ICA2025OSLO
14th IWA Specialised Conference on
Instrumentation, Control and Automation

# Application of artificial intelligence with daily operational management data at sewage treatment plants for more energy-saving operations

F. Nishimura*, P. Chun*, and Taira Hidaka*

* Kyoto University, 1-2 Yumihama, Otsu, Shiga, JAPAN, nishimura.fumitake.3n@kyoto-u.ac.jp

Contact information (Fumitake Nishimura, professor, nishimura.fumitake.3n@kyoto-u.ac.jp, 520-0811, +81-77-527-6225) corresponding author

** The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, JAPAN, chun@g.ecc.u-tokyo.ac.jp

*** Kyoto University, C1-2, Kyoto-Daigaku-Katsura, Nishikyo-ku, Kyoto, JAPAN, hidaka.taira.4e@kyoto-u.ac.jp

**Abstract:** Sewerage has been developed with the major objectives of "inundation control," "improvement of public health", and "conservation of water quality in public water areas". In recent years, in addition to these purposes, the practical application of the recovery of resources and energy of sewage such as phosphorus recovery and methane fermentation has been attempted. On the other hand, although the facilities are becoming obsolete, it is difficult to renew the facilities due to the financial situation and the declining state of the population. Therefore, it is required to appropriately manage the existing facilities. In order to achieve these objectives and perform more appropriate value-added operation management, it is necessary to collect the necessary and sufficient information and utilize it. In this study, the possibility of various information that sewerage has is considered. The water quality of effluent from a wastewater treatment plant is predicted by machine learning using the operation & management data acquired every day as an example of information utilization. The ideal ways of sewerage as an information base in the catchment area are considered and discussed as well. As a result, it was found that it is possible to predict the quality of the processed water using deep learning and decision tree methods, and that it is possible to extract specific information on how to manage the operation for energy conservation, using daily operation management data.

**Keywords**: sewerage, daily operational data, machine learning, energy-saving operations

## Introduction

In Although the actual sewage treatment system is relatively simple in terms of its structure, it is difficult to understand the whole system just by accumulating an understanding of the phenomena of each elemental process, because the sewage to be treated and the microbial metabolism that is the key to the treatment reaction are complex systems. On the other hand, it is thought that it is worthwhile to try to understand the response characteristics of a sewage treatment facility from a large amount of input and output information. In addition, if it is possible to understand the response, it is possible to perform appropriate operation management accordingly, and there is the possibility of further improving the current system management, for example from the perspective of saving resources and energy.

In this paper, we first considered the role that sewage systems can play in watershed management, and the further possibilities in the information age to come. As an example of technology that can be used to improve and enhance the operation and management of actual sewage treatment plants, we attempted to apply machine learning technology, which has seen rapid technological progress in recent years. Specifically, we examined the reproducibility of treated water quality by having the machine learn the inflow sewage water quality and quantity, rainfall data, and operating conditions. We considered that there are some water quality items that are highly reproducible and others that are difficult to reproduce, and we discussed the effectiveness of machine learning and further improvements.

## Methods

(1) Targeted sewage treatment system

The sewage treatment plant used as an example is A sewage treatment plant in the Kansai region. Here, the system is a diversion system, and treatment is carried out using the anaerobic anoxic aerobic method and rapid sand filtration. The facility scale is 75,000 m3/day. In addition to organic

matter (BOD, COD), this sewage treatment plant also uses the A2O (Anaerobic Anoxic Oxic) method to reduce nitrogen and phosphorus (TN, TP). The treatment plant also has a sludge treatment system, and the sludge from the primary sedimentation tank and the sludge drawn from the final sedimentation tank are each concentrated, mixed, and then incinerated in an incinerator after undergoing a dehydration process. The water treatment system at this treatment plant uses only the A2O method, and advanced treatment methods have been used since the plant was first constructed. A removal system using coagulant (PAC) is also incorporated for phosphorus. The biological reaction tank section is made up of six tanks arranged in parallel. The entire system includes return water from the sludge treatment system, but in this case the inflow water to the biological reaction tank is treated as the outflow water from the first sedimentation tank, including the return water.

(2) Operational management data and calculation methods used

The data collected includes monthly and daily reports, as well as rainfall data published by the Ministry of Land, Infrastructure, Transport and Tourism, and covers a period of approximately seven years.

**Results and Discussion**

**Figure 1** shows the results of machine learning for some of the treated water quality parameters (for example, results of TN). Here, the system was trained using input data from the six years prior to March 31, 2016, and then used to predict the water quality for the year from April 1, 2016 (FY2016). The graphs show the changes in each water quality parameter over time. The results on the left are the results of deep learning calculations using a neural network with 10 hidden layers and 100 nodes for each, while the results on the right are the results of machine learning using the random forest method. The random forest method is a type of machine learning that can be applied to a wide range of uses, and while it is not deep learning, it is one of the excellent machine learning methods that can produce high-precision results with a very simple program and little parameter tuning. The number of decision trees was set at 300, and the CART method was used for the decision tree partitioning method. No parameter tuning was performed for the deep learning calculation results. For each measurement day, hourly measurement data was treated as the average value for the day of observation, and for weekly measurement data, supplementary values were used. Each measurement day was treated as a single data set and used for calculations. One model each of deep learning and random forest was used to predict the nine target variables.

