Journal of Water Process Engineering 16 (2017) 283–289

Contents lists available at ScienceDirect

Journal of Water Process Engineering

journal homepage: www.elsevier.com/locate/jwpe

Predicting the effluent quality of an industrial wastewater treatment plant by way of optical monitoring

Jani Tomperi [a,*], Elisa Koivuranta [b], Kauko Leiviskä [a]

[a] Control Engineering, University of Oulu, Oulu, Finland
[b] Fibre and Particle Engineering, University of Oulu, Oulu, Finland

ARTICLE INFO

Article history:
Received 9 August 2016
Accepted 21 February 2017

Keywords:
Activated sludge process
Cross-validation
Digital imaging
Modelling
Pulp and paper industry
Variable selection

ABSTRACT

Wastewater samples taken from the aeration tank of a full-scale activated sludge plant were analyzed using an automatic optical monitoring device. Five variable selection methods were utilized to find the optimal subsets of input variables to develop predictive models for the important parameters of the wastewater treatment process efficiency and the quality of the effluent, including suspended solids, biochemical oxygen demand, chemical oxygen demand, total nitrogen and total phosphorus. The dependencies between the selected variables were also inspected. The study showed that the models based solely on the optical monitoring variables can be used to predict the level of the effluent quality parameters hours before the traditional sampling and analyses. Thus, predictive modelling based on the optical monitoring variables is a potential tool to be used assistance in a process control, keeping the process in a stable operating condition and avoiding environmental risks and economic losses.

1. Introduction

In pulp and paper industry, water is not only used as raw material but also for cooling and lubricating the machines. Thus huge amount of water flow through the processes and the amount of produced wastewater is large even though for economic and environmental reasons water is recycled as much as possible. For example, Finnish Forest Industries [1] states that nowadays in Finland from five to 50 cubic meters of water is used to produce one metric ton of pulp, and producing one metric ton of paper from seven to 15 cubic meters of water is consumed. The wastewaters of a pulp and paper industry are most commonly treated in a complex biological process, an activated sludge plant (ASP), where several physical, chemical and microbiological factors simultaneously affect the purification process. The operation of the treatment process depends on the condition of the biomass which is very sensitive to internal and external disturbances. A good bacterial balance enables the high pollution removal rate, low suspended solids in the effluent and a good settling properties of the sludge, which are considered as the key elements of an efficiently operating process. However, the quantity and quality of the incoming wastewater may vary drastically especially when wastewaters from other processes in the mill area (for example chemical processes) are led to the same wastewater treatment process. The major changes affect the condition of biomass detrimentally and because the recovery from the disturbances is slow the effects on the process operation and purification results are long-lasting. The disturbances in the bacterial balance will most often be shown as dysfunctional flocculation and settling properties which may have serious environmental and economic effects [2].

The main goal of a wastewater treatment is to remove organic compounds, excessive nutrients and toxicants from the treated water that can be reused or discharged to waterways. Stringent limitations to effluent discharges defined by authorities and constantly rising operating costs force the wastewater treatment plant operation to be optimized. However, an advanced control requires accurate monitoring of the process. A modern activated sludge plant has a wide range of on-line and off-line measurements but the common process measurements do not include sufficient information on the special features of the wastewater or give an early-warning information on the process efficiency and the effluent quality. Biological oxygen demand (BOD) and chemical oxygen demand (COD) are measures of the amount of dissolved oxygen required to oxidize the organic substances in wastewater. Nitrogen (N) and phosphorus (P) are useful nutrients but as large amounts in discharged wastewaters they cause eutrophication. Together with suspended solids (SS) the above-mentioned parameters are often used to indicate the effluent quality and the efficiency of a wastew-

* Corresponding author at :University of Oulu, Control Engineering, P.O. Box 4300, FI-90014 University of Oulu, Finland.
E-mail address: jani.tomperi@oulu.fi (J. Tomperi).

http://dx.doi.org/10.1016/j.jwpe.2017.02.004
2214-7144/

ater treatment process. However, measuring SS, a sludge volume index (SVI), nutrients content, or amount of organic substances from the effluent only show the prevailing quality of the wastewater. On this account, there is a demand for using new on-line monitoring tools and methods for assessing the process and predicting the upcoming quality of the effluent. Optical monitoring of floc morphological characteristics is a potential tool to be used as assistance in process control as it gives fast objective information about the quality of wastewater and the state of the treatment process, reveals some of the reason for settling problems and combined to a predictive modelling shows the quality of effluent in advance hours before it is noticed by traditional process measurements [3,2–8].

In this study, the results of the automatic optical monitoring of wastewater samples taken from a full-scale industrial active sludge process during a period over one year were utilized to develop predictive models for the effluent quality parameters (suspended solids, biological oxygen demand, chemical oxygen demand, nitrogen and phosphorus). Five variable selection methods were used for selecting the optimal subsets of input variables for each developed model. The study included also a short inspection of dependencies between the selected optical monitoring variables and the quality parameters.

