

Article

Predicting Residual Adsorbable Organic Halides Concentrations in Industrial Wastewater Using Typical Wastewater Parameters

Hannah Milh ¹, Kwinten Van Eyck ², Bert Bastiaens ², Steven De Laet ², Dirk Leysen ², Deirdre Cabooter ³ and Raf Dewil ^{1,*}

¹ Department of Chemical Engineering, Process and Environmental Technology Lab, KU Leuven, J. De Nayerlaan 5, B-2860 Sint-Katelijne-Waver, Belgium; hannah.milh@kuleuven.be

² InOpSys NV, Zandvoortstraat 12A, 2800 Mechelen, Belgium; kwinten.vaneyck@inopsys.eu (K.V.E.); bert.bastiaens@inopsys.eu (B.B.); steven.delaet@inopsys.eu (S.D.L.); dirk.leysen@inopsys.eu (D.L.)

³ Department of Pharmaceutical and Pharmacological Sciences, Pharmaceutical Analysis, KU Leuven, Herestraat 49, 3000 Leuven, Belgium; deirdre.cabooter@kuleuven.be

* Correspondence: raf.dewil@kuleuven.be; Tel.: +32-496-07-49-90

Received: 29 March 2020; Accepted: 13 May 2020; Published: 9 June 2020



Abstract: The aim of this study was to predict the residual adsorbable organic halides (AOX) concentration in an industrial wastewater using conventional, easy-to-measure wastewater parameters. In a pilot test unit, the wastewater was subjected to ozonation at various intensities, resulting in an AOX-removal and hence varying AOX concentrations. In first instance, the parameters used for modeling were selected using Pearson and Spearman correlations. Secondly, multiple linear regression (MLR) was used as a modeling tool to predict both the soluble and total AOX concentration in wastewater samples. To prevent overfitting, a 10-fold cross-validation was carried out. It was found that both the soluble and the total AOX concentration can be predicted using typical wastewater parameters. The measured parameters were pH, chloride concentration, Water-Soluble Organic Carbon concentration (WSOC), UV-VIS spectrum, turbidity, and Solids Removable by Filtration (SRF). Out of these parameters, the following parameters were found to be significant for prediction of the total AOX concentration: turbidity; SRF; UV-VIS absorbance at 200; 227, and 250 nm; and pH. UV-VIS absorbance at 200 and 227 nm and turbidity of the wastewater were found to contribute significantly to the final model. For the soluble AOX concentration, the significant parameters were turbidity; SRF; absorbance at 200, 227, and 250 nm; pH, and chloride concentration. Here, UV-VIS absorbance at 200 and 227 nm were found to contribute significantly to the final model. The obtained final models had an adjusted R^2 of 0.921 and 0.916 for the total and soluble AOX, respectively. As a result of the obtained models, both AOX concentrations can be predicted using parameters that are easier to determine. This allows for a significant reduction in wastewater sampling and analysis time and offers the opportunity to optimize the ozone dosing in the wastewater treatment process in the future.

Keywords: AOX prediction; multiple linear regression modelling; ozonation

1. Introduction

Adsorbable organic halides (AOX) are organic compounds containing halogen atoms (Cl, Br, and I) [1,2]. AOX are mainly formed in industrial processes using chlorine or chlorinated chemicals, such as bleaching. For example, the production of AOX is a major problem arising in the paper and pulp industries [3]. Because AOX are unwanted side products of industrial processes, they usually end up in the resulting wastewater. AOX can have carcinogenic or mutagenic properties and should, therefore, be removed from wastewater [4]. Conventional wastewater treatment processes are often

insufficient for realizing complete AOX removal [5]. Therefore, several other removal techniques have been proposed in the literature, including adsorption [5,6], biological treatment [7–9], nano and ultrafiltration [10,11], and advanced oxidation processes (AOPs). AOPs employ the production of highly reactive radicals (hydroxyl radicals, $\cdot\text{OH}$) to eliminate target substances through oxidation. Examples of AOPs used for AOX degradation are ozonation (alone or combined with UV irradiation) [12], peroxone treatment [13,14], and Fenton oxidation [15].

The analytical determination of AOX concentration is time-consuming and expensive. Furthermore, a real-time and continuous assessment of AOX concentration during the ozonation process could enable one to determine at which point AOX concentration has dropped below the preset limit and is hence useful as a control tool. This facilitates the optimization of the reaction time and hence the ozone and electricity consumption of the ozonation system. Therefore, the development of a more-straightforward determination method is highly desirable, preferentially via analysis tools that are (i) fast, (ii) nondestructive, (iii) environmentally friendly, and (iv) do not require any additional chemicals [16]. Previous studies have already suggested UV-VIS spectroscopy as an adequate tool to monitor the quality of wastewater, meeting these conditions [17]. The possibility of direct in-line and online monitoring is a big advantage that enables a fast analysis. Previous research has shown the applicability of UV-VIS as a predictor for the estimation of different wastewater parameters such as COD, BOD, TSS, and TOC [17–19].

It is well known that datasets resulting from water monitoring usually have a multidimensional structure, containing a large number of (often multicollinear) variables, making the choice of the right modeling technique very important [17]. To this end, different modeling techniques can be used. For example, Jiao et al. [20] used a simple multiple linear regression model to estimate the concentration of thiophanate-methyl using UV-VIS spectroscopy [20]. Other authors have used more complicated modeling techniques such as PCA (Principal Component Analysis) and PLS (Partial Least Square) regression using UV-VIS spectroscopy only [21,22]. Besides UV-VIS, other spectroscopic methods, such as fluorescence spectroscopy and near infrared spectroscopy, have also been employed [17]. The combined use of spectroscopy and other wastewater parameters is investigated less. For example, Lv et al. [23] used multiple linear regression to estimate the toxicity of pharmaceutical wastewater using a combination of volatile fatty acids concentrations and UV-VIS spectroscopy [23].

The aim of this paper was to estimate the residual AOX concentration (soluble and total AOX) via modeling based on the following experimental wastewater parameters: pH, chloride concentration, Water-Soluble Organic Carbon concentration (WSOC), UV-VIS spectrum, turbidity, and Solids Removable by Filtration (SRF). A dataset was obtained by applying ozonation to transform AOX compounds into biodegradable compounds. The AOX concentration was analyzed at various time points during the ozonation process. From this dataset, the wastewater parameters suitable for model building were selected using Pearson and Spearman correlations. Modeling was done by performing stepwise multiple linear regression using forward selection, eventually obtaining a multiple linear regression model.

