Assessment of the reliability of the operation of a sewage treatment plant using Monte Carlo simulation

Karolina Migdal1), Agnieszka Operacz1), Iryna Vaskina2), Paulina Śliż2), Jorge Tavares4,5), Adelaide Almeida4,5,6), Michał Migdal7)

1) University of Agriculture in Krakow, Faculty of Environmental Engineering and Land Surveying, Department of Sanitary Engineering and Water Management, al. Mickiewicza 24/28, 30-059 Kraków, Poland
2) Sumy State University, Faculty of Technical System and Energy Efficient Technologies, Department of Applied Ecology, Sumy, Ukraine
3) Cracow University of Economics, Institute of Spatial Management and Urban Studies, Department of Spatial Management, Kraków, Poland
4) Polytechnic Institute of Beja, Department of Technology and Applied Sciences, Beja, Portugal
5) University of Beira Interior, Faculty of Engineering, Research Unit Fiber Materials and Environmental Technologies (FibEnTech-UBI), Covilhã, Portugal
6) University Nova of Lisbon, Faculty of Science and Technology, Center for Environmental and Sustainability Research (CENSE), Lisbon, Portugal
7) Stalprodukt S.A., Bochnia, Poland

Abstract: The aim of the study was to model the operation of a wastewater treatment plant using the Monte Carlo method and selected probability distributions of random variables. Pollutant indices in treated wastewater were analysed, such as: biological oxygen demand ($BOD_5$), chemical oxygen demand ($COD_{Cr}$), total suspended solids ($TSS$), total nitrogen ($N_{tot}$), total phosphorus ($P_{tot}$). The preliminary analysis of pollution indicators series included the: calculation of descriptive statistics and assessment of biological degradability of wastewater. The consistency of the theoretical distributions with the empirical ones was assessed using Anderson–Darling statistics. The best-fitting statistical distributions were selected using the percent bias criterion. Based on the calculations performed, it was found that the analysed indicators of pollution in treated wastewater were characterised by an average variability of composition for $BOD_5$, $COD_{Cr}$ and $TSS$, and a high variability of composition for $N_{tot}$ and $P_{tot}$. The best fitted distribution was log-normal for $BOD_5$, $TSS$, $N_{tot}$ and $P_{tot}$ and general extreme values for $COD_{Cr}$. The simulation carried out using the Monte-Carlo method confirmed that there may be problems associated with the reduction of nutrients ($N_{tot}$ and $P_{tot}$) in the analysed wastewater treatment plant. Results of values obtained of the risk values of negative control of wastewater treatment plant operation for biogenic compounds, different from 1, indicate that the number of exceedances at the outflow may be higher than the acceptable one.

Keywords: efficiency, mathematical simulation, Monte Carlo simulation, pollution indicators, probability distribution, sewage treatment plant
INTRODUCTION

In recent years, an increasing rate of industrialisation, urbanisation and population growth can be observed, which directly translates into increased environmental pollution. For this reason, sewage treatment has become one of the most important environmental issues [Angelakis, Gikas 2014; Flores-Alsina et al. 2010; Barrou et al. 2018; Taheriyoun, Moradnejad 2015]. Due to the variety of sewage treatment processes, use of chemical reagents, production of sludge and emission of gases, it is very important to carefully analyse the reliability of pollutant removal.

The evaluation of the performance of wastewater treatment plants includes many individual objectives, with the most important being the removal of pollutants that may cause changes in the aquatic environment [Flores-Alsina et al. 2008]. Therefore, the selected alternative method needs to comply with current regulations as well as to minimise the environmental impact on the receiving water body [Gernaey et al. 2004; Gizińska-Górska et al. 2017; Młynski et al. 2018; Todeschini 2016]. Therefore, it is important to study the quality of wastewater after the treatment process, which allows verification of the operation of individual facilities of the process line and the final effect of treatment in relation to the applicable guidelines.

There are many studies concerning the disposal of municipal wastewater at selected treatment plants. They concern both the analysis of wastewater treatment efficiency [Bugajski et al. 2019; Hendren et al. 2013; Kuber et al. 2020; Młynski et al. 2016a, b; Taheriyoun, Moradnejad 2015] and the susceptibility of wastewater to biological decomposition processes [Młynski et al. 2017; Noworolska-Majewska, Bugajski 2019; Plucznik-Koropczuk, Jakubaszek 2012]. The studies and results presented here provide important information for plant operators. The proper functioning of wastewater treatment plants is evidenced by their reliability, which means the plant’s capacity to discharge the expected amount of sewage, to the extent required by their receiver under certain operating conditions of the plant, within the assumed operating time and with random changes in the functional characteristics of the plant [Dlugosz, Gawożek 2013].

