, 3654–3664. [[CrossRef](#)]
8. Lee, J.M.; Yoo, C.; Choi, S.W.; Vanrolleghem, P.A.; Lee, I.B. Nonlinear process monitoring using kernel principal component analysis. *Chem. Eng. Sci.* **2004**, *59*, 223–234. [[CrossRef](#)]
9. Lee, D.S.; Park, J.M.; Vanrolleghem, P.A. Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor. *J. Biotechnol.* **2005**, *116*, 195–210. [[CrossRef](#)] [[PubMed](#)]
10. Moon, T.S.; Kim, Y.J.; Kim, J.R.; Cha, J.H.; Kim, D.H.; Kim, C.W. Identification of process operating state with operational map in municipal wastewater treatment plant. *J. Environ. Manag.* **2009**, *90*, 772–778. [[CrossRef](#)] [[PubMed](#)]
11. Wimberger, D.; Verde, C. Fault diagnosticability for an aerobic batch wastewater treatment process. *Control Eng. Pract.* **2008**, *16*, 1344–1353. [[CrossRef](#)]
12. Baklouti, I.; Mansouri, M.; Hamida, A.B.; Nounou, H.; Nounou, M. Monitoring of wastewater treatment plants using improved univariate statistical technique. *Process Saf. Environ.* **2018**, *116*, 287–300. [[CrossRef](#)]
13. Chow, C.W.K.; Liu, J.; Li, J.; Swain, N.; Reid, K.; Saint, C.P. Development of smart data analysis tools to support wastewater treatment plant operation. *Chemom. Intell. Lab. Syst.* **2018**, *177*, 140–150. [[CrossRef](#)]
14. Cortés, U.; Martínez, M.; Comas, J.; Sánchez-Marré, M.; Poch, M.; Rodríguez-Roda, I. A conceptual model to facilitate knowledge sharing for bulking solving in wastewater treatment plants. *AI Commun.* **2003**, *16*, 279–289.
15. Comas, J.; Rodríguez-Roda, I.; Sánchez-Marré, M.; Cortés, U.; Freixó, A.; Arrá, J.; Poch, M. A knowledge-based approach to the deflocculation problem: Integrating on-line, off-line, and heuristic information. *Water Res.* **2003**, *37*, 2377–2387. [[CrossRef](#)]
16. Huang, Y.C.; Wang, X.Z. Application of fuzzy causal networks to waste water treatment plants. *Chem. Eng. Sci.* **1999**, *54*, 2731–2738. [[CrossRef](#)]
17. Aulinas, M.; Nieves, J.C.; Cortés, U.; Poch, M. Supporting decision making in urban wastewater systems using a knowledge-based approach. *Environ. Model. Softw.* **2011**, *26*, 562–572. [[CrossRef](#)]
18. Li, D.; Yang, H.Z.; Liang, X.F. Prediction analysis of a wastewater treatment system using a Bayesian network. *Environ. Model. Softw.* **2013**, *40*, 140–150. [[CrossRef](#)]
19. Garvajal, G.; Roser, D.J.; Sisson, S.A.; Keegan, A.; Khan, S.J. Bayesian belief network modelling of chlorine disinfection for human pathogenic viruses in municipal wastewater. *Water Res.* **2017**, *109*, 144–154. [[CrossRef](#)]
20. Grace, J.B. *Structural Equation Modeling and Natural Systems*; Cambridge University Press: Cambridge, UK, 2006.