2. Material and methods

2.1. Data

The data for this study was collected from the activated sludge process of a Finnish pulp and paper mill that treats in addition to the pulp making processes wastewaters from two chemical processes located in the mill area. The average wastewater flow through the treatment process is around 30 000 m^3/day. The unit operations of the wastewater treatment process are intake, screening, preliminary settling, neutralization, aeration, secondary settling and discharge. Wastewater samples for the optical monitoring were taken from the aeration tank and analyzed on the same working day within a couple of hours after sampling. Hydraulic delay in the aeration tank is 20 h at the average flow and in the secondary settling tank around 10 h. The dataset included optical monitoring results and selected process measurements including the effluent quality parameters from a period of 13 consecutive months. The optical monitoring was carried out by a beforehand planed schedule which consisted of regular but sparse sampling periods and more frequent sampling periods when wastewater samples were monitored nearly daily. In overall, dataset included 54 measurement times.

Before modelling, the dataset including several different type of measurements was scaled between [−2,2] using the nonlinear scaling method based on generalized moments, norms and skewness [9]. Before the scaling, the dataset was inspected and missing values were replaced with interpolation.

2.2. Optical monitoring and image analysis

To replace a laborious, slow and subjective method to study wastewater samples manually using a microscope, a small-scale automatic optical monitoring device was designed [10] and applied to analyze the wastewater samples taken from the aeration tank of the industrial ASP. Wastewater samples were diluted with deionized water at a ratio of 1:200 and pumped through a cuvette which was imaged with a high resolution charge-coupled device (CCD) camera. The CCD camera image sensor was 5.0 mm * 3.7 mm (1392 * 1040 pixels) with a pixel size of 3.6 μm * 3.6 μm. One video from a wastewater sample contained approximately 250–350 images.

The developed imaging system measures and analyses several morphological features of the flocs and filaments. Image processing and analyses methods and the mathematical formulas of the calculated size and shape parameters are presented in details in Koivuranta et al. [10]. For this study, the number of analyzed parameters were limited to the most suitable ones based on the preliminary laboratory studies carried out during the development period of the optical monitoring device. The parameters were calculated as an average of the values for individual objects over a single image. In addition to size parameters such as equivalent diameter, floc area and filament length, the calculated shape parameters included for example fractal dimension, form factor and roundness. The equivalent diameter is the diameter of a circle with an area equal to the object's area. The form factor is affected by the irregularity or roughness of the object's boundary and it is 1.0 for a perfect circle and below 1.0 for any other shape because objects with more irregular boundaries have a longer perimeter per surface area and therefore have smaller form factors. Roundness is defined as the ratio between the area of an object and the area of a circle with a diameter equal to the object's length. Roundness is also 1.0 for a perfect circle. The aspect ratio, which varies from 1.0 (for a perfect circle) to infinity, describes how elongated an object is [11].

In the following results, the amount of filaments is presented as a ratio of total filament length and floc area, where the total filament length is the sum of filament length of all the individuals present in the image. The number of small objects was calculated based on the size distribution, where each object was assigned to a size category based on its equivalent diameter. The size distribution was calculated as the sum of the distributions of individual images. In the image analysis, the threshold value for floc area was 100 μm^2 because the boundaries of the smaller objects may not have been sharp enough due to the resolution of the camera. The limit value for the small objects was equivalent diameter of 25 μm.

2.3. Variable selection

Variable selection is a practical way to reduce the amount of variables available and to choose the optimal input variables of a predictive model from a large dataset. Input variables that include noise, are correlated to each other or have no significant relationship with the output variable increase the computational complexity and reduce the prediction result of the model. In model development, a good principle is to keep the amount of input variables decent. Using too many input variables increase the risk to develop an over-fitted model which has an excellent training result but is not usable with a new upcoming data. In this work, five variable selection methods (correlation based selection, forward selection, stepwise selection, a genetic algorithm (GA) and a successive projections algorithm (SPA)) were utilized to find the optimal subsets of input variables for modelling the SS, BOD, COD, total nitrogen and total phosphorus in the effluent.

Variable selection methods can be roughly grouped into wrapper and filter methods. In a filter method, variables are selected or deleted according to the formed ranking which is based on the correlation coefficients. Filter methods are very efficient but the developed model is seldom optimal. In a wrapper method, a subset of variables is assessed according to their usefulness to a given predictor. Wrapper methods wrap around an appropriate learning machine which is employed as the evaluation criterion, such as prediction or classification error. Wrappers often give better results but are slower than filters. Correlation-based selection and a successive projections algorithm are classified as filters and forward selection and a genetic algorithm are classified as wrappers [12,13]. For a very large dataset one variable selection method can be used for the variable elimination before the final variable selection by another method. In Sorsa et al. [14] a successive projections algorithm was