2. Materials and Methods

2.1. Experimental Set-Up and Wastewater Sampling

The wastewater stream investigated in this particular study originates from a chemical production process. The AOX present in this waste stream can be divided into two categories: particulate AOX and soluble AOX. The particulate AOX is removed using coagulation-flocculation, followed by flotation. When the treatment process is executed above its hydraulic capacity, some of the particulate AOX will not be removed, and hence will be present in the wastewater treatment influent, causing problems in the wastewater treatment process (the presence of soluble AOX is not an issue in this specific case). A possible solution, which enables the removal of the particulate AOX before it ends up in the wastewater influent, is the installation of an ozonation unit. Therefore, a pilot ozonation unit was

placed on the plant site and ozonation was performed on effluent from the coagulation-flocculation step. Because of the fluctuating characteristics of this effluent, samples from the effluent tank were taken at various time points. This way, a comprehensive dataset could be obtained, containing data points resulting from ozonation experiments on different effluents, representing the real fluctuation in characteristics of the effluent. The ozonation was carried out on wastewater streams containing both particulate and soluble AOX. To ensure the discharge limit for total AOX (particulate and soluble AOX) is not exceeded, both parameters will be taken into account in this work. To differentiate between the soluble and particulate AOX content in the wastewater, the AOX concentration in filtered and unfiltered wastewater samples is determined. This results in the soluble AOX fraction and the sum of soluble and particulate AOX content (total AOX).

The ozonation experiments were conducted using a containerized pilot plant provided by Air-Liquide. This pilot plant consisted of an ozone generator, ozone gas analyzers, an ozone destructor, and a stainless steel reaction vessel of 250 L. The ozone generator was fed with high-grade oxygen gas and had a maximal capacity of 450 g O₃/h (5.3 w%). The ozone uptake of the reaction mixture was monitored using an ozone gas analyzer at both the inlet and the off gas. The off gas was sent through an ozone destructor unit to prevent discharge of ozone to the environment. The ozonation process was executed in batch mode. For each experiment, the reactor was filled with the wastewater effluent as previously described. During the ozonation process, wastewater samples were manually taken from the reaction tank at defined time intervals and analyzed as further specified. Because the aim is to obtain a diverse set of data points for developing and training an accurate and predictive model, the sampling time or method was varied throughout the experiments. In total, nine experiments were conducted (indicated as “Batch” in Table S1, corresponding samples for each experiment are indicated as “Sample” in Table S1). Depending on the sampling time, the ozone dose ranged from 0 to 4161 g O₃/m³.

2.2. Analysis Methods

Several conventional wastewater parameters were determined conform their corresponding DIN/EN/ISO norms: pH (DIN EN ISO 10523), soluble and total AOX (DIN EN ISO 9562), WSOC (DIN EN 1484), SRF (DIN EN 872), chloride concentration (DIN ISO 15923-1), and turbidity (DIN EN ISO 7027). The UV-VIS spectrum was determined using a Specord 200 photometer from Analytik Jena AG (Jena, Germany).

2.3. Dataset

The wastewater parameters used to estimate the residual total and soluble AOX concentrations are pH, chloride concentration, WSOC, UV-VIS spectrum, turbidity, and SRF. A screening of the UV-VIS spectra was carried out to make an appropriate selection of wavelengths to prevent multicollinearity. Each sample was diluted 25 times with ultrapure water to obtain a UV-VIS spectrum below the absorbance limit. It was found that for all samples, the UV-VIS spectra followed a similar trend: for every sample, a clear peak with a maximum at a wavelength of 220–230 nm was observed. An example of a UV-VIS spectrum is shown in Figure 1. Because of the presence of an obvious peak, it was chosen to divide the UV-VIS spectra in three regions: (i) 200–210 nm: in front of the peak; (ii) 220–230 nm: at the peak maximum; and (iii) 240–250 nm: after the peak. These regions are indicated in Figure 1. Because the peak maximum was situated at 227 nm, the absorbance at this wavelength was selected as one of the independent variables. In order to reduce the possible appearance of multicollinearity, wavelengths in the other two regions were chosen based on their distance relative to 227 nm. In this way, the following parameters were used in the model: absorbance at 200 nm, absorbance at 227 nm, absorbance at 250 nm, turbidity, SRF, pH, WSOC, and chloride ion concentration. The used dataset is shown in Table S1.

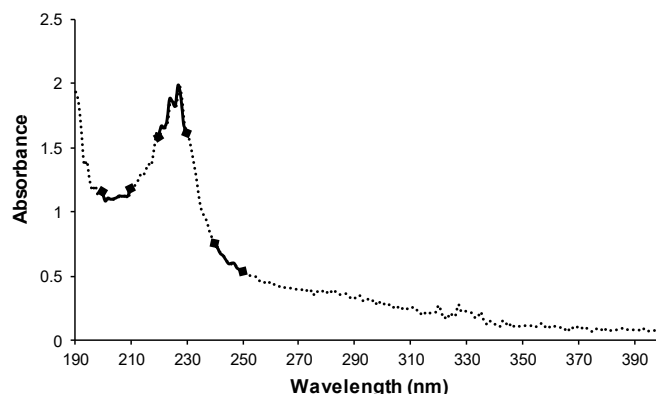


Figure 1. Typical UV-VIS spectrum (1/25 dilution).

2.4. Model Development and Validation

The independent predictor variables were obtained by calculating both Pearson and Spearman correlations for every variable in the dataset with the dependent variables (total AOX and soluble AOX concentration). In this way, the relationship between the predictor variables and the dependent variable was elucidated. If a significant correlation was found, this independent variable was used to build the model. Thereafter, the models were obtained using stepwise multiple linear regression with forward selection. This means that the independent predictor variables were sequentially added to the model, and with each additional term, the model was tested both on the significance of each term ($p < 0.05$) and on meeting the assumptions of multiple linear regression (normal distribution of residuals, homoscedasticity, no multicollinearity, and linear relationship between dependent and independent variables). The improvement of the model was monitored by calculating R^2 and adjusted R^2 values for each model, as well as the Akaike Information Criterion (AIC). It was chosen to exclude interaction terms in the model, because of the absence of a specific intuitive scientific reason to do so [24].

After obtaining the model, 10-fold cross-validation was used to validate the model. In this type of validation, the dataset is divided in 10 sections. One of the 10 sections is chosen as “external” data section, and the other nine sections are used as training data to obtain the model. The obtained model is then validated using the external data, and the differences between the fitted and the real values are calculated. This is done 10 times, using a different section as external data set every time. All statistical analyses were performed in the statistical software RStudio [25].

