Environmental Technology & Innovation 27 (2022) 102638

Contents lists available at ScienceDirect

Environmental Technology & Innovation

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

Analysis of particle size distribution in municipal wastewaters

Matteo Cornacchia [a,b], Gabriele Moser [c], Ezio Saturno [b], Andrea Trucco [c], Paola Costamagna [a,*]

[a] Department of Chemistry and Industrial Chemistry (DCCI), University of Genoa, Via Dodecaneso 31, 16146 Genoa, Italy
[b] EXXRO S.r.l., Lungobisagno Istria 14, 16141 Genoa, Italy
[c] Department of Electrical, Electronics and Telecommunication Engineering and Naval Architecture (DITEN), University of Genoa, Via Opera Pia 11a, 16145, Genoa, Italy

ARTICLE INFO

Article history:
Received 24 November 2021
Received in revised form 6 April 2022
Accepted 27 April 2022
Available online 6 May 2022

Keywords:
Laser diffraction
Modeling
Municipal wastewater
Particle size distribution (PSD)
Probability density function (PDF)

ABSTRACT

Innovative membrane filtration plants for municipal wastewaters are being developed and need the support of reliable filtration models in the designing phase. In the past, semi-empirical filtration models for membrane processes have been proposed. At present, the most prominent works point out the importance of particle poly-dispersity in the development of reliable models but fail into the implementation of probability density functions (PDFs) capable of an accurate fitting of the experimental particle size distribution (PSD). We report the experimental PSDs of two different municipal wastewater samples, obtained through the laser diffraction technique. The experimental results show that the laser diffraction technique can characterize wastewater particle dimensions both in the colloidal and supra-colloidal regions. The experimental study is complemented by a comparative analysis in which many PDFs are used to fit the experimental PSDs through a least-squares approach. Some of these PDFs are proposed here for the first time to fit experimental wastewater PSDs. Among the PDFs considered for the statistical modeling, the three-parameter lognormal and the Burr PDFs are demonstrated to provide satisfactory fitting, whereas the other considered functions fail. This result is confirmed by the analysis of both the available wastewater samples.

1. Introduction

Municipal wastewater, or sewage, is a type of wastewater produced by a community, consisting mostly of gray and black waters. It is a complex matrix containing organic, inorganic, and biological compounds that can be found either in suspended particles or in solution. Particles can be divided into categories based on their size, i.e. dissolved (< 0.001 μm), colloidal (0.001–1 μm), supra-colloidal (1–10^2 μm), and settleable (> 10^2 μm) (Dulekgurgen et al., 2006). Most of the suspended particulate consists of organic material, whereas the soluble part consists mainly of inorganic compounds. However, since the soluble fraction is the predominant part, the greater portion of the total organic carbon (TOC) is identifiable among the soluble compounds (Rickert and Hunter, 1971).

The particle size distribution (PSD) represents the dimensional distribution curve of a powder, a granular material, or a suspended particles sample. It provides a list of standard values that defines the relative amount, by mass, volume, or number, according to size (Jillavenkatesa et al., 2001). In general, there are several options for determining the PSD of a sample. The choice of method depends upon several factors such as size range, sample quantity, required accuracy, time

* Corresponding author.
E-mail address: paola.costamagna@unige.it (P. Costamagna).

https://doi.org/10.1016/j.eti.2022.102638

consumption, and analysis costs. Traditional methods such as sieving, centrifugation, and gel filtration are not suitable for the specific characterization of wastewaters since the first two are both not very accurate and limited in the dimensional range, whereas the last is time-consuming and cumbersome. Flow field fractionation (FFF) and high-performance liquid chromatography (HPLC) allow to detection of particles also in the soluble region but they are expensive and limited to particle sizes smaller than 1 μm. They alone are not sufficient to determine the overall PSD of a sewage sample (Arimi, 2018). Electrical impedance counting method instruments can operate in the size range 0.01–10 μm and they are documented to be suitable to analyze agricultural wastewater (Chavez et al., 2004) and slaughterhouse effluent (Sanchis et al., 2003). The main issue with this technology is that it suffers from clogging phenomena and it has to work with dilute samples. On the other hand, techniques based on spectroscopic methods such as laser diffraction can identify particles in a very wide range that can go from 0.01 μm to 1 mm, depending on the analysis conditions and the configuration of the instrument, thus allowing to characterize the whole insoluble region of interest (Azema et al., 2002; Hocaoglu and Orhon, 2013).