In the first six years, the calculations were based on the actual measured data used in the training, so the calculation results necessarily matched the actual measured values well, and highly accurate calculations were performed. In the seventh year, data that was not used in the training was input, so the calculation results showed a deviation from the actual measured values compared to the results of the previous six years. For each water quality parameter, the predicted values for water temperature, UV, pH and TP were calculated more accurately, while the calculated results for VIS and SS showed a deviation from the actual measured values(**Figure 2**). The reason for the good prediction results for water temperature is thought to be that the annual change patterns are similar each year. On the other hand, there are a small number of values in the training data that greatly exceed the average value for SS, and it is thought that these had an effect in lowering the prediction accuracy for the verification data. For VIS, the range of the actual measured values for the verification data (0-2.4) is smaller than the range of the actual measured values for the training data (0 -5.3) was smaller than the range of the measured values of the training data (0-2.4), and although the range of the predicted values was within the range of the measured values, it was thought that the correlation coefficient (Pearson's product-moment correlation coefficient) between the measured and predicted values decreased due to the influence of other factors (input data values). In places where the values change significantly over time, even if the result is from the training period, the deviation between the measured values and calculated values is large. This

trend seems to be greater for the Random Forest method than for Deep learning over the entire calculation period, which includes the training and prediction periods (the discrepancy between the black line [calculated values] and red line [measured values] in the figure shows that the red line is drawn behind the black line in many places).

In the results of this calculation, the correlation coefficient between the measured value and the predicted value was higher for the Random Forest method than for the Deep Learning method over the prediction period, and the Random Forest method was able to calculate more accurately. However, since the parameters were not tuned in Deep Learning, there is thought to be room for further improvement in calculation accuracy.

Using the system after machine learning, we simulated the processing characteristics when the operating conditions were changed. Here, as an example, the left diagram shows the change over time in the treated water when the DO in the nitrification tank was changed by 0.5 mg/L in the range of -2.0 to +2.0 mg/L compared to the actual operating value, and the right diagram shows the average difference from the actual observed value for each DO control value condition(**Fig. 3**). The lower the DO control value is set below the actual control value, the more nitrogen removal progresses, and conversely, when the DO control value is set higher than the current value, the total nitrogen concentration increases, indicating that denitrification is not progressing as a whole. It has been shown that nitrogen removal can be promoted by lowering the DO value below the current value. Since nitrogen removal progresses by reducing the amount of aeration, it has been shown that more energy-efficient operation can be achieved by applying machine learning technology(**Figure 4**).

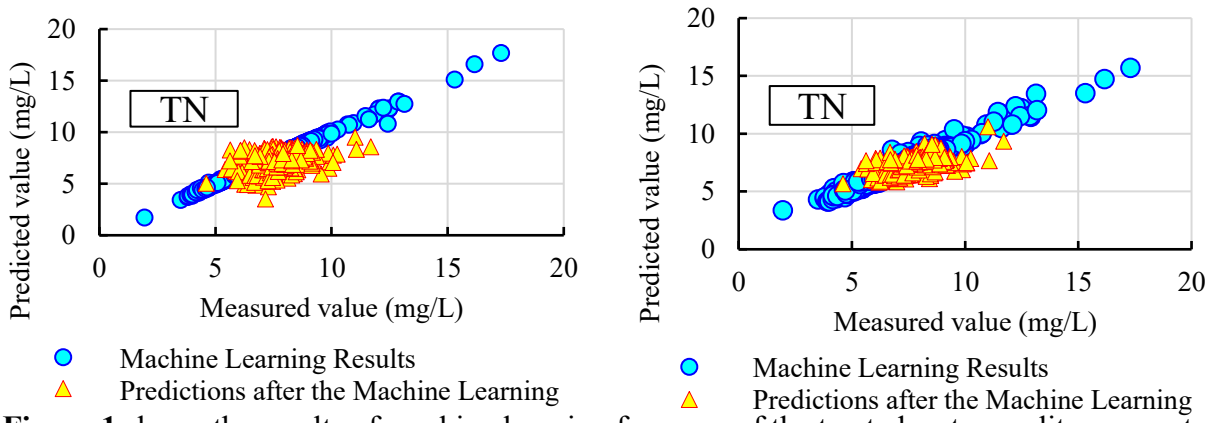


**Figure 1** shows the results of machine learning for some of the treated water quality parameters (for example, results of TN)

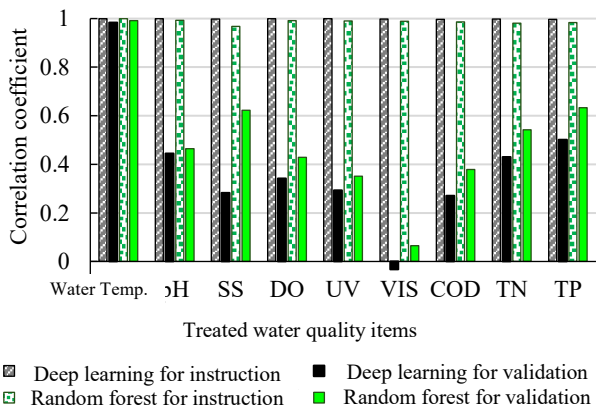


**Figure 2.1** Correlation coefficient between actual and predicted values
Prediction of the situation for 2016 using data from the six years from 2010 to 2015