3. Results and Discussion

3.1. Dataset Exploration

The obtained Pearson and Spearman correlations for each variable in the dataset with both the total and soluble AOX concentration are shown in Tables 1 and 2, respectively. These correlations indicate the association between the dependent variable and each independent variable. To perform this calculation, the null hypothesis assumes no association between these variables (both Spearman and Pearson correlations are equal to 0). This way, correlation values with a significant p -value ($p < 0.05$) imply a correlation of the predictor variable with the dependent variable different from 0. In Tables 1 and 2, insignificant ($p > 0.05$) correlations are marked in grey. A significant correlation was found between turbidity; SRF; absorbance at 200, 227, and 250 nm; pH, and total AOX concentration. Hence, these variables were used for model building. For the soluble AOX concentration, only the WSOC was found to be insignificant. Therefore, turbidity, SRF, absorbance at 200, 227, and 250 nm; pH, and chloride concentration were used for model building.

Table 1. Spearman and Pearson correlations for the correlation of the independent variables with the total adsorbable organic halides (AOX) concentration.

	Total AOX						
	Turbidity	SRF	Abs. (200 nm)	Abs. (227 nm)	Abs. (250 nm)	Chlorides pH	WSOC
Spearman's correlation (ρ)	0.815	0.668	−0.416	0.837	0.712	0.151	0.715
Significance	0.000	0.000	0.000	0.000	0.000	0.167	0.000
Pearson's correlation (r)	0.809	0.714	−0.257	0.778	0.640	0.261	0.699
Significance	0.000	0.000	0.018	0.000	0.000	0.016	0.000

The gray numbers indicate the insignificant correlations.

Table 2. Spearman and Pearson correlations for the correlation of the independent variables with soluble AOX concentration.

	Soluble AOX						
	Turbidity	SRF	Abs. (200 nm)	Abs. (227 nm)	Abs. (250 nm)	Chlorides pH	WSOC
Spearman's correlation (ρ)	0.536	0.385	−0.462	0.878	0.472	0.366	0.496
Significance	0.000	0.001	0.000	0.000	0.000	0.001	0.000
Pearson's correlation (r)	0.547	0.364	−0.423	0.838	0.441	0.512	0.493
Significance	0.000	0.002	0.000	0.000	0.000	0.000	0.000

The gray numbers indicate the insignificant correlations.

3.2. Obtaining the Model

Following the Pearson and Spearman correlations, the contribution of each predictor variable was tested for its significance in the predicting model. During the process of stepwise linear regression, SRF, absorbance at 250 nm, and pH were found to not have made a significant contribution and were hence removed from the model for total AOX.

The full model equation obtained for the estimation of the total AOX concentration is depicted in Equation (1). This model has an R^2 of 0.926 and an adjusted R^2 of 0.921. The dependent variable (total AOX concentration) is log-transformed to meet the homoscedasticity assumption of the model. As shown in Equation (1), the turbidity and the absorbances at 200 and 227 nm were found to contribute significantly to the estimation of the total AOX concentration. The absorbance at 227 nm was found to be significant in both the linear and the quadratic term. To prevent multicollinearity, this variable was mean-centered. Equation (1) can be transformed into a more useful equation, without centered variables and log-transformed dependent variable as shown in Equation (2).

$$\ln([\text{total AOX}]) = 8.735236 + 2.932502 \times 10^{-6} \times (\text{Turbidity})^2 + 0.8033485 \times \text{Abs.}(227 \text{ nm}), \text{centered} - 0.1679748 \times (\text{Abs.}(227 \text{ nm}), \text{centered})^2 - 0.4910289 \times (\text{Abs.}(200 \text{ nm}))^2 \quad (1)$$

$$[\text{total AOX}] \left(\frac{\mu\text{g}}{\text{L}} \right) = e^{6.362228} \times e^{2.932502 \times 10^{-6} \times (\text{Turbidity})^2} \times e^{2.002824 \times \text{Abs.}(227 \text{ nm})} \times e^{-0.3008305 \times (\text{Abs.}(227 \text{ nm}))^2} \times e^{-0.4910289 \times (\text{Abs.}(200 \text{ nm}))^2} \quad (2)$$

The assumptions of multiple linear regression for this model are met. Figures 2 and 3 show the QQ-plot for the distribution of the residuals and the plot of the residuals as a function of the fitted values to test homoscedasticity, respectively. In a QQ-plot, the normality of the residuals is tested by plotting these residuals against the theoretical quantiles calculated from a normal distribution. If the residuals are normally distributed, this should result in the plot of a straight line (which is the case as seen in Figure 2). Figure 3 tests for homoscedasticity: the assumption that the variance of the residuals is independent from the dependent variable. If the homoscedasticity assumption is met, the residuals are equally distributed over the range of fitted values and no trend can be observed (the residuals do not have a specific shape). This can be seen more easily by the red line plotted in the figure, which is a LOWESS (LOcally WEighted Scatterplot Smoothing) curve. A LOWESS curve fits a smooth curve to the data (using nonparametric strategies) to give an impression of the distribution of the residuals. Ideally, when the residuals are equally distributed, the LOWESS curve should follow a horizontal

line at a residuals value of zero (indicated by the dotted line). If this is the case, the positive and negative residuals balance each other out and thus are equally distributed. As can be seen in Figure 3, the LOWESS curve is quite smooth, with a small drop at the end. This small drop can be explained by the very small number of data points in that region and the presence of a larger residual (indicated as observation 66 on the plot). Hence, the overall conclusion can be made that the homoscedasticity assumption is met. Multicollinearity was tested by calculating the Variance Inflation Factors (VIFs) for every variable. Table 3 shows the obtained VIFs for the model. The threshold for multicollinearity was set at $VIF > 2$.

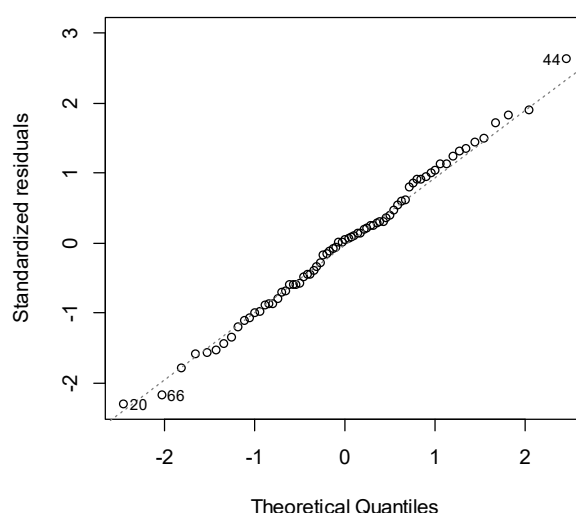


Figure 2. QQ-plot for the obtained model: total AOX.