The PSD can reasonably be considered as the fingerprint of wastewater. Some studies claim that PSD can be a convenient tool in the assessment of operating and design conditions in coagulation-flocculation processes (Chavez et al., 2006), others emphasize the correlation between PSD and the evaluation of parameters such as total suspended solids (TSS), turbidity, and chemical oxygen demand (COD) (Chavez et al., 2004; Dulekgurgen et al., 2006; Hocaoglu and Orhon, 2013). Furthermore, PSD study and modeling can serve as a powerful tool to understand the behavior and efficiency of wastewater filtration processes such as filter beds (Adin and Alon, 1993) and in modeling the functioning of membrane processes, as evidenced in the works of Foley et al. (1995), and Broeckmann et al. (2006). Both Foley and Broeckmann explicitly stress the role of particle poly-dispersity in membrane crossflow and hollow fiber filtration. They advance a successful approach that combines the classical filtration theory with the mathematical PSD formalization, which results in mathematical models able to describe key filtration behaviors such as the decreases in transmembrane flux and the increase in the cake thickness and specific resistance. In those works, a Gaussian probability density function (PDF) is simply considered to express the number fraction of particles in defined diameter ranges. Nevertheless, the Gaussian density has some drawbacks. First, its support ranges from $-\infty$ to ∞ and the density value is greater than zero over the entire domain, which is unrealistic for the particle diameter. Furthermore, the Gaussian PDF shows an even symmetry around the maximum, whereas PSDs typically have a skewed shape (Azema et al., 2002; Levine et al., 1985). This must be taken into account in the development of more reliable models for membrane filtration of municipal wastewater. It is therefore necessary to identify one or more PDFs capable of formalizing the PSDs of municipal wastewaters. The results of the above-mentioned models are expected to be highly influenced by the accuracy of the PSD function implemented.

Despite the remarkable advances made in PSD analysis technologies in recent years, the studies on the use of PDFs to model the PSD of municipal wastewaters are still few. In the papers focused on the black and gray waters, the PDF is assumed to follow the power-law (Adin and Alon, 1993; Alon and Adin, 1994; Arimi, 2018; García-Mesa et al., 2010a,b; Lawler, 1997) or the lognormal models (Kuśnierz and Wiercik, 2016; Oliveira et al., 2012). In particular, García-Mesa et al. (2010a,b) hypothesized a relationship between the two parameters of the power-law distribution and the values of COD and TSS characterizing the wastewater. Broadening the analysis to the urban run-off water, Selbig and Fienen (2012) proposed the Rosin–Rammler PDF in addition to the power-law and lognormal ones. Finally, very recently, Liu and Sansalone (2020) observed that gamma and lognormal PDFs are the best options for modeling the PSD of particulate matter contained in urban waters. Unfortunately, none of the papers specifically addressing wastewater presents a comparative analysis of different PDFs, with the related fitting techniques, to establish which is the best option for modeling the size distribution of such particles.

In this work, on the one hand, we present experimental results of PSD of municipal wastewater obtained through laser light scattering technique. This technique is less hindered by clogging than electroresistance counting methods. Furthermore, it can determine a much wider range of sizes than traditional sieving systems or image capture systems (Dapkunas and Jillavenkatesa, 2001). On the other hand, the experimental study is complemented by a comparative analysis in which many PDFs, both already used and never proposed for wastewater, are fitted to the experimental histograms by a least-squares approach. In particular, the PDFs are compared through appropriate metrics, encompassing both fitting accuracy and statistical consistency with experimental data.

