Analysis of Kinetic Parameters for the Water Treatment in Tokyo Bay Based on Machine Learning

Chenyang Lin*

Department of Civil and Environmental Engineering, Waseda University, Tokyo, Japan

* Corresponding Author Email: chenyang.lin@unswalumni.com

Abstract. Water contamination is an environmental problem that is being prompt in the global society. Exposure to contaminated water can hurt public health, due to the industrial, biological, and anthropogenic effects. Hence the advanced water treatment method has been raised to solve the contamination problem. Many techniques are being obtained to develop a method in engineering to solve the contamination problem on a site, and machine learning is also a tool to help with a high decision efficiency of the method construction. This study mainly focuses on the application of the MLP regressor and the MLP classifier on the water treatment in Tokyo Bay and the decision-making process during the water treatment construction. It has indicated that in the Tokyo Bay area, the aerobic bioremediation method works well. The Biological Biomimetic Membrane has the best predictability in the MLP classifier modal, and the Aerobic Bioremediation is appropriate for the selected observation sites in Tokyo Bay. Recent tests have shown that the real-time parameters selected in the field and the amount of data extracted have a significant impact on the water treatment method recommendation. In this paper, by applying the techniques of machine learning to specific scenarios, it provides new perspectives on the way of choosing conventional water treatment methods and improves the decision-making efficiency to a certain extent.

Keywords: Water Treatment, Biomimetic Membrane, Machine Learning.

1. Introduction

Due to the increasing attention to water quality issues, water treatment technology has been receiving more and more attention from various countries and research institutions. Pollutants in water have the potential to harm aquatic ecosystems, contaminate sources of drinking water, and reduce biodiversity. Conventional water treatment systems frequently are unable to adequately treat persistent water pollutants, which leads to incomplete mineralization [1]. This can have several negative effects, including the ongoing pollution of water bodies, possible health risks for those who depend on untreated water sources, and the destruction of ecosystems. People's awareness of water treatment technology is growing considering the seriousness of the problems with water quality. To remove obstinate pollutants and protect the well-being of communities and ecosystems, people are actively looking for novel approaches to water treatment and are growing increasingly concerned about the quality of the water they drink.

The water treatment plays a significant role in the bay. This is because, among multiple industrial facilities, the bay is one of the most important parts. At the same time, the bay has a high probability of being contaminated. Hence, the water treatment in the bay has been prompted recently. The largest industrial port in Japan is in Tokyo Bay, which bears a significant amount of maritime traffic and, consequently, generates a multitude of pollution-related issues, including industrial and biological contamination. Therefore, it is crucial to select appropriate water treatment methods for Tokyo Bay. However, considering the multitude of water treatment methods and the complex aquatic environment in the region where Tokyo Bay is located, it is necessary to analyze and predict the types and quantities of pollutants, and ultimately select suitable water treatment methods.

The main methods employed in the treatment of the Bay region primarily include pre-treatment, biological treatment, reverse osmosis, distillation, and other approaches. These methods have been applied in various research studies. According to a study regarding the fluctuation of nutrients, it points out that the reduction of nitrogen and phosphorus in the water is significantly benefited by the
new type of water treatment system installed near Tokyo Bay [1]. Furthermore, the oxygen content in water also has a certain degree of influence on water pollution. Under the action of microorganisms, oxygen can affect the content of nitrogen and phosphorus elements in water. In September, Tokyo Bay experienced a significant hypoxic zone, which can be reduced by decreasing total nitrogen (TN) and total phosphorus (TP) levels, thereby preventing red tides [2]. In addition, Total Organic Carbon (TOC) will also impact the water pollution surrounding Tokyo Bay. The carbon cycling in the water has also been analyzed regarding coastal urbanization. It has been noted that, even from the 1970s to 2011, the carbon input from the river decreased, and the sedimentation rate remained the same [3]. Hence, the optimization of carbon cycling in sedimentation is also emphasized. Removing easily degradable organic carbon is also one of the future tasks in changing carbon cycling in Tokyo Bay. Based on the previous studies, it is recommended to obtain data analysis, especially machine learning, which is an effective tool for facilitating decision-making in engineering. Machine learning models can minimize process interruptions and produce savings in wastewater treatment plants by optimizing operations [4]. Given the abundance of data and trend analysis in previous studies, there has been a scarcity of research utilizing machine learning for predictive analysis of Tokyo Bay. Hence, it can be used for the analysis in Tokyo Bay to generate a hybrid treatment for water pollution.