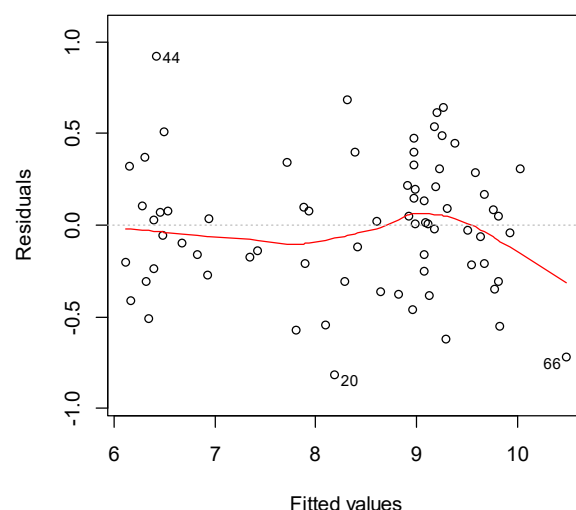


Figure 3. Residuals vs. fitted values for the obtained model: total AOX.

Table 3. Variance Inflation Factors (VIFs) for the obtained model: total AOX.

Independent Variable	VIF
Abs. (227 nm), centered	1.7
(Abs. (227 nm), centered) ²	1.1
(Abs. (200 nm)) ²	1.1
Turbidity ²	1.5

Similarly, turbidity; SRF; absorbance at 200, 227, and 250 nm; pH; and chloride concentration were used to obtain the model to predict the soluble AOX concentration. The modeling process identified

a significant contribution of only the absorbance at 227 nm and the absorbance at 200 nm. Again, the absorbance at 227 nm was found to contribute to the model in both the linear and the quadratic term, and so this variable was mean-centered. To prevent homoscedasticity, the dependent variable (soluble AOX concentration) was log-transformed.

The full model equation for the estimation of the soluble AOX concentration is depicted in Equation (3). This model has an R^2 of 0.919 and an adjusted R^2 of 0.916. For the obtained model, the values of R^2 and adjusted R^2 are approximately equal, which indicates that the data fit the model well and all included predictor variables have a substantial influence on the dependent variable.

$$\ln([\text{soluble AOX}]) = 8.5337786 + 0.8708687 \times \text{Abs.}(227 \text{ nm}), \text{centered} - 0.2289705 \times (\text{Abs.}(227 \text{ nm}), \text{centered})^2 - 0.4530129 \times (\text{Abs.}(200 \text{ nm}))^2 \quad (3)$$

This results in Equation (4) for uncentered predictor variables and [soluble AOX] instead of $\ln([\text{soluble AOX}])$.

$$[\text{soluble AOX}] \left(\frac{\mu\text{g}}{\text{L}} \right) = e^{5.873231} \times e^{2.366463 \times \text{Abs.}(227 \text{ nm})} \times e^{-0.401316 \times (\text{Abs.}(227 \text{ nm}))^2} \times e^{-0.4530129 \times (\text{Abs.}(200 \text{ nm}))^2} \quad (4)$$

Checking the assumptions led to the results shown in Figures 4 and 5 and Table 4 for testing the normal distribution of residuals, homoscedasticity, and multicollinearity, respectively. As can be seen in Figure 4, the data in the QQ-plot for the model for the soluble AOX concentration slightly differ from the desired straight line (especially in the upper right-hand corner). However, execution of the Shapiro–Wilk’s normality test confirmed the normality of the residuals of the model.

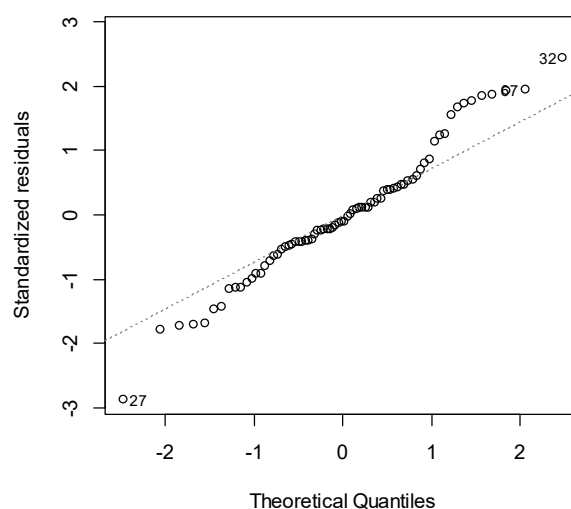


Figure 4. QQ-plot for the obtained model: soluble AOX.

Table 4. VIFs for the obtained model: soluble AOX.

Independent Variable	VIF
Abs. (227 nm), centered	1.2
(Abs. (227 nm), centered) ²	1.1
(Abs. (200 nm)) ²	1.1

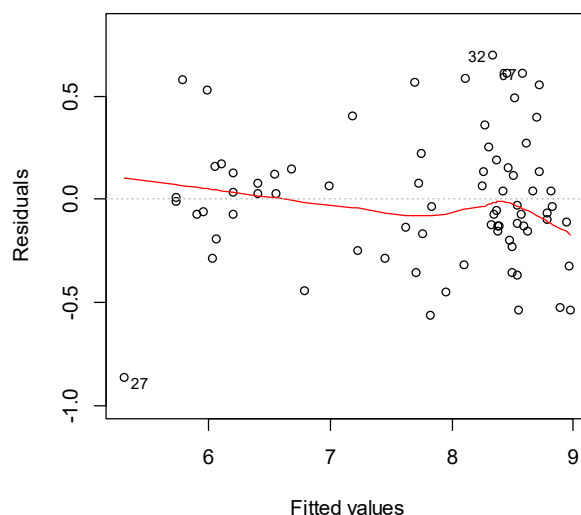


Figure 5. Residuals vs. fitted values for the obtained model: soluble AOX.

3.3. Validation

3.3.1. Confidence and Prediction Intervals

To evaluate the accuracy of the model, the prediction and confidence intervals were calculated. The 95% confidence interval attributes for the interval where, at specific values of the independent variables, the chance of obtaining the mean value is 95%. The prediction interval on the other hand, attributes for where an external observation is situated, considering the variability of the data matrix. There is a 95% chance that this data point is situated in this interval. Both the confidence and prediction interval per observation are shown in Figure 6 for the model predicting the total AOX concentration.