One of the most commonly used methods of assessing the performance of wastewater treatment plants is to determine reliability indices of the facility’s operation by comparing the values of pollutant indices in the effluent (arithmetic mean) to the permissible values set out in the relevant Regulation. The presented assumption is highly simplified, as it is based only on an empirical sample of analysed random variables [Młynski et al. 2019]. According to Olyar et al. [2018], the reliability of wastewater treatment plant operation is mainly considered as the percentage of time in which the expected concentration of pollutants in the effluent at the outfall meets the standards. Based on the variability of the pollutant indicators in the treated effluent, mathematical simulations of the operation of wastewater treatment plants can be carried out, which provides a more complete picture of the assessment of the plant’s performance compared to a model based solely on short observation series for the given indicators. Simulations of wastewater treatment plant operation are often considered research problems, both at the stage of design, modernisation or operation of such facilities. However, the research so far with the use of mathematical models focuses on the simulation of pollutant values at the outflow with an assumed specific form of probability distribution, e.g. Weibull [Bugajski et al. 2016; Marzec 2017]. It should be emphasised that this represents a large generalisation because pollution indicators have an underdetermined character. Hence, it can be used more than one probability distribution to describe them. To do this, the best-fitting functions should be selected from the group of distributions [Młynski et al. 2019]. Monitoring the operation of wastewater treatment plants with the use of probability distributions allows to present the analysed phenomenon in a wider range, thanks to modelling complex dependencies and forecasting their values at a given time. The simulation values obtained can be helpful in assessing the risk of improper functioning of the treatment plant.

With this in mind, the paper presents a simulation of the operation of a wastewater treatment plant using selected statistical distributions, which has allowed the determination of the size of reliability indexes of the facility’s operation based on the predicted values, reflecting the factors affecting the operation of the facility.

MATERIALS AND METHODS

STUDY OBJECT

Reliability analysis was carried out on the example of a sewage treatment plant located in the city of Tarnów in southern Poland. A mechanical-biological wastewater treatment plant is a facility that uses active sludge in the biological phase to remove pollutants from wastewater. The plant receives wastewater from individual users as well as industrial wastewater. Total population equivalent for wastewater treatment plant is 460 000. Design capacity of the treatment plant is 86 400 m³·d⁻¹. The process of mechanical wastewater treatment includes: screens, horizontal sand traps, radial primary clarifiers. The process of biological decomposition of pollutants takes place in biological reactors with active sludge, followed by the phase of clarification of treated sewage in secondary radial settling tanks. At the end of the process, the treated wastewater is discharged to the receiving body, i.e. Biała.

DATA USED

The analysis was made on the basis of the data covering the values of pollution indicators: $BOD_5$ (biochemical oxygen demand), $COD_{Cr}$ (chemical oxygen demand), total suspended solids (TSS), total nitrogen ($N_{tot}$) and total phosphorus ($P_{tot}$) in the period from 2015 to 2020 for raw and treated sewage. The analysis was made according to the following stages: preliminary statistical analysis of the series of pollutant indicators, assessment of the susceptibility of wastewater to biological decomposition processes, selection of the theoretical function, best-fit empirical distribution, modelling of the values of pollutant indicators in raw and treated wastewater, determination of the values of reliability indicators of the operation of the treatment plant based on the results of the simulation carried out. According to the current Rozporządzenie [2019] for wastewater treatment plants serving more than 100,000 PE, the permissible values for: $BOD_5$ – 15 mg O₂·dm⁻³, $COD_{Cr}$ – 125 mg O₂·dm⁻³, TSS – 35 mg O₂·dm⁻³, $N_{tot}$ – 10 mg·dm⁻³, $P_{tot}$ – 1 mg·dm⁻³. The required number of control samples is 24 samples per year, while the number of non-
compliant samples is 3. Samples of raw and treated sewage for physicochemical tests were collected twice a month, in the period from 2015 to 2020, which gives 159 samples of raw and 159 samples of treated sewage that were analysed in the period from 2015 to 2020.

**PRELIMINARY ANALYSIS OF POLLUTION INDICATORS**

The preliminary analysis of pollution indicators series included the calculation of descriptive statistics. The following descriptive statistics were determined: mean (mean), minimum (min) and maximum (max) values; measures of dispersion – standard deviation (SD) and coefficient of variation (CV).