This article aims to explore how the choice of water treatment techniques in Tokyo Bay is affected by the amount of dissolved oxygen. A thorough analysis that considers the regional distribution of pollution levels is required due to the West Coast's persistently high nutrient content. Furthermore, this research will take a machine learning and data analytic viewpoint. With the assistance of this study, it is aimed to achieve a better ability to predict appropriate water treatment techniques based on the dissolved oxygen content of the water. This paper also attempts to explain the proportionate links among various water treatment techniques.

2. Investigation and Analysis of The Current Situation of Water Pollution and Water Treatment

2.1. Water Contamination Investigation

The dissolved oxygen in water is closely related to pollution, especially organic waste, wastewater discharge, and excessive nutrient sources, which may lead to a decrease in dissolved oxygen in water. The low-oxygen area in the water has a significantly negative impact on the aquatic organisms. Hence the level of DO is usually used for assessing the health of the water quality and providing guidance for future water treatment strategies.

From the theoretical research, the dissolved oxygen in environmental water will generally be maintained at a certain level, which is 10mg/L [5]. The biological effects are caused by low oxygen concentration in water, which will result in a large number of fish dying under this condition, and it will cause certain pollution to the water body.

‘According to the EC, DO in water shouldn't drop below 4 mg/L for cyprinid species and 6 mg/L for salmonid species’ [6]. Therefore, when predicting whether dissolved oxygen will have an impact on the water body in the subsequent analysis, this standard can be considered for future judgment. Furthermore, changes in dissolved oxygen can also cause variations in the metal content of water, thereby affecting water quality.

‘Heavy metals, particularly those attached to organic matter sediments, are released into the overlying water when the dissolved oxygen concentration is low, that is, less than 7 mg/L, and vice versa’ [7]. From the investigations, the main contamination that is made by metal can contain Ni, Cr, Cd, and Pb. These metals not only cause ecological damage but also exacerbate the situation of limited water resources.

According to the study in the Kelantan River Basin, the Dissolved Oxygen (DO) has decreased from 2013 until now and the amount of DO is detected to be 6.1-6.6 mg/L. Dissolved oxygen (DO) is predicted to decline as a result of the rise in suspended sediment content, which is linked to an
increase in manufacturing numbers [8]. The other test in Košice basin displays the findings of a physico-chemical analysis of groundwater observed over one year demonstrating the result that the DO varies from 5.7 to 8.6 mg/L, and the legislative limits should be more than 5mg/L [9]. It was summarized from this study that lower DO levels can result in anaerobic conditions, which are harmful to aquatic organisms.

In addition, a study that has been conducted by Liu et.al was dedicated to using tree-based machine learning for prediction of the water pollution. The DO is one of the indicators that shows the pollution levels in the water and this parameter can be used as input in machine learning. Such prediction experiments typically require setting some conventional values as benchmark references. It was then given that under the temperature of 20 Celsius degree, the concentration of the DO is 9.21 mg/L [5]. It can also be concluded that the decline of the DO will result in contamination including carbon hydrates and organic compounds. At the biological level, different concentrations of dissolved oxygen can also have different ecological impacts. Regarding inorganic substances, metals are a significant influencing factor. The response measures required for these pollutants caused by different levels of dissolved oxygen are also varied.

The next section will introduce some methods and applications that can be used to address and treat pollutants.

2.2. Water Treatment Analysis

Specific water treatment techniques have been designed for the special cases. By analyzing past research on the treatment of different pollutants, dissolved oxygen in the water and other factors in this case, some advice will be given on how to treat the water in the Bay.