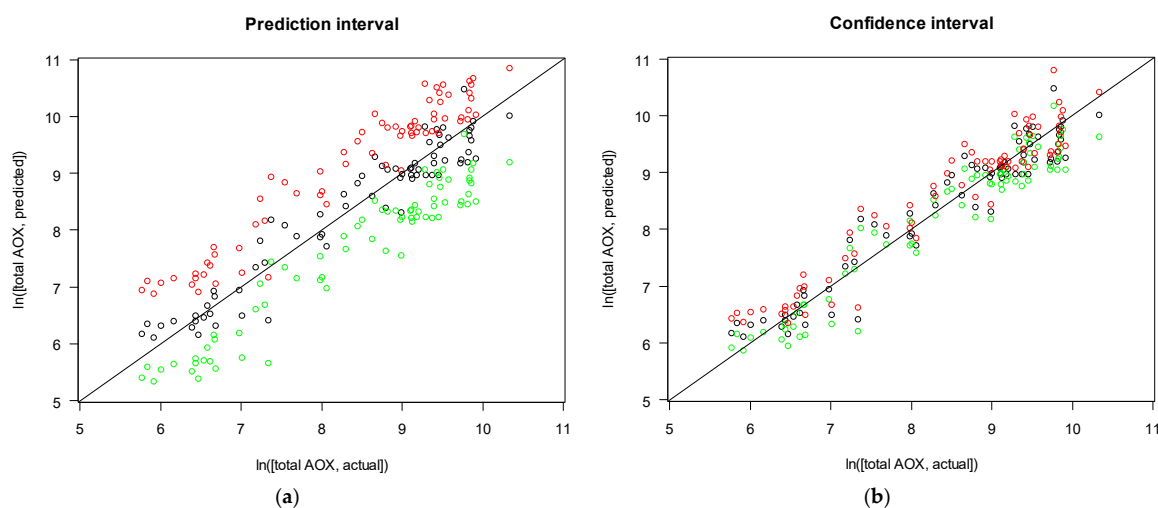


Figure 6. Prediction interval (a) and confidence interval (b) per observation for the model predicting the total AOX concentration (minimum points are indicated in green, maximum points in red, and fitted points in black).

This figure shows a clear correlation between the predicted $\ln([total\ AOX])$ and the actual $\ln([total\ AOX])$ values. Only 5 out of 73 observations are situated outside $\pm 7.5\%$ of the predicted total AOX values. In Figure 6, it is clear that both the confidence and the prediction intervals change for every observation. This is due to the difference in the values of the different predictor variables. Therefore, both intervals (confidence and prediction) were calculated for each individual parameter. In this calculation, the intervals were calculated for a change in the respective predictor variable,

while the other predictor variables remained constant (mean of the predictor variable). To have a good understanding of the intervals, the obtained values were back-transformed, so the real absorbance and total AOX concentration values were obtained. The results are shown in Figure 7. As shown in this figure, the performance of the model decreases, depending on the value of the predictor variables. In the first plot, showing the influence of the absorbance at 227 nm, it is clear that the width of the confidence interval increases rapidly at an absorbance > 2.5 . In this dataset, the percentage of values of absorbance > 2.5 is 5.5%, so the probability of having a data point above 2.5 is low. Next to this, this plot is made for mean values of turbidity and absorbance at 200 nm, and this will not always be the case for observations with a high absorbance at 227 nm (similar conclusions can be drawn for all three plots). In the second plot, showing the influence of the turbidity, it is clear that the confidence interval increases rapidly (deviation of 20% from the fitted line) at a turbidity > 601 FNU. In this dataset, the percentage of values of turbidity > 601 FNU is 5.5%, so the probability of having a data point above 601 FNU is low and the conclusion is similar to that of the absorbance at 227 nm. In case of the absorbance at 200 nm, the confidence interval increases (deviation of 20% from the fitted line) at absorbances lower than 0.675 and higher than 1.4. The percentage of values of the dataset in this range is 8.2%, so the same conclusion can be made.

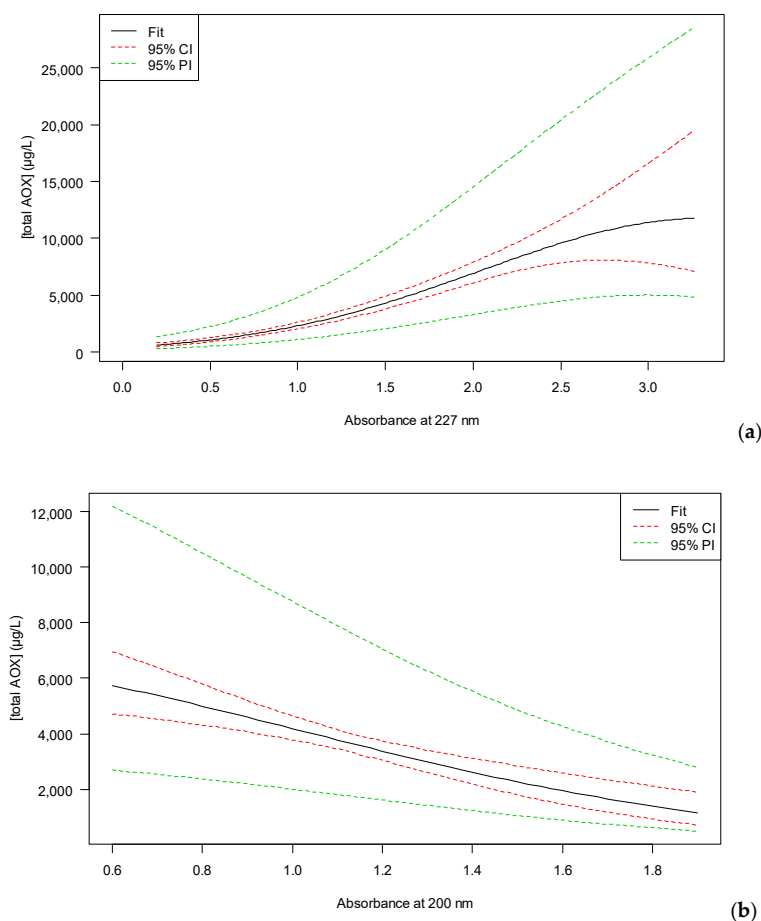


Figure 7. Cont.

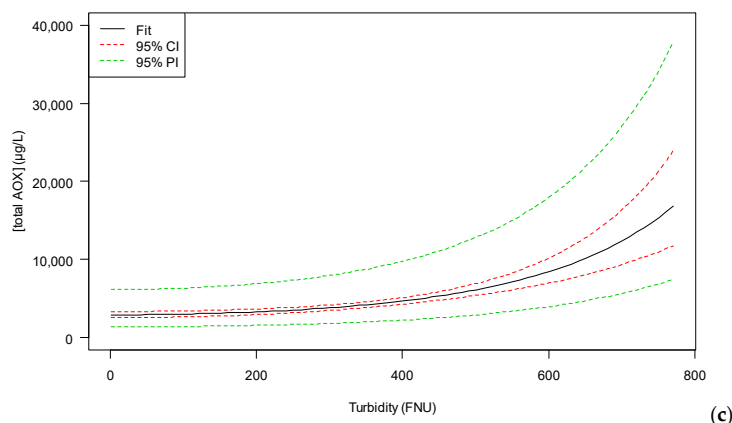


Figure 7. Confidence (CI) and prediction intervals (PI) for each individual predictor variable for the model predicting the total AOX concentration: (a) absorbance at 227 nm; (b) absorbance at 200 nm and (c) turbidity.