The final aim of this work is to identify PDFs suitable to fit PSD experimental data, to be implemented into membrane filtration models. The filtration models will be used for designing advanced treatment plants for black and gray wastewaters based on membrane filtration. Advanced treatment plants all based on membrane processes are now attracting attention due to the ever-growing attention towards water recovery and possible reuse. Thanks to the new composite materials and membrane configurations, it is now possible to treat fluids with a high content of suspended solids (Hankins and Singh, 2016). Reliable models can be appropriately exploited in the investigation of the most suitable membrane characteristics, the design of the plant configuration, and in choosing the operating parameters. Applications are expected to be relevant in the naval sector, for cruise ships in particular, which are currently responsible for discharging 4 billion liters of black and gray waters, not treated or only partially treated, in the oceans (Rothkopf, 2014). Indeed, the technologies currently available for the onboard treatment of sewage present costs, management complexity, and results that are not entirely satisfactory (Anon, 2020b).

This paper is organized as follows. Section 2 "Experimental" reports the details of the pre-treatment processes carried out on two municipal wastewater samples along with a description of the characterization and PSD analysis performed.

Section 3 "Methods" presents an introduction to the most relevant PDFs being addressed to model the experimental PSDs, the fitting procedure, and the quantitative measures used to assess statistical consistency. Section 4 "Results and Discussion" reports the outcomes of the characterization and PSD analysis; these results are followed by the statistical modeling comparison. Section 5 "Conclusions" summarizes the objectives of this work and presents the conclusions that can be drawn.

2. Experimental

2.1. Pre-treatment filtrations

In the experimental part of this work, characterization and PSD analysis were carried out on municipal wastewater samples, category CER200304 according to the regulation 2014/955/UE, collected in a small municipality, Magliolo, in the province of Savona (Italy). The samples, therefore, consisted of a combination of gray and black waters. The samples were collected directly from the sewerage, and no pre-treatment was applied before that moment, except for a very coarse mechanical filtration with a very large mesh grid (Standard Mesh $\ll$ 3.5, Opening $\gg$ 2 cm). Two different specimens were withdrawn respectively in August 2020 and March 2021. The choice of sampling on two distant dates lies in the seasonality of the effluent composition and concentration due to the weather conditions and the fluctuating resident population. The town is a popular tourist resort that triples the number of inhabitants in summer, the climate is typical of the Ligurian hinterland, Cfb according to the Köppen and Geiger climate classification (Anon, 2020a), with conspicuous rainfall in winter. Based on the above factors, it is reasonable to expect some difference in the size statistics of the particles contained in the two specimens.

The experiments involve pre-treatment filtration of the sewage, of the type that is normally performed to obtain a fluid suitable as a feed for a hypothetical membrane treatment. Preliminary characterization was carried out which highlighted the presence of few but large particles in the fluids. The removal of large particles, larger than 10 μm, is essential to reduce the frequency of membrane module washings and to protect and extend the membrane life. Since membrane processes are high pressure-driven and membranes are characterized by thin sheets and small pores, without adequate pre-filtering steps, such systems can be irreversibly damaged from large particles and fouling can occur. Although few, particles having a large mass could also interfere in the laser diffraction detection performed in this study.

For these reasons, several dead-end filtrations were performed on the sample by using filter bags of decreasing retention sizes: 80 μm, 50 μm, and 25 μm. A last dead-end filtration was performed using an industrial paper filter with a nominal retention size of 10 μm. The liquid resulting from this last stage was the sample for the Microfiltration or Ultrafiltration process.

The overall number of particles blocked by the pre-filters was estimated in an approximated way. *A posteriori*, this number was verified to be significantly lower than the number of particles in the sample.

2.2. Characterization analysis

The two samples retrieved after the pre-treatment steps were analyzed for relative densities, and total suspended solids. Density analysis: UNI EN 13040:2002. Total suspended solids analysis: APAT CNR IRSA 2090 B Man 29 2003.