**ASSESSMENT OF BIOLOGICAL DEGRADABILITY OF WASTEWATER**

The assessment of the susceptibility of wastewater inflow to the treatment plant in Tarnów and after the biological treatment process was made on the basis of the ratio of COD$_{cr}$/BOD$_5$. For this purpose, the mean annual values of both indicators of the content of organic compounds were compared. This assessment was based on the relationships available in the literature describing the measures of biodegradability of organic pollutants present in wastewater according to the following relationships [Klimuk, Lejkowska 2003; Miersch, Sikora 2010]:

1. COD$_{cr}$/BOD$_5 < 2.0$ easy biodegradability
2. COD$_{cr}$/BOD$_5 = 2.0–2.5$ mean bio-degradability
3. COD$_{cr}$/BOD$_5 = 2.5–5.0$ low biodegradability
4. COD$_{cr}$/BOD$_5 > 5.0$ non-degradable matter

Additionally, based on the mean annual concentrations of $N_{tot}$ and $P_{tot}$, both in the sewage before entering the technological system and after treatment, the susceptibility of the wastewater to the processes of biological removal of biogenic compounds was determined. For this purpose, quotients of the values of $N_{tot}/BOD_5$ and $P_{tot}/BOD_5$ were determined and then compared to the values reported in the literature, which indicate that denitrification (5) and defosphatation (6) processes occur most effectively when [Heidrich, Witkowski 2015; Janosz-Rajczyk 2008; Lomotowski, Sz mundane 1999]:

5. $N_{tot}/BOD_5 < 0.25$
6. $P_{tot}/BOD_5 < 0.04$

**THEORETICAL DISTRIBUTION FITTING**

Before the Monte Carlo simulation, a three-stage intermediate analysis was performed in the analysis: adjusting the theoretical distribution to empirical distributions, assessing the consistency of the theoretical distributions with the empirical distributions of the analysed pollution indicators and selecting the best-fitted theoretical distributions.

On the basis of observation series of the analysed indicators of pollution in raw and treated wastewater, an attempt was made to determine the best-fitting theoretical distributions to the empirical distributions of the random variables under study. The following theoretical distributions were analysed: Johnson SU, Weibull, Gaussian mixture model (GMM), general extreme values (GEV), half-normal distribution, log-normal, Pareto, normal distribution, triangular, Rayleigh. The distributions of each distribution $f(x)$ are described by the following functions [Alam et al. 2018; Glickman, Xu 2008; Jagiello et al. 2016; Walejka et al. 2014; Wang (ed.) 2010; Yu et al. 2012; Zoran, Weiss 2011]:

- **Johnson SU distribution:**
  \[ f(x; \xi, \lambda, \gamma, \delta) = \frac{1}{\lambda} \exp \left\{ \frac{1}{2} y^2 \right\} \]
  \[ y = \frac{x - \xi}{\lambda} + \delta \]
  \[ \delta = \frac{\gamma}{\lambda} \]
  where: $x = \text{variable}$, $\lambda = \text{scale parameter}$, $\delta = \text{shape parameter}$, $\xi = \text{location parameter}$.

- **Weibull distribution:**
  \[ f(x; \alpha, \beta) = \frac{\beta}{\alpha} \left( \frac{x}{\alpha} \right)^{\beta-1} \exp \left( -\left( \frac{x}{\alpha} \right)^\beta \right) \]

- **GMM distribution:**
  \[ f(x) = \sum_{k=1}^{K} \omega_k f_k(x; \theta_k) \]

- **GEV distribution:**
  \[ f(x) = \frac{1}{\beta} \left( 1 + \frac{y}{\mu} \right)^{-1/\beta} \exp \left( -\left( 1 + \frac{y}{\mu} \right)^{-1/\beta} \right) \]

- **Half-normal distribution:**
  \[ f(x) = \frac{2}{\sqrt{2\pi}\sigma} \exp \left( -\frac{x^2}{2\sigma^2} \right) \text{ if } x \geq 0 \]

- **Log-normal distribution:**
  \[ f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(\ln(x) - \mu)^2}{2\sigma^2} \right) \]

- **Pareto distribution:**
  \[ f(x) = \frac{c}{x_{\text{ref}}^{c+1}} \]

where: $c = \text{shape parameter}$.
• Normal distribution
\[ f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x - \mu)^2} \]  
(16)

• Triangular distribution
\[ f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(m-a)} & \text{for } a \leq x \leq m \\ \frac{2(b-x)}{(b-a)(b-m)} & \text{for } m \leq x \leq b \end{cases} \]  
(17)

where: \( m = \) mode, \( a = \) lower limit, \( b = \) upper limit.

• Rayleigh distribution
\[ F(x; \lambda) = 1 - e^{-\lambda x^2}, \quad x \geq 0, \lambda > 0 \]  
(18)

\[ f(x; \lambda) = 2\lambda^2 x e^{-\lambda x^2}, x \geq 0, \lambda > 0 \]  
(19)

where: \( \lambda = \) scale parameter.