Figure 8 shows the prediction and confidence interval per observation for the model obtained for predicting the soluble AOX concentration.

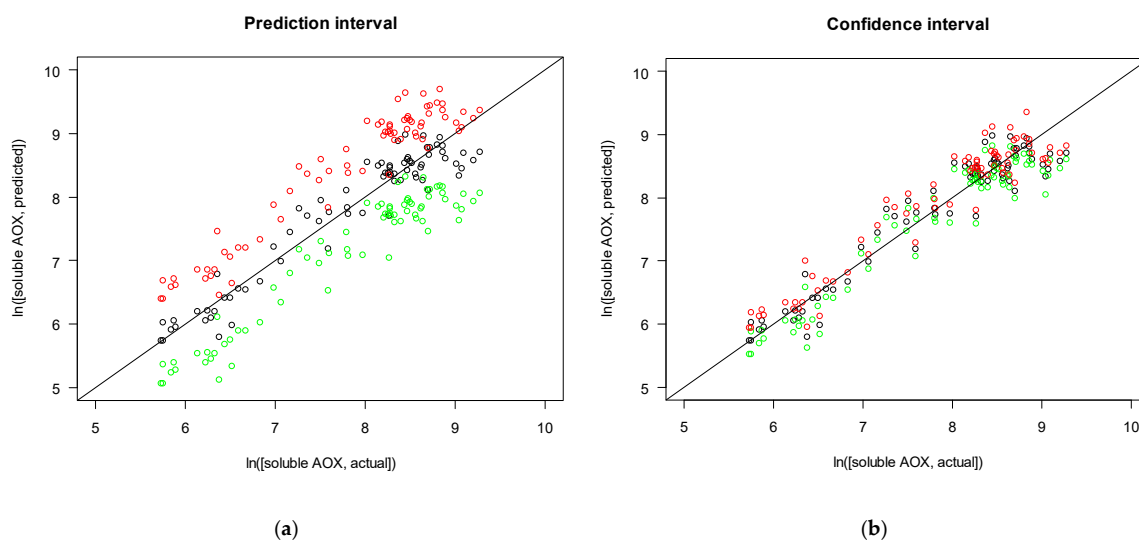


Figure 8. Prediction interval (a) and confidence interval (b) per observation for the model predicting the soluble AOX concentration (minimum points are indicated in green, maximum points in red, and fitted points in black).

These figures show a clear correlation between the predicted $\ln([\text{soluble AOX}])$ and the actual $\ln([\text{soluble AOX}])$ values, which indicates a good model fit. Only 4 out of 76 observations are outside $\pm 7.5\%$ of the predicted soluble AOX values. Similarly, as previously specified, both intervals (confidence and prediction) can be calculated for the prediction of the soluble AOX concentration, for each individual parameter. The results are shown in Figure 9. In the first plot, showing the influence of the absorbance at 227 nm, it is clear that the confidence interval increases rapidly at an absorbance > 2.5 . The percentage of values of absorbance > 2.5 in this dataset is 5.3%, so the probability of having a data point above 2.5 is low. In case of the absorbance at 200 nm, the confidence interval increases (deviation of 20% from the fitted line) at an absorbance > 1.525 . The percentage of values of this dataset in this range is 2.6%, so the same conclusion can be drawn as above.

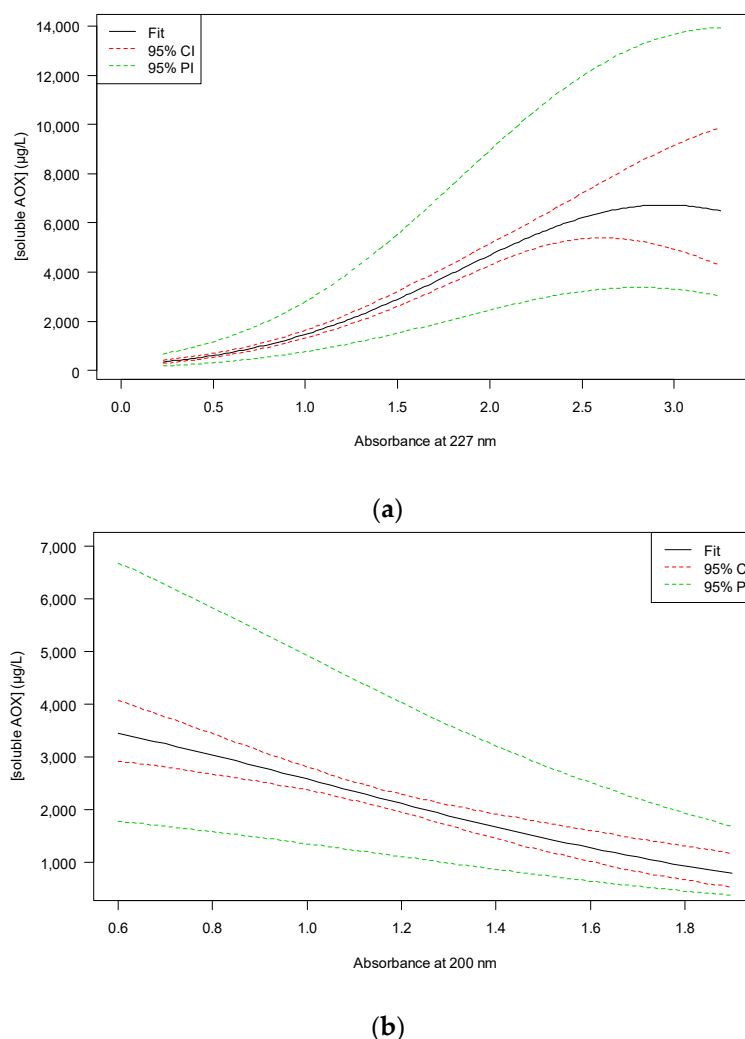


Figure 9. Confidence (CI) and prediction intervals (PI) for each individual predictor variable for the model predicting the soluble AOX concentration: (a) absorbance at 200 nm and (b) absorbance at 200 nm.

3.3.2. Cross-Validation

10-fold cross-validation was carried out to assess the performance of the model. The results are reported as the mean square error between the fitted and the real values. For the model obtained for the total AOX concentration, a mean squared error of 0.146 (log-transformed) was obtained. To visualize this error, the fit $\pm$ mean squared error (back transformed) can be added to the plots showing the confidence and prediction intervals. This is shown in Figure S1. In these plots, the fitted interval for cross-validation lies close to the 95% confidence interval of the plots.

Cross-validation resulted in a mean squared error of 0.114 (log-transformed) for the soluble AOX concentration. To visualize this error, the fit $\pm$ mean squared error (back transformed) is added to the plots showing the 95% confidence and prediction intervals (Figure S2). A similar conclusion can be made: the cross-validation interval lies close to the 95% confidence intervals.