2.3. Particle size analysis with laser diffraction

The PSD analyses were executed with a Malvern Mastersizer 3000™. This instrumentation exploits the laser diffraction technique (ISO 13320:2020) to provide the particle size distributions of a sample. This method is applied to particle sizes ranging from approximately 0.1 μm to 3.5 mm. Only with special conditions, the applicable size range can be extended above 3.5 mm and below 0.1 μm. In a laser diffraction measurement, a laser beam passes through a suspended particles sample. These particles diffract or scatter the light at certain angles depending on the particle size. Larger particles diffract light at small angles relative to the laser beam and smaller ones diffract light at large angles. The diffracted beam is measured by detectors that assess the angular variation in intensity. The size of the particles is then calculated using the Mie theory of light scattering that converts the light intensity. The particle size is reported as a volume equivalent sphere diameter and the overall particle size distribution is obtained by collecting and analyzing all the diffracted beams (Ryzak and Bieganowski, 2011; Sochan et al., 2014).

The instrument specifically used was able to provide the percent volume-based distribution of particles.

3. Methods

It must be considered that the PSD is a simplified description of the widespread particles in a sample. Practically all particles differ from each other not only in size but also in shape. Particles have random complex and irregular shapes that are generally reduced to "spheres" or "cubes" for convenience. For this reason, the expression "sphere-equivalent diameter" is widely used. This definition given, the size of the measured particle is expressed by the diameter of a spherical element that shows the same response to a certain sizing measurement technique. Therefore, the measurement of a PSD

depends to some extent upon the physical principle of the chosen procedure because a given measurement technique evaluates a physical property of the particle that can be related to its size (Herdan and Smith, 1960).

The most common graphical representation of a PSD is the histogram. Indeed, the size distribution is commonly obtained into a certain number of size classes. The histogram provides a good portrait of how the particles are distributed among the various sizes, where the width of each rectangle represents the considered size interval, and the height represents the number of particles in that interval. To make it possible to compare multiple PSDs a normalization is generally performed by dividing the values of each interval by the total number of counted particles. The entire size range is summarized into a series of contiguous particle size intervals and the relative amount of particles belonging to each interval is assigned (Hinds, 1999).

The ideal situation would be to obtain a histogram with regular size intervals. Unfortunately, in many cases, the analytical sizer instrument provides inhomogeneous size widths and this affects the representation of the size distribution because the height of any interval is dependent on the width of that interval.

For example, supposing to double a given interval width could result in obtaining a height roughly twice that of the initial interval. To address this issue, the data histogram is normalized by dividing the number of particles in each interval by the total number of particles and by the width of that interval, providing a function with a unitary area, in which the heights of the columns with different widths are comparable (Raabe, 1971).

Such a normalized histogram represents a discretized experimental PDF. On one hand, it is a legitimate PDF *per se* (nonnegative and with unitary integral), although it describes the behavior of the statistics of the observation in a piecewise constant manner. On the other hand, it can be used to identify underlying continuous PDFs, defined by parametric equations, suitable for modeling the PSD. This task requires a fitting procedure that estimates, for each considered continuous PDF, the values of the involved parameters. The description of PSDs through a PDF is of paramount importance to enable comparison between different distributions, to perform statistical analyses, and to implement physically-based models of treatment plants (e.g., to optimize the design of physical–chemical pre-treatment and final membrane wastewater processes). For these reasons, several PDFs have been proposed over the years to mathematically model the PSD experimental observations, depending on the nature of the examined particles (Devore and Berk, 2012).

3.1. PDFs

This section provides a brief introduction to the PDFs (Johnson et al., 1994) most frequently adopted to model PSDs, including those specifically used in previous literature to model particle distribution in municipal wastewater, and those that are proposed in this paper, jointly with a fitting method, for the same purpose. All considered PDFs are meant to characterize non-negative random variables, so their supports are either $[0, +\infty)$ or subsets of this interval.