The decrease in dissolved oxygen in the water reflects the deterioration of water quality, resulting in the death of fish and other aquatic organisms, and decomposition in the water body, thus further aggravating water pollution. Hence the study has introduced a method for treating the animal death condition, which is to use a biological biomimetic membrane made of aquaporins used in wastewater treatment. ‘Aquaporins provide other roles in the transportation of several other tiny solutes and gasses’ [10]. This allows the technology to selectively filter harmful substances without contamination because proteins can be degraded under certain conditions. Organic matter is one of the main sources of Microbial metabolism in water. Microorganisms consume oxygen in the process of breaking down organic matter. Therefore, by degrading too much organic matter in the water can also improve the water quality problems caused by too low dissolved oxygen. Among them, compounds are typically consumed in water through aerobic degradation. However, during this process, the dissolved oxygen content in the water is inevitably reduced to a certain extent. ‘Several types of aromatic compounds are under aerobic biodegradation, which passes via a crucial intermediate’ [11]. The reduction of oxygen content typically imposes certain limitations on water treatment methods that require a certain level of oxygen content. Hence, it has also been addressed that the dissolved oxygen is required to be at least 2mg/L for the aerobic bioremediation technique being used [12].

Water pollution caused by the pollution of metals, especially some heavy metals, should also be a concern, relevant methods need to be developed to specifically address the issue of metal contamination in water resulting from the detection of low oxygen levels in the water. The presence of oxygen is essential for water quality, and it can combine with metals to form metal compounds, thereby reducing the impact of metallic elements. ‘The heightened reactivity of nanoscale zero-valent iron (nZVI) makes it a material of interest in these domains’ [13]. The previous research has presented that the nZVI can be used for removing heavy metals, including the reduction and oxidation of Cr, As and Cu. ‘The oxidation of the nZVI can be under aerobic, which is 8mg/L and also be anaerobic, which is less than 3mg/L’ [14]. Because of the uncertainties in the aquatic environment, nZVI is generally recommended for use in the aerobic environment, whereas there are many extremes in the actual conditions, the possibility of choosing to use in an oxygen-free environment is therefore not ruled out.
The biological biomimetic membrane has the characteristics of environmental protection, high efficiency, high separation, and high selectivity, but it has higher requirements for dissolved oxygen because of its good Biocompatibility. The biodegradation method has the features of low cost and sustainability. Even though it has certain requirements for the concentration of dissolved oxygen, it is not overly demanding. It still has good versatility. The method of nZVI has a good compatibility to remove the metal in the water and it can also be used for the treatment of organic and inorganic compounds. As an emerging technology, nZVI has the best versatility among the other technologies and is relatively easy to obtain. However, the commonly used pump and treatment method in water treatment plants will also be considered in this study.

3. Methodology

3.1. Site Investigation

In this study, previous investigations regarding water pollution, such as biological contamination and the treatment and the latest data will be integrated to analyze and research the water quality of Tokyo Bay and provide recommendations for water treatment methods [11-13]. The topography and water resource situation of Tokyo Bay will be introduced and real-time and up-to-date data from relevant observation departments will be obtained [13]. From these data, the information on dissolved oxygen, pH value, and other relevant data in the water body of Tokyo Bay will be investigated. The above data source is from the website of the Tokyo Bay Environmental Information Center [13].

According to the previous studies, it can be concluded that in the Tokyo Bay mouth, the tidal currents are in domination. Moreover, in comparison, the amount of seawater flowing into Tokyo Bay from external sources is significantly higher than the amount flowing into the bay from rivers [14]. Various indications suggest that the environment on the eastern side of Tokyo Bay resembles that of a marine environment. Under the influence of seawater, pollutants tend to accumulate more in Tokyo Bay. For example, volatile methylsiloxanes (VMS) are considered to be a common pollutant in water. The distribution pattern of its detection in water can to some extent reflect the relatively concentrated areas of pollutants in Tokyo Bay under the influence of seawater.