The Anderson–Darling (A–D) test was used to assess the compliance of the theoretical distributions with the empirical distributions of the analysed pollution indicators. The A–D statistic is more sensitive over the entire range of distributions and is more likely to detect differences between distributions, than it gives a better estimate of compliance than the other tests. The hypotheses for the A–D test are as follows: \( H_0 \) data has a specific distribution, \( H_1 \) data does not have a specific distribution. The Anderson–Darling statistics is described by the correlations [Evans et al. 2017; Jantschi, Bolboaca 2018; Zeng et al. 2015]:
\[ A - D = -n - \sum_{i=1}^{n} \left( \frac{2i - 1}{n} \ln[F(X_i)] + \ln[1 - F(X_{n+1-i})] \right) \]  
(20)

where: \( n = \) observations number, \( F = \) theoretical cumulative distribution, \( F_n = \) empirical cumulative distribution.

Since the critical values for the A–D statistic depend on the type of probability distribution being tested, hence the verification of the hypotheses was based on the \( p \)-value of the test. If the \( p \)-value was less than \( \alpha = 0.05 \), then rejected the null hypothesis that the data come from that distribution.

**SELECTION OF BEST-FIT THEORETICAL DISTRIBUTIONS**

If random variables could be described by more than one theoretical distribution, an identification of the best-fitting theoretical distributions to the empirical distributions was made using the percent bias criterion (PBIAS), which describes the Equation (21) [Abdul et al. 2020]:
\[ PBIAS = \frac{\sum_{i=1}^{n} (Q_{ob} - Q_{es})}{\sum_{i=1}^{n} Q_{ob}} \times 100 \]  
(21)

where: \( Q_{ob} = \) observed values, \( Q_{es} = \) estimated values.

The PBIAS describes the average tendency of estimated values to be greater or less than their observed values. The closer the PBIAS value is to 0 the greater the agreement of the compared quantities, positive values indicate underestimated of the model and negative values indicate overestimation [Gupta et al. 1999; Miorias et al. 2007; Yuan et al. 2014]. According to guidelines, PBIAS value < ±10% indicates a very good fit; ±10% ≤ PBIAS < ±15% – good; ±15% ≤ PBIAS < ±25% – satisfactory; and PBIAS ≥ ±25% – unsatisfactory fit [Archibald et al. 2014; Dongjian et al. 1983; Miorias et al. 2007; Singh et al. 2004; Van Liew et al. 2003].

**MODELLING OF POLLUTANT INDICATOR VALUES**

Modelling of the values of pollutant indices in treated wastewater was carried out using the Monte-Carlo method. It is defined as the use of a sequence of random numbers to build a sample from a hypothetical population, from which it is possible to determine statistical estimators of the parameters of the problem sought to be solved [Halton 1970]. In practice, it consists in the formulation of a stochastic model describing a real phenomenon, and then repeated implementation of this model using randomly generated variables (according to the assumed probability distribution) and statistical analysis of the obtained results. The Monte-Carlo method is widely used in analyses related to the operation of water and wastewater management facilities, as indicated by the works of Hendren et al. [2013], Schaubberger et al. [2013], Barton et al. [2015]. In this study, simulations were performed for best-fit statistical distributions describing the empirical distributions of the analysed pollution indicators. One simulation cycle consisted of the generation of 365 random values of pollution indicators. For each generated set, a 24-element sample was determined at random. In each cycle, the number of observations from the simulation which did not meet the criterion of technological effectiveness (i.e. where the value from the simulation was greater than the permissible value) was determined and the result of the control of operation of the treatment plant was determined (control indicator \( \text{CF} = 1 \) when the treatment plant was assessed negatively, i.e. when the number of samples not meeting the requirements exceeded the permissible value of the number of samples which may not meet such requirements; \( \text{CF} = 0 \) otherwise). Each of the simulation cycle was then repeated 100 times.

**DETERMINATION OF RELIABILITY COEFFICIENTS FOR THE OPERATION OF TREATMENT PLANTS**

On the basis of the simulation carried out, the reliability estimators of the operation of the wastewater treatment plant were determined: the technological efficiency index \( R \), the reliability coefficient \( CR \) and the risk of negative control of the operation of the wastewater treatment plant \( \text{Rr} \). Calculations were carried out according to the formulas [Andraka, Dzenis 2013]:
\[ R = \frac{N_{sym} - N_{Xperm}}{N_{sym}} \]  
(22)

where: \( N_{sym} = \) number of simulations carried out, \( N_{Xperm} = \) number of samples from simulations not meeting the limit values for the pollutant indicator analysed:
\[ CR = m_e \]  
(23)

where: \( m_e = \) the average value of a given indicator for the simulations carried out, \( N_{perm} = \) the limit value of the indicator concerned.
The analysis of the results carried out in this study includes preliminary analysis of the data, assessment of the susceptibility of wastewater to biological decomposition processes, selection of the best-fitting theoretical to empirical distribution, modelling of the values of pollutant indices in raw and treated wastewater, together with an analysis of the correctness of the two-stage treatment processes.