3.1.1. Power-law function and generalized Pareto distribution

Some previous papers (Adin and Alon, 1993; Alon and Adin, 1994; Arimi, 2018; García-Mesa et al., 2010a,b; Lawler, 1997) used a power-law function to model the particle size distribution in wastewater:

$$\widetilde{f_X}(x) = Ax^{-\beta}, \quad \beta > 0, A > 0, x > 0 \tag{1}$$

where x is the particle size, A and β are empirical constants and $\widetilde{f_X}$ represents a non-normalized PDF, i.e., a nonnegative function that models some partial behavior of the probability distribution of the random variable X (e.g., its asymptotic behavior for $x \to +\infty$) but whose integral over $[0, +\infty)$ is generally non-unitary. As discussed in Johnson et al. (1994), the normalized version of this function is the Pareto distribution of the first kind:

$$f_X(x) = ak^a x^{-(a+1)}, \quad k > 0,\ a > 0,\ x \geq k \tag{2}$$

where a and k are, respectively, the shape and scale parameters of the function, and f_X represents the PDF of an absolutely continuous random variable X. A more general expression, allowing a continuous range of possible shapes that includes the Pareto distribution as a special case, is the generalized Pareto PDF:

$$f_X(x) = \frac{1}{\sigma}\left[1 + \frac{\xi(x-\mu)}{\sigma}\right]^{\left(-\frac{1}{\xi}-1\right)}, \quad \sigma > 0, \mu \in \mathbb{R}, \xi \neq 0 \tag{3}$$

where ξ, σ and μ are, respectively, the shape, scale, and location parameters of the function. The support is $x \geq \mu$ when $\xi > 0$, and $\mu \leq x \leq \mu - \sigma/\xi$ when $\xi < 0$. Given the aforementioned focus on non-negative random variables, the additional condition $\mu \geq 0$ is accepted.

3.1.2. Two- and three-parameter lognormal distribution

The lognormal PDF has been widely adopted to model PSDs (Liu and Sansalone, 2020) and, more recently, has been specifically used to model the wastewater PSD as well (Kuśnierz and Wiercik, 2016; Oliveira et al., 2012). It is defined as:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \quad \sigma > 0,\ \mu \in \mathbb{R}, x > 0 \tag{4}$$

where σ and μ are, respectively, the standard deviation and the mean of the natural logarithm of the considered random variable X.

It is possible to introduce an additional parameter to translate the PDF along the x-axis, moving from the two-parameter to the three-parameter lognormal distribution:

$$f_X(x) = \frac{1}{(x-\theta)\,\sigma\sqrt{2\pi}} \exp\left\{-\frac{[\ln(x-\theta)-\mu]^2}{2\sigma^2}\right\}, \quad \sigma > 0,\ \mu \in \mathbb{R}, x > \theta \tag{5}$$

where θ is the location parameter . The condition $\theta \geq 0$ is accepted to ensure again that X is non-negative.

3.1.3. Rosin-Rammler distribution

A very common PDF used to model PSDs is the Rosin–Rammler distribution (Bayat et al., 2015; Liu and Sansalone, 2020; Selbig and Fienen, 2012), also known as the Weibull distribution, which includes the exponential and Rayleigh PDFs as special cases. It can be written as follows:

$$f_X(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{x}{\lambda}\right)^{k}\right], \quad k > 0,\ \lambda > 0,\ x > 0 \tag{6}$$

where k and λ are the shape and scale parameters, respectively.

3.1.4. Gamma distribution

Another PDF proposed to describe the PSD in different fields is the gamma distribution (Liu and Sansalone, 2020; Piro et al., 2010; Sansalone and Ying, 2008). Using a shape-scale parametrization, it can be defined as:

$$f_X(x) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} \exp\left(-\frac{x}{\theta}\right), \quad k > 0,\ \theta > 0,\ x > 0 \tag{7}$$

where k and θ are the shape and scale parameters, respectively, and $\Gamma(k)$ is the gamma function.