21. Arhonditsis, G.B.; Stow, C.A.; Steinberg, L.J.; Kenney, M.A.; Lathrop, R.C.; McBride, S.J.; Reekhow, K.H. Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecol. Model.* **2006**, *192*, 385–409. [[CrossRef](#)]
22. Grace, J.B.; Bollen, K.A. Representing general theoretical concepts in structural equation models; the role of composite variables. *Environ. Ecol. Stat.* **2008**, *15*, 191–213. [[CrossRef](#)]
23. Grace, J.B.; Anderson, T.M.; Olff, H.; Scheiner, S.M. On the specification of structural equation models for ecological systems. *Ecol. Monogr.* **2010**, *80*, 67–87. [[CrossRef](#)]
24. Grace, J.R.; Schoolmaster, D.R., Jr.; Guntenspergen, G.R.; Little, A.M.; Mitchell, B.R.; Miller, K.M.; Schweiger, E.W. Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere* **2012**, *3*, 1–44. [[CrossRef](#)]
25. Capmourteres, V.; Anand, M. Assessing ecological integrity: A multi-scale structural and functional approach using Structural Equation Modeling. *Ecol. Indic.* **2016**, *71*, 258–269. [[CrossRef](#)]
26. Pajunen, V.; Luoto, M.; Soinen, J. Unravelling direct and indirect effects of hierarchical factors driving microbial stream communities. *J. Biogeogr.* **2017**, *44*, 2376–2385. [[CrossRef](#)]
27. Hatami, R. Development of a protocol for environmental impact studies using causal modeling. *Water Res.* **2018**, *138*, 206–223. [[CrossRef](#)]
28. Villeneuve, B.; Valette, J.P.; Souchon, Y.; Usseglio-Polatera, P. Direct and indirect effects of multiple stressors on stream invertebrates across watershed, reach and site scales: A structural equation modelling better informing on hydromorphological impacts. *Sci. Total Environ.* **2018**, *612*, 660–671. [[CrossRef](#)] [[PubMed](#)]
29. Sanches Fernandes, L.F.; Fernandes, A.C.P.; Ferreira, A.R.L.; Cortes, R.M.V.; Pacheco, F.A.L. A partial least squares—Path modeling analysis for the understanding of biodiversity loss in rural and urban watersheds in Portugal. *Sci. Total Environ.* **2018**, *626*, 1069–1085. [[CrossRef](#)] [[PubMed](#)]
30. Zou, S.; Yu, Y.S. A general structural equation model for river water quality data. *J. Hydrol.* **1994**, *162*, 197–209. [[CrossRef](#)]
31. Ariana, E.S.G.; Melissa, A.K.; Curtis, J.R. Examining the relationship between ecosystem structure and function using structural equation modelling: A case study examining denitrification potential in restored wetland soils. *Ecol. Model.* **2010**, *221*, 761–768. [[CrossRef](#)]
32. He, J. Probabilistic Evaluation of Causal Relationship between Variables for Water Quality Management. *J. Environ. Inform.* **2016**, *28*, 110–119. [[CrossRef](#)]
33. Zhu, L.; Zhou, H.; Xie, X.; Li, X.; Zhang, D.; Jia, L.; Wei, Q.; Zhao, Y.; Wei, Z.; Ma, Y. Effects of floodgates operation on nitrogen transformation in a lake based on structural equation modeling analysis. *Sci. Total Environ.* **2018**, *631–632*, 1311–1320. [[CrossRef](#)]
34. Moreira, J.F.; Cabral, A.R.; Oliveira, R.; Silva, S.A. Causal model to describe the variation of faecal coliform concentrations in a pilot-scale test consisting of ponds aligned in series. *Ecol. Eng.* **2009**, *35*, 791–799. [[CrossRef](#)]
35. Anderson, J.C.; Gerbing, D.W. Structural equation modeling in practice: A review and recommended two-step approach. *Psychol. Bull.* **1988**, *103*, 411. [[CrossRef](#)]
36. Santibáñez-Andrade, G.; Castillo-Argüero, S.; Vega-Peña, E.V.; Lindig-Cisneros, R.; Zavala-Hurtado, J.A. Structural equation modeling as a tool to develop conservation strategies using environmental indicators: The case of the forests of the Magdalena river basin in Mexico City. *Ecol. Indic.* **2015**, *54*, 124–136. [[CrossRef](#)]
37. Jöreskog, K.G.; Sörbom, D. *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Square Methods*; Scientific Software, Inc.: Mooresville, IN, USA, 1986.
38. Jöreskog, K.G.; Sörbom, D. *LISREL 8: User's Reference Guide*; Scientific Software International, Inc.: Chicago, IL, USA, 2001.
39. Quinn, G.P.; Keough, M.J. *Experimental Design and Data Analysis for Biologists*; Cambridge University Press: New York, NY, USA, 2002.
40. Wright, S. The method of path coefficient. *Ann. Math. Stat.* **1934**, *5*, 161–215. [[CrossRef](#)]
41. Belkhir, L.; Narany, T.S. Using Multivariate Statistical Analysis, Geostatistical Techniques and Structural Equation Modeling to Identify Spatial Variability of Groundwater Quality. *Water Resour. Manag.* **2015**, *29*, 2073–2089. [[CrossRef](#)]
42. Hou, D.; Al-Tabbaa, A.; Chen, H.; Mamic, I. Factor analysis and structural equation modelling of sustainable behavior in contaminated land remediation. *J. Clean. Prod.* **2014**, *84*, 439–449. [[CrossRef](#)]

43. Hair, J.F.J.; Anderson, R.E.; Tatham, R.L.; Black, W.C. *Multivariate Data Analysis*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1998.
44. Hooper, D.; Coughlan, J.; Mullen, M. Structural equation modeling: Guidelines for determining model fit. *Electron. J. Bus. Res. Methods* **2008**, *6*, 53–60.
45. Doll, W.J.; Xia, W.; Torkzadeh, G. A confirmatory factor analysis of the end-user computing satisfaction instrument. *MIS Quart.* **1994**, *18*, 357–369. [[CrossRef](#)]
46. Baumgartner, H.; Homburg, C. Applications of Structural Equation Modeling in Marketing and Consumer Research: A review. *Int. J. Res. Mark.* **1996**, *13*, 139–161. [[CrossRef](#)]
47. Ryberg, K.R. Structural equation model of total phosphorus loads in the Red River of the north basin, USA and Canada. *J. Environ. Qual.* **2017**, *46*, 1072–1080. [[CrossRef](#)]
48. Estiri, H. A structural equation model of energy consumption in the United States: Untangling the complexity of per-capita residential energy use. *Energy Res. Soc. Sci.* **2015**, *6*, 109–120. [[CrossRef](#)]
49. Dang, H.L.; Li, E.; Nuberg, I.; Bruwer, J. Understanding farmer’s adaptation intention to climate change: A structural equation modelling study in the Mekong Delta, Vietnam. *Environ. Sci. Policy* **2014**, *41*, 11–22. [[CrossRef](#)]
50. Wang, K.; Qiao, Y.; Li, H.; Zhang, H.; Yue, S.; Ji, X.; Liu, L. Structural equation model of the relationship between metals in contaminated soil and in earthworm (*Metaphire californica*) in Hunan Province, subtropical China. *Ecotoxicol. Environ. Saf.* **2018**, *156*, 443–451. [[CrossRef](#)] [[PubMed](#)]
51. Ullman, J.B. Structural equation modeling: Reviewing the basics and moving forward. *J. Personal. Assess.* **2006**, *87*, 35–50. [[CrossRef](#)] [[PubMed](#)]
52. Bagozzi, R.P.; Yi, Y. Specification, evaluation and interpretation of structural equation models. *J. Acad. Mark. Sci.* **2012**, *40*, 8–34. [[CrossRef](#)]



© 2019 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/>).