PRELIMINARY DATA ANALYSIS

In order to characterise the dynamics of changes in the quality of raw and treated sewage, descriptive statistics were determined: location and dispersion measures for the analysed values of pollutant indexes in sewage. The results of the analyses are presented in Table 1.

The results of the preliminary analysis are presented in Table 1. Based on it, it was found that in the raw sewage the difference between the extreme values (min. and max.) was respectively: for $\text{BOD}_5 \sim 85\%$, $\text{COD} \sim 85\%$, $\text{TSS} \sim 88\%$, $N_{\text{tot}} \sim 76\%$, $P_{\text{tot}} \sim 88\%$. Additionally, it can be concluded that the composition of wastewater flowing into the Tarnów WWTP did not differ significantly from the typical composition for domestic wastewater from southern Poland, as indicated by the results of studies conducted by Kaczor [2009] and Młynski et al. [2020]. The values of coefficient of variation ($C_v$) for individual indicators of pollution indicate an average variability of the composition of wastewater flowing into the treatment plant. In the case of treated wastewater, the characteristic values of pollutant indicators are presented in Table 1 and Figure 1. The disproportion between the values of pollutants from the organic group was respectively for $\text{BOD}_5 \sim 84\%$ and $\text{COD} \sim 81\%$, and for $\text{TSS}$ was at the level of 81%. For biogenic compounds, the differences were at the level of 93% for $N_{\text{tot}}$ and 90% for $P_{\text{tot}}$. In the case of organic pollutants in the period of analysis (2015–2020) exceedances of limit values can be observed. For $\text{BOD}_5$ there were 8 exceedances of the limit value (15 mg O$_2$·dm$^{-3}$) and for $\text{COD}$ there were two observations above 125 mg O$_2$·dm$^{-3}$. For $\text{TSS}$, two samples were recorded throughout the observation period that did not meet the limit value of 35 mg·dm$^{-3}$. In the case of biogenic compounds as many as 31 exceedances of the standard values (10 mg·dm$^{-3}$) were recorded for total nitrogen in treated wastewater and 10 samples which exceeded the permissible value of total phosphorus concentration in the outflow (1 mg·dm$^{-3}$).

The recorded exceedances of values of pollutant indices could have been caused by a varied course of factors which significantly shape the effectiveness of treatment processes, i.e. the volume of wastewater inflow or the temperature of wastewater. Another factor may be the size of the pollutant load in the effluent entering the plant. The quality and quantity of wastewater is strongly influenced by the nature of urban households. People, who did not have access to the sewage treatment are accustomed to saving water. The result is the concentration of higher amount of pollutants in less water used. In addition, trends in household water consumption result in high heterogeneity in wastewater discharge, which exposes treatment systems to unstable operation [Wałęga et al. 2018]. Analysing the value of the $C_v$ values, we can observe an average variability in the composition of organic pollutants and total suspended solids, and a high variability in the composition for total nitrogen and total phosphorus at the outflow in wastewater after biological treatment processes. As shown by the results of studies conducted by other authors, this is a characteristic feature of treated wastewater [Bugański et al. 2016; Chmielowski et al. 2017; Kaczor et al. 2015; Młynski et al. 2020; Wąsk, Chmielowski 2013].

### RESULTS AND DISCUSSION

GENERAL INFORMATION

The analysis of the results carried out in this study includes preliminary analysis of the data, assessment of the susceptibility of wastewater to biological decomposition processes, selection of the best-fitting theoretical to empirical distribution, modelling of the values of pollutant indices in raw and treated wastewater, together with an analysis of the correctness of the two-stage treatment processes.

Table 1. Preliminary data analysis for observation series of pollutant indicators in raw and treated wastewater