3.2. Data Resources

According to 'The Distribution of The VMS in Tokyo Bay and Its Historical Pollution [15]', the Mass loading of the VMSs near the Tokyo harbour is 2500 kg/y. In addition, the VMSs concentration around the Tokyo harbour is also higher than the bay mouth. After the year 1994 to 2013, the pollution depth rose from 0 to 600 cm [15]. Hence, it can be concluded that the degree of pollution in the industrial and residential areas on the inner side of Tokyo Bay is higher and the pollution increased rapidly after year 1994, and the observation site at Urayasu Oki, Kenmigawa Oki, Chiba Port No.1 Light and Kawasaki artificial island should be chosen.

Fig 1. Distribution of Tokyo Bay water quality monitoring stations [13].
According to the Tokyo government's provided Tokyo Bay water quality monitoring sites from Fig. 1. It has shown the location of the sites Urayasu Oki, Kenmigawa Oki, Kawasaki Artificial Island, Chiba Port and the sites above should be mainly considered.

The data of the DO from the Observation Point in Tokyo Bay, which are Urayasu Oki, Kenmigawa Oki, Chiba Port No.1 Light and Kawasaki artificial island [13]. It can be summarized that the DO in each of the observation points is different and has specific features. The port of Urayasu Oki shows that the DO is around 9 mg/L. The Do is 12mg/L at Kenmigawa Oki and the number is fluctuated. The DO at the Kenmigawa Oki has the lowest level when it is in the afternoon considering the production schedule in the industry harbor. The Do is stable at Chiba Port No.1 Light, DO at the upper and middle level is 10mg/L and 6 mg/L at the bottom level. For the Kawasaki artificial island, the DO concentration at 10 mg/L. In addition, it is notable that the DO at the top level of the water is higher than the DO at the lower level [13]. It is worth noting that this study used a relatively limited number of observation sites and site data for analysis. Although there may be some numerical differences in the dissolved oxygen of water bodies in different locations, these values can still be addressed within the same water treatment method. From an economic perspective, the same water treatment method can still be used to solve the issue.

3.3. Method

3.3.1. Construction Model

In this study, the aim is to utilize the commonly used programming language Python, along with the corresponding machine learning library Pandas, to analyze the oxygen content in water. Through this analysis, this paper aims to provide recommendations and predictions for water treatment methods.

<table>
<thead>
<tr>
<th>Treatment Method</th>
<th>Do Level (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nZVI</td>
<td>&lt;3 or &gt;8</td>
</tr>
<tr>
<td>Pump and Treat</td>
<td>-</td>
</tr>
<tr>
<td>Biological Biomimetic Membrane</td>
<td>&gt;2</td>
</tr>
<tr>
<td>Aerobic Bioremediation</td>
<td>4-6</td>
</tr>
</tbody>
</table>

Based on the previous literature reviews and summary, the proper DO level for the treatments has been summarized in Table 1. At different levels of oxygen content, there will be some fluctuations in the water treatment techniques used. In the previous section, there are a total of 4 potential water treatments that can be used to apply in Tokyo Bay including the nZVI, pump and treat, biological biomimetic membrane and aerobic bioremediation. It has been shown that the DO for the aerobic bioremediation is 4-6 mg/L, and the DO for the nZVI can be aerobic and anaerobic, which can be less than 3 mg/L and more than 8mg/L. For the biological biomimetic membrane, the DO level is proper to be more than 2 mg/L. Since traditional pump and treat is a physical treatment method, there are no additional requirements for the level of dissolved oxygen.

3.3.2. Data Processing Procedure

The DO content of the four main observation sites will be extracted where pollutants are believed to be more concentrated. Firstly, the machine learning libraries are imported into the Python environment, which are Pandas, sklearn, MLPClassifier StandardScaler, and finally the result will be opened in an Excel file by the OS function. Pandas are for data processing. It will help with the data import and output. The sklearn is for the model training and data preprocessing. It will provide the realization of the MLP for classification and regression. StandardScaler is a technique to make sure the features such as DO in this case have similar scales. Since the parameter in this study is DO, so the pretreatment procedure is not the most significant step. The data retrieved from Tokyo Bay is on January the 18th, since the data is obtained each hour in 24 hours, the amount of data is enough for the analysis. Firstly, the 4 observation sites are a whole system, and for each of the water treatment
technologies, there is a percentage to describe the probability of obtaining that method, so multiple water treatment methods can be obtained based on the data. The DO data and the four water treatment methods will be entered first, and the intervals for the DO described earlier will be defined, followed by data standardization and data splitting, and then a training model will be constructed, after that, the model is evaluated, the data is predicted, and the results are presented.