3.1.5. Burr distribution

The Burr Type XII distribution is a very flexible family of functions that fit a wide range of empirical data and makes it possible to express a broad spectrum of distribution shapes, lying between two asymptotic limiting cases: the Rosin–Rammler distribution and the Pareto distribution of the first kind (Johnson et al., 1994). The three-parameter version of this PDF can be written as follows:

$$f_X(x) = \frac{ck}{\lambda}\left(\frac{x}{\lambda}\right)^{c-1}\left[1+\left(\frac{x}{\lambda}\right)^{c}\right]^{-k-1}, \quad k > 0,\ \lambda > 0,\ c > 0,\ x > 0 \tag{8}$$

where c and k are the shape parameters, and λ is the scale parameter of the function.

3.2. PDF fitting procedure

To find the parameter values that best fit a given PDF to the normalized histogram of the experimental PSD, a least-squares problem can be defined and solved by an appropriate algorithm. Let D_i be the center of the ith discretized bin in the aforementioned normalized histogram of the experimental particle diameters, $i = 1, 2, \ldots, N$, where N is the total number of bins, and let h_i be the height of the normalized histogram bar centered in D_i. The fitting can be performed by finding the parameter vector $\boldsymbol{\phi}_{opt}$ that solves the following least-squares problem:

$$E^2 = \min_{\boldsymbol{\phi}\in\Phi}\left\{\sum_{i=1}^{N}\left[f_X(D_i|\boldsymbol{\phi}) - h_i\right]^2\right\} \tag{9}$$

where $\boldsymbol{\phi}$ is the vector containing the parameters of the PDF – which is conveniently denoted here as $f_X(\cdot|\boldsymbol{\phi})$ to emphasize its dependence on $\boldsymbol{\phi}$ – and E^2 is the minimum square distance between the PDF and the normalized histogram. This square distance is attained when $\boldsymbol{\phi} = \boldsymbol{\phi}_{opt}$. For each considered PDF, the parameter vector $\boldsymbol{\phi}$ takes values in a domain $\Phi \subset \mathbb{R}^P$, where P is the number of parameters of the PDF and Φ is determined by the conditions discussed in the previous section (e.g., $\Phi = \boldsymbol{\phi} = (\sigma, \mu) : \sigma > 0, \mu \in \mathbb{R}$ for the log-normal PDF). We emphasize that, in Eq. (9), both $f_X(D_i|\boldsymbol{\phi})$ and h_i $(i = 1, 2, \ldots, N)$ are samples of PDFs, although with different analytical behaviors: a smooth PDF in the case of $f_X(\cdot|\boldsymbol{\phi})$ and a piecewise constant function in the case of the normalized histogram. Indeed, this histogram provides an empirical nonparametric approximation of the PSD through a staircase function, which is inconvenient whenever a smooth model is necessary and – most critically – may strongly depend on the employed specific realization of data samples. From this perspective, the model-fitting problem addressed through Eq. (9) can be thought of as a regularization task based on the assumption of a smooth parametric family $f_X(\cdot|\boldsymbol{\phi})$ underlying the PSD behavior.

To solve the optimization problem in Eq. (9), a nonlinear algorithm should be used: the trust-region-reflective algorithm, described by Coleman and Li (1996), has been chosen because it is widely adopted as a reliable option, especially when the equation system is not underdetermined (i.e., when $P \leq N$), as in the case considered here.

In addition to the value of E^2, the similarity between the normalized histogram and the fitted PDF can be evaluated by the (sample) linear correlation coefficient, ρ, $\rho \in [-1, 1]$, defined by the Pearson's equation (Ott and Longnecker, 2016) as follows:

$$\rho = \frac{\sum_{i=1}^{N} \left(f_X\left(D_i|\boldsymbol{\phi}_{opt}\right) - \overline{f_X}\right)\left(h_i - \overline{h}\right)}{\sqrt{\sum_{i=1}^{N} \left(f_X\left(D_i|\boldsymbol{\phi}_{opt}\right) - \overline{f_X}\right)^2 \sum_{i=1}^{N} \left(h_i - \overline{h}\right)^2}} \tag{10}$$

where $\overline{f_X}$ and $\overline{h}$ are, respectively, the sample means of $f_X\left(D_i|\boldsymbol{\phi}_{opt}\right)$ and h_i, $i = 1, \ldots, N$.