3.4. Influence of Predictor Variables

The influence of the predictor variables on the model predicting the total AOX concentration can be assessed based on their individual contribution to R^2 and adjusted R^2 . These contributions were obtained by calculating the R^2 and adjusted R^2 of the model, leaving one predictor variable out of the model. The obtained R^2 and adjusted R^2 were then subtracted from the R^2 and adjusted

R^2 obtained with the full model. This was done for all predictor variables. The result is shown in Figure 10. The influence of the variables can be ranked as follows (from high to low): Abs. (227 nm)–Turbidity²–(Abs. (200 nm))²–(Abs. (227 nm))².

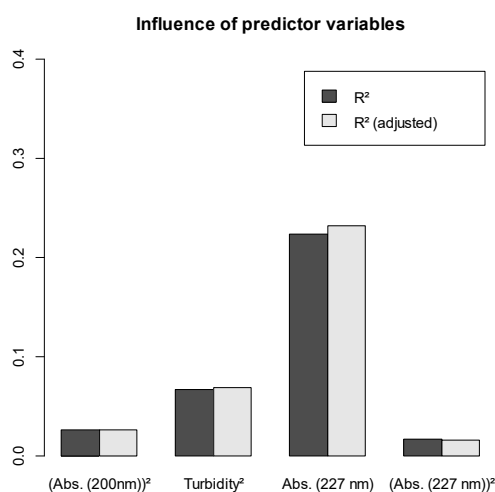


Figure 10. Influence of predictor variables: total AOX.

The influence of the predictor variables for the model predicting the soluble AOX concentration was assessed similarly and is shown in Figure 11. The influence of the variables can be ranked as follows (from high to low): Abs. (227 nm)–(Abs. (227 nm))²–(Abs. (200 nm))².

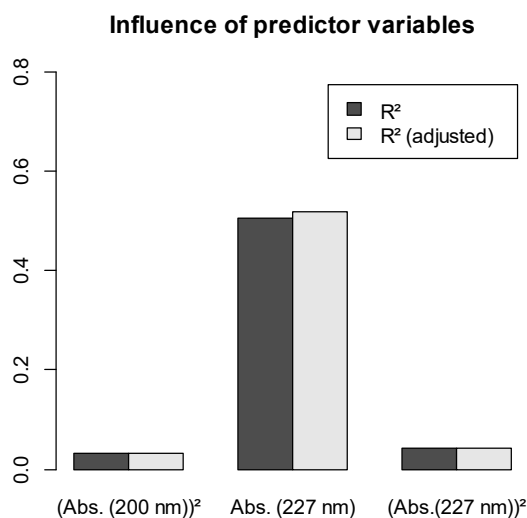


Figure 11. Influence of predictor variables: soluble AOX.

The similarities as well as the differences in influence of the parameters between both models can easily be explained. In both models, the absorbance at 227 nm is the most influential parameter. This wavelength corresponds with the maximum of the absorbance peak of the wastewater samples. For the model of the total AOX, the turbidity is also of importance. This can be explained by the presence of both soluble and particulate AOX in these samples. The presence of the particulate AOX has a big influence on the turbidity of the wastewater, whereas these suspended particles are filtered out in the soluble AOX samples. This explains the presence and absence of these parameters in the model for the total and soluble AOX, respectively. For the absorbance at 200 nm, no direct explanation can be found. During the ozonation process, the organic compounds present in the wastewater stream are oxidized and transformed into other organic compounds, which can readily influence the UV-VIS spectrum and can influence the absorbance at 200 nm.

4. Conclusions

An accurate model for the prediction of the total AOX concentration, meeting all model assumptions, with an R^2 of 0.926 and an adjusted R^2 of 0.921 was developed, containing the UV absorbance at 227 nm, the UV absorbance at 200 nm, and the turbidity of the wastewater. Confidence and prediction intervals were reported. Next to this, cross-validation was carried out to validate the model and the influence of the predictor variables was assessed.

A similar methodology for the prediction of soluble AOX was followed. An accurate model, meeting all assumptions, with an R^2 of 0.919 and an adjusted R^2 of 0.916 was found, containing the absorbance at 227 nm and the absorbance at 200 nm of the wastewater. Similarly, confidence and prediction intervals were reported and cross-validation was carried out. The most influential predictor variables were also defined.

These models can be beneficial for the wastewater treatment process in many ways. In first instance, the sampling frequency necessary for process monitoring can be minimized to the minimal required amount, because in-line monitoring of the UV-VIS spectra and turbidity will be sufficient for the determination of both soluble and total AOX concentrations. Secondly, the fast prediction of the AOX concentration can be useful for the optimization of the ozone dosage in the ozonation step. Currently, ozonation is carried out using a fixed ozone dose and fluctuations in AOX concentrations are resolved by applying an ozone overdose. Working at an ozone overdose can be omitted with the provided in-line monitoring system, because the ozone dosage can more easily be altered by a fast and easy determination of the residual AOX concentrations.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4441/12/6/1653/s1>, Table S1: Used dataset, Figure S1: Visualization of cross validation error for the prediction of the total AOX concentration, Figure S2: Visualization of cross validation error for the prediction of the soluble AOX concentration.

Author Contributions: H.M. conducted the analysis of the data as well as the data modeling, and wrote the paper; R.D. and D.C. contributed to the writing of the paper; and K.V.E., B.B., S.D.L. and D.L. have obtained the experimental data. All authors have read and agreed to the published version of the manuscript.

Funding: H.M. holds a PhD fellowship of the Research Foundation—Flanders (FWO) (11D7418N). The research was partially funded by the Industrial Research Council of KU Leuven (C24E/19/040 and IOFM/16/007).