During the data processing procedure, MLP artificial neural networks will be used as models. MLP can better fit the relationship between parameter features and targets. It is mainly used to solve the nonlinear problems encountered in this study. MLPRegressor will be used to analyze the percentage of the usage for each treatment and the MLPClassifier has been generated to predict the specific treatment on each site. Analyzing the problems of each water treatment method can be a regression problem while identifying specific types is a classification problem. Therefore, in regression problems, MSE is used to evaluate model performance, while in classification problems, accuracy is a more important factor.

### 3.3.3. Result

#### Table 2. Summarized data including the Mean DO in the observation sites and the percentage of usage for the water treatments.

<table>
<thead>
<tr>
<th>Observation Site</th>
<th>Mean DO</th>
<th>Treatment Method</th>
<th>Percentage of Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urayasu Oki</td>
<td>10.16333333</td>
<td>nZVI</td>
<td>3.11%</td>
</tr>
<tr>
<td>Kenmigawa Oki</td>
<td>14.09041667</td>
<td>Pump and Treat</td>
<td>-1.00%</td>
</tr>
<tr>
<td>Chiba Port No.1 Light</td>
<td>8.985416667</td>
<td>Biological Biomimetic Membrane</td>
<td>91.78%</td>
</tr>
<tr>
<td>Kawasaki artificial island</td>
<td>9.63</td>
<td>Aerobic Bioremediation</td>
<td>5.54%</td>
</tr>
</tbody>
</table>

According to Table 2, the mean DO value for the four observation sites will be entered as 10.16333333, 14.09041667, 8.985416667, 9.63. It can be concluded that the DO at the Kenmigawa Oki has the highest level and the Kawasaki artificial island has the lowest level of the DO. The Kawasaki artificial island in this case should be considered as the observation point with the highest priority. At the same time, the table also shows the proportion that different water treatment solutions may have in the overall water treatment process. Finally, the result is going to be nZVI: 3.11%; pump and treat: -1.00%; biological biomimetic membrane: 91.78%; aerobic bioremediation: 5.54%. This result also fits well with the analysis of the topography of Tokyo Bay. As most of Tokyo Bay is ocean, it is not recommended to use the pump and treat method for large-scale water treatment. As a new emerging technology, nZVI is often limited by its cost and production capacity, and in most cases, it is only applied to more extreme situations, therefore, in overall analysis, the recommended usage proportion of nzvi is relatively low. Aerobic bioremediation is also not as versatile as the biological biomimetic membrane in terms of systematic analysis. However, although some water treatment methods have not been strongly recommended, they still have a certain recommendation rate, so they can be used as auxiliary means.

Hence for the overall treatment for the observation site, the suggested water treatment method is the biological biomimetic membrane, and it can be used with aerobic bioremediation for better performance. Mean-Square Error (MSE) has been used to estimate the performance of this model.
Fig 2. Scatter Plot to compare and analyse the actual values and predicted values of the water treatment method.

Fig. 2 shows that the prediction value for the aerobic bioremediation is higher than the nZVI and the biological biomimetic membrane has the highest actual value, which is 1.0 and the predicted value is 0.4. So, it is predicted to be the best treatment for the overall observation sites. The predicted values for the Biological Biomimetic Membrane and the Aerobic Bioremediation are 0.20 and 0.25 and the actual values are 0. The reason for this situation is likely that, although the study selected 24 hours of real-time monitoring data for machine learning training, the fluctuation of the data is not very high within the sampling interval. Therefore, the results will show a two-level differentiation style, and the unrecommended method will not be adopted by the model.

For the specific observation site, the result might be different from the overall estimation, hence the previous code has obtained MLPRegressor, which can be used to predict the percentage of each water treatment method. The MLPClassifier can be used for the analysis of the classification. This modelling method can be used to predict the specific output treatment methods for each site. The accuracy is a standard to evaluate the performance of this model. The result has been displayed in Excel and the most recommended technique by this code is aerobic bioremediation based on the accuracy standard.