<table>
<thead>
<tr>
<th>Pollutant indicator</th>
<th>Min.</th>
<th>Mean</th>
<th>Max.</th>
<th>SD</th>
<th>$C_v$ (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw sewage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{BOD}_5$</td>
<td>140.00</td>
<td>518.93</td>
<td>940.00</td>
<td>182.55</td>
<td>0.35</td>
</tr>
<tr>
<td>$\text{COD}_{\text{cr}}$</td>
<td>266.00</td>
<td>970.67</td>
<td>146.00</td>
<td>294.37</td>
<td>0.30</td>
</tr>
<tr>
<td>$\text{TSS}$</td>
<td>140.00</td>
<td>476.43</td>
<td>1200.00</td>
<td>165.62</td>
<td>0.35</td>
</tr>
<tr>
<td>$N_{\text{tot}}$</td>
<td>36.40</td>
<td>80.09</td>
<td>149.00</td>
<td>18.23</td>
<td>0.23</td>
</tr>
<tr>
<td>$P_{\text{tot}}$</td>
<td>2.93</td>
<td>7.56</td>
<td>24.50</td>
<td>2.64</td>
<td>0.35</td>
</tr>
<tr>
<td>Treated sewage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{BOD}_5$</td>
<td>3.50</td>
<td>7.90</td>
<td>23.00</td>
<td>3.27</td>
<td>0.41</td>
</tr>
<tr>
<td>$\text{COD}_{\text{cr}}$</td>
<td>27.50</td>
<td>69.41</td>
<td>146.00</td>
<td>20.26</td>
<td>0.29</td>
</tr>
<tr>
<td>$\text{TSS}$</td>
<td>6.80</td>
<td>15.07</td>
<td>35.00</td>
<td>5.81</td>
<td>0.39</td>
</tr>
<tr>
<td>$N_{\text{tot}}$</td>
<td>3.30</td>
<td>10.57</td>
<td>50.60</td>
<td>8.06</td>
<td>0.76</td>
</tr>
<tr>
<td>$P_{\text{tot}}$</td>
<td>0.21</td>
<td>0.59</td>
<td>2.00</td>
<td>0.31</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Explanations: $SD =$ standard deviation, $C_v =$ coefficient of variation, $BOD_5 =$ oxygen demand, $COD_{cr} =$ chemical oxygen demand, $TSS =$ total suspended solids.

Source: own study.

© 2022. The Authors. Published by Polish Academy of Sciences (PAN) and Institute of Technology and Life Sciences – National Research Institute (ITP – PIB).

This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/3.0/).
decomposition of biogenic compounds, it can be observed that the determined relations between $N_{\text{tot}}$ and $BOD_5$ and between $P_{\text{tot}}$ and $BOD_5$ provided appropriate conditions for the processes of biological decomposition of nitrogen compounds by denitrification, and the removal of phosphorus compounds by defosfatation. After the process of biological treatment of wastewater, an increase in the values of the determined ratios can be observed, which indicates a significant decrease in the amount of biologically degradable substances and in the effectiveness of the processes of removing biogenic compounds (Fig. 2). Similar results with observations were also noted in the work by MŁYŃSKA et al. [2017], where the authors obtained very similar values of quotients of analysed indicators of pollution. The obtained quotient values indicate the proper functioning of the mechanical-biological treatment plant in Tarnów.

**VERIFICATION OF THEORETICAL PROBABILITY DISTRIBUTIONS**

The fit of the theoretical distributions to the empirical distributions of the observational series of the pollution indicators analysis in raw and treated sewage was determined by the following functions: Johnson SM, Weibull, GMM, GEV, half-normal, log-normal, normal, Pareto, triangular, Rayleigh. The fit of the distributions was assessed using the Anderson–Darling test for a fixed significance level of $\alpha = 0.05$. When the obtained $p$-value was below the assumed significance level, it was found that the theoretical functions did not match the empirical functions for the analysed indicator measurement series. The obtained results of the analysis are presented in Table 2.
Based on the results presented in Table 2, it was found that the observation series for the $BOD_5$ indicator can be described by the three probability distributions analysed (GMM, GEV, log-normal). This is indicated by the $p$-value for the A–D test statistics taking values above the assumed significance level $\alpha = 0.05$. For the second indicator from the organic group ($COD_{Cr}$), the A–D test showed that a $p$-value above 5% occurs for the distributions: GMM, GEV, Johnson SU, log-normal, Normal, Pareto, Weibull, which means that the hypothesis $H_0$ stating that the theoretical distribution analysed is consistent with the empirical distribution should be accepted. With regard to total suspended solids, a $p$-value above 0.05 indicates the possibility of using the distributions: GMM, GEV, Johnson SU. On the other hand, in the case of indexes from the biogenic group ($N_{tot}$ and $P_{tot}$), statistical analysis showed the conformity of the theoretical distribution with the empirical distribution of these indices for the functions: GMM, GEV, log-normal.

**SELECTION OF THE BEST-FITTING STATISTICAL DISTRIBUTION**

Based on the analysis of the matching of the theoretical and empirical distributions, it was found that the observation series of pollution indicators in treated wastewater can be described by more than one theoretical distribution. Calculations were performed only for those functions for which the A–D criterion confirmed their conformity with the empirical distribution of random variables. The selection of the best-fitting distribution was made on the basis of the PB criterion and the quantile-quantile fit plot. The results of the analyses are summarised in Table 3, and Figure 3 shows the best-fit statistical distributions.