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

References

1. Xie, Y.; Chen, L.; Liu, R.; Tian, J. Reduction of AOX in Pharmaceutical Wastewater in the Cathode Chamber of Bio-Electrochemical Reactor. *Bioresour. Technol.* **2018**, *265*, 437–442. [\[CrossRef\]](#)
2. Müller, G. Sense or No-Sense of the Sum Parameter for Water Soluble “Adsorbable Organic Halogens” (AOX) and “Absorbed Organic Halogens” (AOX-S18) for the Assessment of Organohalogenes in Sludges and Sediments. *Chemosphere* **2003**, *52*, 371–379. [\[CrossRef\]](#)
3. Farooqi, I.H.; Basheer, F. Treatment of Adsorbable Organic Halide (AOX) from Pulp and Paper Industry Wastewater Using Aerobic Granules in Pilot Scale SBR. *J. Water Process Eng.* **2017**, *19*, 60–66. [\[CrossRef\]](#)
4. Savant, D.V.; Abdul-Rahman, R.; Ranade, D.R. Anaerobic Degradation of Adsorbable Organic Halides (AOX) from Pulp and Paper Industry Wastewater. *Bioresour. Technol.* **2006**, *97*, 1092–1104. [\[CrossRef\]](#)
5. Osman, W.H.W.; Abdullah, S.R.S.; Mohamad, A.B.; Kadhum, A.A.H.; Rahman, R.A. Simultaneous Removal of AOX and COD from Real Recycled Paper Wastewater Using GAC-SBBR. *J. Environ. Manag.* **2013**, *121*, 80–86. [\[CrossRef\]](#)
6. Qin, C.; Liu, B.; Huang, L.; Liang, C.; Gao, C.; Yao, S. Adsorptive Removal of Adsorbable Organic Halogens by Activated Carbon. *R. Soc. Open Sci.* **2018**, *5*, 181507. [\[CrossRef\]](#)
7. Kumar, A.; Dhall, P.; Kumar, R. Biological AOX Removal of Pulp Mill Plant Effluent by *Pseudomonas Aeruginosa*—Bench Study. *J. Environ. Eng. Landsc. Manag.* **2013**, *21*, 296–304. [\[CrossRef\]](#)
8. Tiku, D.K.; Kumar, A.; Chaturvedi, R.; Makhijani, S.D.; Manoharan, A.; Kumar, R. Holistic Bioremediation of Pulp Mill Effluents Using Autochthonous Bacteria. *Int. Biodeterior. Biodegrad.* **2010**, *64*, 173–183. [\[CrossRef\]](#)

9. Raj, A.; Kumar, S.; Haq, I.; Singh, S.K. Bioremediation and Toxicity Reduction in Pulp and Paper Mill Effluent by Newly Isolated Ligninolytic *Paenibacillus* Sp. *Ecol. Eng.* **2014**, *71*, 355–362. [\[CrossRef\]](#)
10. Klüpfel, A.M.; Glauner, T.; Zwiener, C.; Frimmel, F.H. Nanofiltration for Enhanced Removal of Disinfection By-Product (DBP) Precursors in Swimming Pool Water-Retention and Water Quality Estimation. *Water Sci. Technol.* **2011**, *63*, 1716–1725. [\[CrossRef\]](#)
11. Vinder, A.; Simonič, M.; Novak, P.Z. Influence of Surfactants on the Removal of AOX Using Micellar-Enhanced Ultrafiltration. *Int. J. Environ. Res.* **2014**, *8*, 205–212.
12. Olsńska, U.; Markowicz, P. Removal of Adsorbable Organic Halides from Water Containing Bromide Ions by Conventional and Advanced Oxidation. *Ozone Sci. Eng.* **2016**, *38*, 452–464. [\[CrossRef\]](#)
13. Catalkaya, E.C.; Kargi, F. Dehalogenation, Degradation and Mineralization of Diuron by Peroxone (Peroxide/Ozone) Treatment. *J. Environ. Sci. Health Part A Toxic/Hazard. Subst. Environ. Eng.* **2009**, *44*, 630–638. [\[CrossRef\]](#)
14. Luyten, J.; Sniegowski, K.; Van Eyck, K.; Maertens, D.; Timmermans, S.; Liers, S.; Braeken, L. AOX Removal from Industrial Wastewaters Using Advanced Oxidation Processes: Assessment of a Combined Chemical-Biological Oxidation. *Water Sci. Technol.* **2013**, *68*, 2048–2054. [\[CrossRef\]](#)
15. Catalkaya, E.C.; Kargi, F. Color, TOC and AOX Removals from Pulp Mill Effluent by Advanced Oxidation Processes: A Comparative Study. *J. Hazard. Mater.* **2007**, *139*, 244–253. [\[CrossRef\]](#)
16. Chen, B.; Wu, H.; Li, S.F.Y. Development of Variable Pathlength UV-Vis Spectroscopy Combined with Partial-Least-Squares Regression for Wastewater Chemical Oxygen Demand (COD) Monitoring. *Talanta* **2014**, *120*, 325–330. [\[CrossRef\]](#)
17. Platikanov, S.; Rodriguez-Mozaz, S.; Huerta, B.; Barceló, D.; Cros, J.; Batle, M.; Poch, G.; Tauler, R. Chemometrics Quality Assessment of Wastewater Treatment Plant Effluents Using Physicochemical Parameters and UV Absorption Measurements. *J. Environ. Manag.* **2014**, *140*, 33–44. [\[CrossRef\]](#)
18. Li, J.; Tong, Y.; Guan, L.; Wu, S.; Li, D. Optimization of COD Determination by UV-Vis Spectroscopy Using PLS Chemometrics Algorithms. *Optik* **2018**, *174*, 591–599. [\[CrossRef\]](#)
19. Nadiri, A.A.; Shokri, S.; Tsai, F.T.C.; Asghari Moghaddam, A. Prediction of Effluent Quality Parameters of a Wastewater Treatment Plant Using a Supervised Committee Fuzzy Logic Model. *J. Clean. Prod.* **2018**, *180*, 539–549. [\[CrossRef\]](#)
20. Jiao, L.; Dong, D.; Zheng, W.; Zhao, X.; Zhang, S.; Shen, C. Determination of Thiophanate-Methyl Using UV Absorption Spectra Based on Multiple Linear Regression. *Optik* **2014**, *125*, 183–185. [\[CrossRef\]](#)
21. Brito, R.S.; Pinheiro, H.M.; Ferreira, F.; Matos, J.S.; Lourenço, N.D. In Situ UV-Vis Spectroscopy to Estimate COD and TSS in Wastewater Drainage Systems. *Urban Water J.* **2014**, *11*, 261–273. [\[CrossRef\]](#)
22. Lourenço, N.D.; Menezes, J.C.; Pinheiro, H.M.; Diniz, D. Development of PLS Calibration Models from UV-Vis Spectra for TOC Estimation at the Outlet of a Fuel Park Wastewater Treatment Plant. *Environ. Technol.* **2008**, *29*, 891–898. [\[CrossRef\]](#)
23. Lv, L.; Li, W.; Yu, Y.; Meng, L.; Qin, W.; Wu, C. Predicting Acute Toxicity of Traditional Chinese Medicine Wastewater Using UV Absorption and Volatile Fatty Acids as Surrogates. *Chemosphere* **2018**, *194*, 211–219. [\[CrossRef\]](#)
24. Andersen, P.K.; Skovgaard, L.T. *Regression with Linear Predictors*; Springer: New York, NY, USA, 2010; Volume 84.
25. Team, R.C. R: A Language and Environment for Statistical Computing. Available online: <http://finzi.psych.upenn.edu/R/library/dplR/doc/intro-dplR.pdf> (accessed on 8 June 2020).