Fig 3. Heat map of the confusion matrix.

The confusion matrix heatmap displayed in Fig. 3 reflects the classification performance of the model on all test sets. It can be observed that there is a larger number of samples for nZVI after testing. Therefore, from this perspective, nZVI is a recommended water treatment method.
4. Discussion

This article uses two different artificial neural network methods of MLP to propose recommendations for water treatment methods in the Tokyo Bay area. However, in the analysis, the results provided by these two methods are like the actual values, while some are different from the initial assumptions. Therefore, by predicting the proportion of each water treatment method and analyzing the water treatment methods for each site through MLP classification, specific water treatment methods can be provided for specific sites. When analyzing different observation sites, due to the fluctuation of the DO values in the data and the inherent differences in the data extracted for different observation sites. The study found that different machine-learning algorithms predicted similar results for these sites. When analyzing the overall water treatment plan, all data is uniformly summarized. Therefore, the result will have better universality. Moreover, since water treatment methods can be combined, the treatment plan has diversity and is more closely related to actual situations.

Due to changes in the extracted data and the fluctuation of the amount of oxygen dissolved, the 2 types of machine learning prediction outcomes for distinct observation points are produced while examining different observation points. The reasons for this result can be analyzed as follows: First, among numerous observation sites, four sites with high concentrations of pollutants were selected for analysis based on the reports of VMSs. However, not enough observation sites were chosen to extract data, so there is still room for improvement in terms of data quantity. In addition, for the analysis of MLP classification, to ensure timeliness, the data from January 18th was selected for this study, and the data was extracted every hour. Although enough data was obtained in this way, the time interval of the data can still be further reduced. In summary, the prediction results generated by MLP classification analysis may be biased towards specific treatment methods such as nZVI and the Aerobic Bioremediation, and may not apply to some other methods, resulting in convergence of results from observation sites. From another perspective, although there are visible significant differences in the DO values of the observation sites, the water quality level of the sites is still located in a safe range. So, the final convergence of results can also be well explained.

To make the results more refined, more data and the selection of more observation sites will help with subsequent research. In addition, when selecting potential water treatment measures, the latest technology can also be combined to develop a more suitable treatment method that meets the needs of society.

5. Conclusion

In this study, an MLP artificial neural network was used to assist in the analysis, based on real-time data provided by the Tokyo Bay Observation Stations, to predict the recommended water treatment methods for Tokyo Bay from both overall and local perspectives. For systematic analysis, MLPRegressor was selected. For each site analysis, MLPClassifier is selected. In the systematic analysis, not only ML was used to predict the recommended water treatment methods for the Tokyo Bay region, but also the recommendation was analysed to rate different water treatment methods, for subsequent potential composite applications. The analysis of different sites can provide theoretical suggestions for specific areas. This study demonstrates the application of MLP in the prediction of water treatment technologies, where the regression analysis demonstrates the percentage of different water treatments recommended to be used and that Biological Biomimetic Membrane is the most appropriate water treatment technology for Tokyo Bay in the overall view in the regression analysis. In classifying analysis, the most appropriate treatment is Aerobic Bioremediation for the observation sites. In addition, the Biological Biomimetic Membrane has the best predictability in the MLP classifier model.

This study provides a new perspective for the engineering strategy and implementation of water treatment, increases the possibility of processing more data, and further increases the accuracy of water treatment method selection. Some limitations can be solved in future research. The first is the
limitation of observation sites. More observation sites will provide more data for better analysis. Another limitation is the limited amount of data that can be extracted from different regions, which prevents further comprehensive analysis from multiple dimensions other than dissolved oxygen. Finally, the monitoring device can be improved to provide data with closer time intervals to make the final prediction results will more accurate.

In future research, more diversified methods can be used to predict water treatment methods, including but not limited to selecting more advanced machine learning models, combining multiple machine learning models, and increasing the sample size to obtain more accurate prediction results.

References