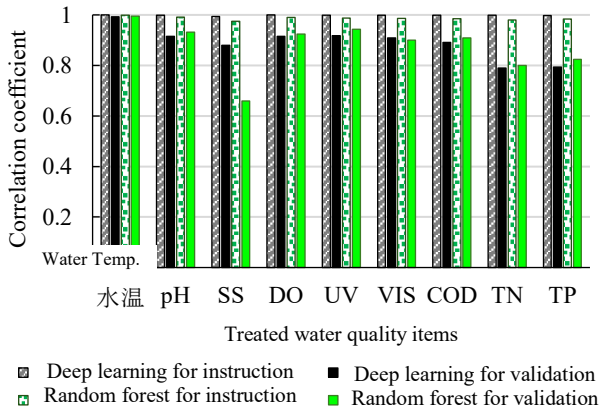


**Figure 2.2** Correlation coefficient between actual and predicted values
70% of all data are used for machine learning, and use the remaining 30% for predictions

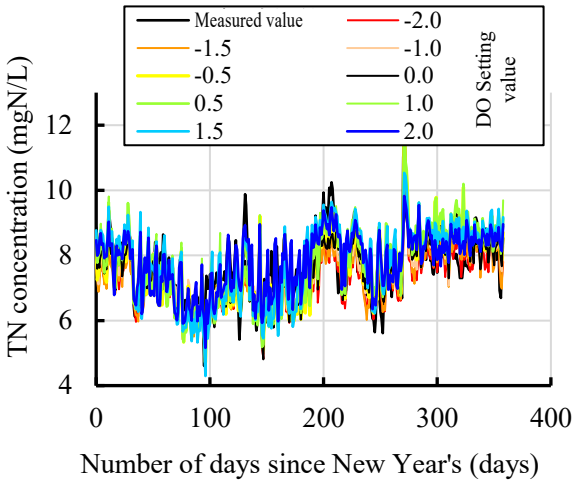


**Figure 3** Predicted value of the TN concentration of the processed water when the set DO value is changed

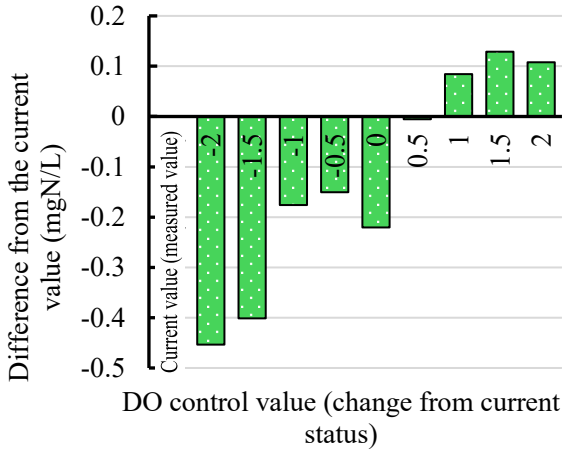


**Figure 4** Average change in treated water TN when the set DO value is changed

## Conclusions

Based on 7 years of daily and monthly data collected from a real sewage treatment plant, we constructed a detailed deep learning model and calculated the output data (predicted data) using the inflow sewage characteristics (water volume, water quality) and operational conditions at the sewage treatment plant as input conditions. We then used the Random Forest method, which is a type of machine learning, to calculate the data. We confirmed that both models could predict the training data with a high degree of accuracy, and then used the learned models to predict the verification data. Depending on the water quality item, the results showed that the accuracy was relatively high for water temperature, UV, pH and TP, but there was room for improvement for other items. It was also found that there was room for parameter tuning in the deep learning model, and that there was potential for further improvements in prediction accuracy, such as incorporating observation results from several days ago in a time series. It was also found that lowering the DO set value from the current level could promote denitrification by 0.4 to 0.5 mgN/L.

## Significance and impact of the presentation on this topic of the conference

A huge amount of driving data is collected on a daily basis in Japan, but it is only used to check compliance with standards and is not put to any other use. This research shows an example of how to use this data for better driving management, and we can discuss the use of monitoring data for a wide range of different purposes at this conference.

## REFERENCES

Fumitake NISHIMURA, Pang-jo CHUN, Yoshifumi FUJIMORI, et al. 2020 Sewerage as an information base in urban and basin areas -attempts to utilize operation management data-. *Artificial Intelligence and Data Science.* **1**(J1), 278-285

T.K. Ho 1995 Random decision forests, *ICDAR*. **1**, 278-282

Y. Amit and D. Geman 1997 Shape Quantization and Recognition with Randomized Trees, *Neural Computation*, **9**(7), 1545-1588