3.3. Statistical measures of fit for the comparison of PDF models

The statistic computed by various hypothesis tests of whether a data set, made of m samples, is drawn from a given distribution can be used to assess the consistency between a set of wastewater particles and a given PDF, fitted to the normalized histogram by the procedure described above, provided that the mentioned wastewater particles were not included in the histogram used for the fitting of the PDF. Three statistics have been selected for this purpose (Anderson and Darling, 1954; Papoulis and Pillai, 2015) and are briefly recalled in this Section. They represent quantitative measures of the difference between a fitted PDF model and the empirical distribution of a test set. In particular, in all the cases, we assume that the set of the N samples used to estimate each PDF model and the set of the m test samples used to compute the statistic are disjoint.

3.3.1. Kolmogorov–Smirnov statistic

This statistic compares the hypothesized cumulative distribution function (CDF, i.e., the integral of the hypothesized PDF), $F_X(x)$, and the empirical CDF, $\hat{F}(x)$, computed using the available m data samples (Papoulis and Pillai, 2015). It is expressed as the L^∞ distance between these two CDFs:

$$S_{KS} = \sup_{x \geq 0} \left|\hat{F}(x) - F_X(x)\right| \tag{11}$$

where sup stands for supremum. According to this equation, the statistic measures the largest absolute difference between the two CDFs, across all x values.

3.3.2. Anderson–Darling statistic

In this case, the statistics measure the quadratic distance between the two CDFs, weighted by a function that emphasized the distances in the tails of the distribution (Anderson and Darling, 1954). The statistic is defined by the following Lebesgue–Stieltjes integral:

$$S_{AD} = m \int_0^\infty \frac{\left(\hat{F}(x) - F_X(x)\right)^2}{F_X(x)\left(1 - F_X(x)\right)} dF_X(x) \tag{12}$$

Computationally, it is evaluated by numerical integration.

3.3.3. Chi-squared statistic

In this case, the m wastewater particles composing the data sample, are classified into k adjacent, non-overlapping intervals, spanning the diameter domain. For each interval, the expected number of particles, computed through the hypothesized PDF, is compared with the number of particles present in it, as defined in the Pearson's test statistic (Papoulis and Pillai, 2015):

$$S_{C2} = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \tag{13}$$

where e_i and o_i are the expected and observed number of particles into the ith interval, respectively, and are such that:

$$\sum_{i=1}^{k} o_i = \sum_{i=1}^{k} e_i = m \tag{14}$$

4. Results and discussion

4.1. Characterization and PSD analysis

In this Section, the results regarding the characterization and PSD analysis are reported. Pre-treatment filtrations were performed as a preliminary step, as outlined in Section 2.1. Data regarding the test carried out in August 2020 are labeled

Table 1
Density and TSS measurements for the two specimens.