Based on the results presented in Table 3, it was found that for $BOD_5$, the best-fitting theoretical distribution is the log-normal distribution, for which the $PBIAS$ value is equal to $-0.225\%$. For $COD_{Cr}$, the best-fitting theoretical distribution was found to be the GEV distribution, for which the $PBIAS$ value is $-0.094\%$. In the case of total suspended solids and biogenic indicators ($N_{tot}$ and $P_{tot}$), the best-fitting theoretical function is the log-normal distribution, which is confirmed by the $PBIAS$ values. The analysis carried out allowed to explicitly indicate that for the indicator from the organic group ($BOD_5$), biogenic indicators and $TSS$ the best-fitting theoretical distribution is log-normal. In the case of $COD_{Cr}$, such a distribution turned out to be GEV, however the comparable value of the $PBIAS$ criterion for the log-normal distribution indicates its applicability in the description of this indicator. Figure 3 presents the best-fitting statistical distributions describing the variables. The possibility of using the log-normal distribution to monitor the performance of wastewater treatment plants has been described in the works of numerous authors [Andraka, Dejens 2013; Oliveira, von Sperling 2007; 2008], which indicate that it is the log-normal distribution that best captures the statistical variability of quality parameters in treated wastewater.

---

**Table 2.** The value of fitting the analysed theoretical distributions to the empirical distributions of random variables of pollution indicators in treated sewage

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$BOD_5$</th>
<th>$COD_{Cr}$</th>
<th>$TSS$</th>
<th>$N_{tot}$</th>
<th>$P_{tot}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMM</td>
<td>1.352</td>
<td>0.215</td>
<td>0.297</td>
<td>0.939</td>
<td>0.829</td>
</tr>
<tr>
<td>GEV</td>
<td>1.078</td>
<td>0.318</td>
<td>0.357</td>
<td>0.889</td>
<td>0.708</td>
</tr>
<tr>
<td>Johnson SU</td>
<td>-</td>
<td>-</td>
<td>0.370</td>
<td>0.877</td>
<td>0.461</td>
</tr>
<tr>
<td>Log-normal</td>
<td>1.849</td>
<td>0.111</td>
<td>0.636</td>
<td>0.614</td>
<td>0.817</td>
</tr>
<tr>
<td>Normal</td>
<td>6.985</td>
<td>0.000</td>
<td>0.646</td>
<td>0.605</td>
<td>3.533</td>
</tr>
<tr>
<td>Weibull</td>
<td>6.359</td>
<td>0.001</td>
<td>0.855</td>
<td>0.442</td>
<td>2.995</td>
</tr>
<tr>
<td>Triangular</td>
<td>24.397</td>
<td>0.000</td>
<td>6.131</td>
<td>0.000</td>
<td>7.188</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>9.490</td>
<td>0.000</td>
<td>14.474</td>
<td>0.000</td>
<td>7.438</td>
</tr>
<tr>
<td>Pareto</td>
<td>23.211</td>
<td>0.000</td>
<td>24.841</td>
<td>0.000</td>
<td>18.186</td>
</tr>
<tr>
<td>Semi-normal</td>
<td>26.244</td>
<td>0.000</td>
<td>34.700</td>
<td>0.000</td>
<td>25.137</td>
</tr>
</tbody>
</table>

Explanations: $BOD_5$, $COD_{Cr}$, $TSS$ as in Tab. 1, A–D = Anderson–Darling statistics, $p$-value is equal to $\alpha = 0.05$ (theoretical distribution can be applied), GMM = Gaussian mixture model, GEV = general extreme values.

Source: own study.

---

**Table 3.** Analysis of percent bias ($PBIAS$) for the pollutants under study

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$BOD_5$</th>
<th>$COD_{Cr}$</th>
<th>$TSS$</th>
<th>$N_{tot}$</th>
<th>$P_{tot}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A–D</td>
<td>$p$</td>
<td>A–D</td>
<td>$p$</td>
<td>A–D</td>
</tr>
<tr>
<td>GMM</td>
<td>0.171</td>
<td>-0.024</td>
<td>0.233</td>
<td>1.409</td>
<td>0.481</td>
</tr>
<tr>
<td>GEV</td>
<td>0.005</td>
<td>-0.094</td>
<td>0.094</td>
<td>0.015</td>
<td>0.050</td>
</tr>
<tr>
<td>Johnson SU</td>
<td>-</td>
<td>-0.036</td>
<td>0.008</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Log-normal</td>
<td>-0.225</td>
<td>-0.067</td>
<td>-0.097</td>
<td>-0.772</td>
<td>-0.295</td>
</tr>
<tr>
<td>Normal</td>
<td>-</td>
<td>0.306</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Weibull</td>
<td>-</td>
<td>1.025</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Explanations: $BOD_5$, $COD_{Cr}$, $TSS$ as in Tab. 1, GMM, GEV as in Tab. 2, bold – best-fit statistical distributions.

Source: own study.
DETERMINATION OF RELIABILITY COEFFICIENTS FOR THE OPERATION OF THE TARNÓW WASTEWATER TREATMENT PLANT

Based on the results for the best-fitting statistical distributions, values for the pollution indicators were determined using the Monte-Carlo method. The modelling was based on the log-normal distribution for the indicators $BOD_5$, $TSS$ and biogenic group ($N_{tot}$ and $P_{tot}$). For $COD_{Cr}$, the simulation was based on the GEV distribution. On the basis of the simulated values, the indices of technological efficiency ($R$), reliability coefficient ($CR$) and risk of negative control of wastewater treatment plant operation ($R_e$) were calculated. The results are summarised in Table 4.

Based on the results of the simulation carried out, it can be concluded that the wastewater treatment plant in Tarnów is functioning properly in terms of the reduction of organic pollutants and total suspended solids. The values of technological efficiency index $R$ for $BOD_5$ and $COD_{Cr}$, as well as for $TSS$ are very close to each other and equal to unity, which indicates that the number of samples that do not meet the requirements for the...
Table 4. Efficiency indicators of wastewater treatment plant operation determined on the basis of Monte-Carlo simulation

<table>
<thead>
<tr>
<th>Reliability coefficient</th>
<th>Pollutant indicator</th>
<th>(BOD_5)</th>
<th>COD (_{Cr})</th>
<th>TSS</th>
<th>(N_{tot})</th>
<th>(P_{tot})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)</td>
<td>0.967</td>
<td>0.967</td>
<td>0.998</td>
<td>0.674</td>
<td>0.929</td>
<td></td>
</tr>
<tr>
<td>(CR)</td>
<td>0.520</td>
<td>0.553</td>
<td>0.429</td>
<td>0.999</td>
<td>0.575</td>
<td></td>
</tr>
<tr>
<td>(R_e)</td>
<td>0.050</td>
<td>0.090</td>
<td>0.050</td>
<td>0.980</td>
<td>0.270</td>
<td></td>
</tr>
</tbody>
</table>

Explanations: \(BOD_5\), COD \(_{Cr}\), TSS as in Tab. 1, \(R\) = technological efficiency, \(CR\) = reliability coefficient, \(R_e\) = risk of negative control. Source: own study.

The aim of this study was to model the operation of a wastewater treatment plant, using the Monte-Carlo method. The pollutant indexes in the treated wastewater were analysed: \(BOD_5\), COD \(_{Cr}\), TSS, \(N_{tot}\) and \(P_{tot}\). Simulations were performed using the following probability distributions: Johnson SU, Weibull, gaussian mixture model (GMM), general extreme values (GEV), half-normal distribution, log-normal, Pareto, normal, triangular, Rayleigh. The fit of the theoretical distributions together with the empirical distributions was assessed using Anderson–Darling (A–D) statistics. The best-fitting statistical distributions were selected using the percent bias criterion (PBIAS) criterion. Based on the calculations performed, it was found that the analysed indicators of pollution in treated wastewater were characterised by an average variability of composition for organic pollutants and total suspended solids, and a high variability of composition for total nitrogen and total phosphorus. The best fitted distribution was log-normal for \(BOD_5\), TSS, \(N_{tot}\) and \(P_{tot}\) and GEV for COD \(_{Cr}\). The simulation carried out using the Monte-Carlo method confirmed that there may be problems associated with the reduction of nutrients at the analysed wastewater treatment plant. The Monte Carlo simulation method is a useful tool for modelling the technological reliability of a wastewater treatment plant, provided that an appropriate theoretical distribution for the random variables is indicated. This makes it possible to generate new data while retaining existing correlation structures between variables. The values generated in this way can be useful in assessing the performance of the treatment plant and in preparing various possible scenarios of its operation.

REFERENCES


Rozporządzenie Ministra Gospodarki Morskiej i Żeglugi Śródlądowej z dnia 12 lipca 2019 r. w sprawie substancji szczególnie szkodliwych dla środowiska wodnego oraz warunków, jakie należy spełnić przy wprowadzaniu do wód lub do ziemi ścieków, a także przy odprowadzaniu wód opadowych lub roztopowych do wód lub do urządzeń wodnych [Regulation of the Minister of Maritime Economy and Inland Navigation of 12 July 2019 on substances particularly harmful to the aquatic environment and conditions to be met during sewage discharge into the water or into the ground and during rainwater or snowmelt discharge into the water or into the water devices]. Dz.U. 2019 poz. 1311.