Property	Sample I	Sample II
Density [g/L]	999	1002
TSS [mg/L]	312	200

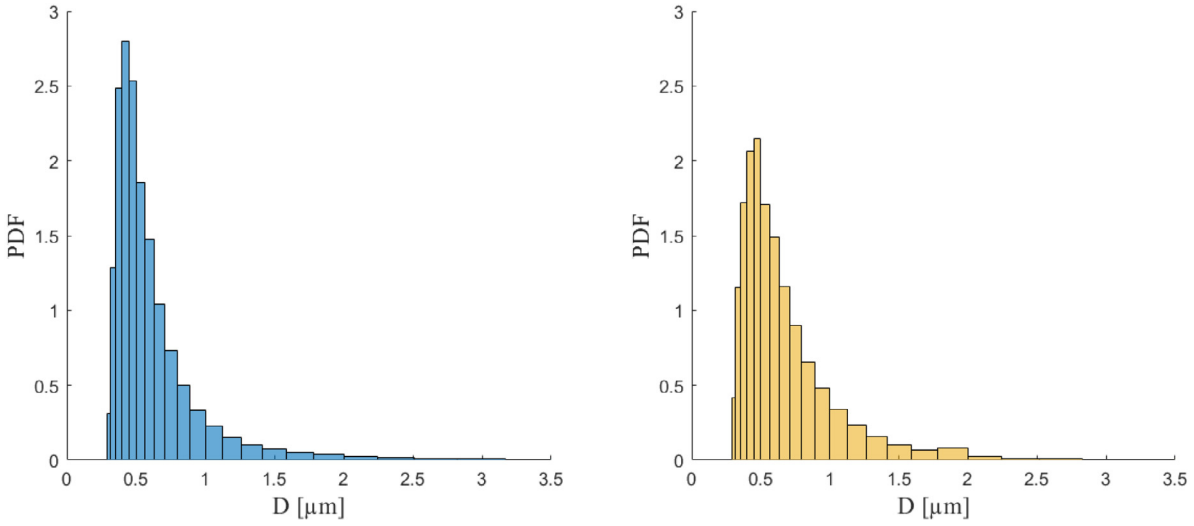

Fig. 1. Normalized histograms of the sphere-equivalent diameter, D, for (left) Sample I and (right) Sample II.

as Sample I, those of March 2021 are labeled as Sample II. Table 1 shows the measurements of the mass density and TSS.

Significant differences can be observed between the concentration of suspended solids in the effluents collected in the summer and winter periods. In the summer season, which is characterized by less rain and a higher resident population, the TSS is higher.

The data collected by the instrument, analyzing Sample I and Sample II, were organized in normalized histograms, applying the procedure described in Section 3. As already mentioned in Section 3, these normalized histograms represent discretized experimental PDFs. The histograms, depicted as a function of the sphere-equivalent diameter, D, are shown in Fig. 1. Although the two histograms are similar in their general shape, non-marginal differences are also evident: while the histogram of Sample I has a higher peak, that of Sample II shows a more gradual descent for diameters greater than those at which the peak is located. The histogram differences between the two specimens are reflected in their mean values and standard deviations: Sample I particles have a sample mean diameter of 0.68 μm and a sample standard deviation of 0.54 μm; the particles of Sample II have a sample mean diameter of 0.74 μm and a sample standard deviation of 0.51 μm.

4.2. Statistical modeling comparison

The results of the wastewater PSD described above have been used to fit different PDFs through the procedure depicted in Section 3.2 and, in turn, the fitted PDFs have been quantitatively compared through the statistics illustrated in Section 3.3. The considered distributions included those mentioned in Section 3.1 as well as many other traditional PDFs. Here, the focus is on the PDFs described in Section 3.1 because they are the ones proposed to model the wastewater particle size or most commonly adopted in the field of PSD, or because they were the best at modeling the experimental data considered in this paper.

The number of particles randomly selected and employed for computing S_{KS}, S_{AD}, and S_{C2}, and not included in the histogram used for the PDF fitting, has been 10^4 out of more than 10^6 particles composing the two original data samples (i.e., Sample I and Sample II).

Fig. 2 shows the normalized histogram of the experimental PSD in comparison with the least-squares fitting of the functions previously used to model the wastewater particles size, i.e., the power-law curve and the two-parameter lognormal PDF. To transform the power-law curve into a PDF, while maintaining the maximum degrees of freedom, the generalized Pareto PDF has been adopted. The results of the least-squares fitting for the Rosin–Rammler and Gamma PDFs are shown in Fig. 3, while Fig. 4 shows the results for the three-parameter lognormal and Burr PDFs. The values of the minimum square distance, E^2, and correlation coefficient, ρ, for the mentioned PDFs and the two data samples are reported in Table 2.

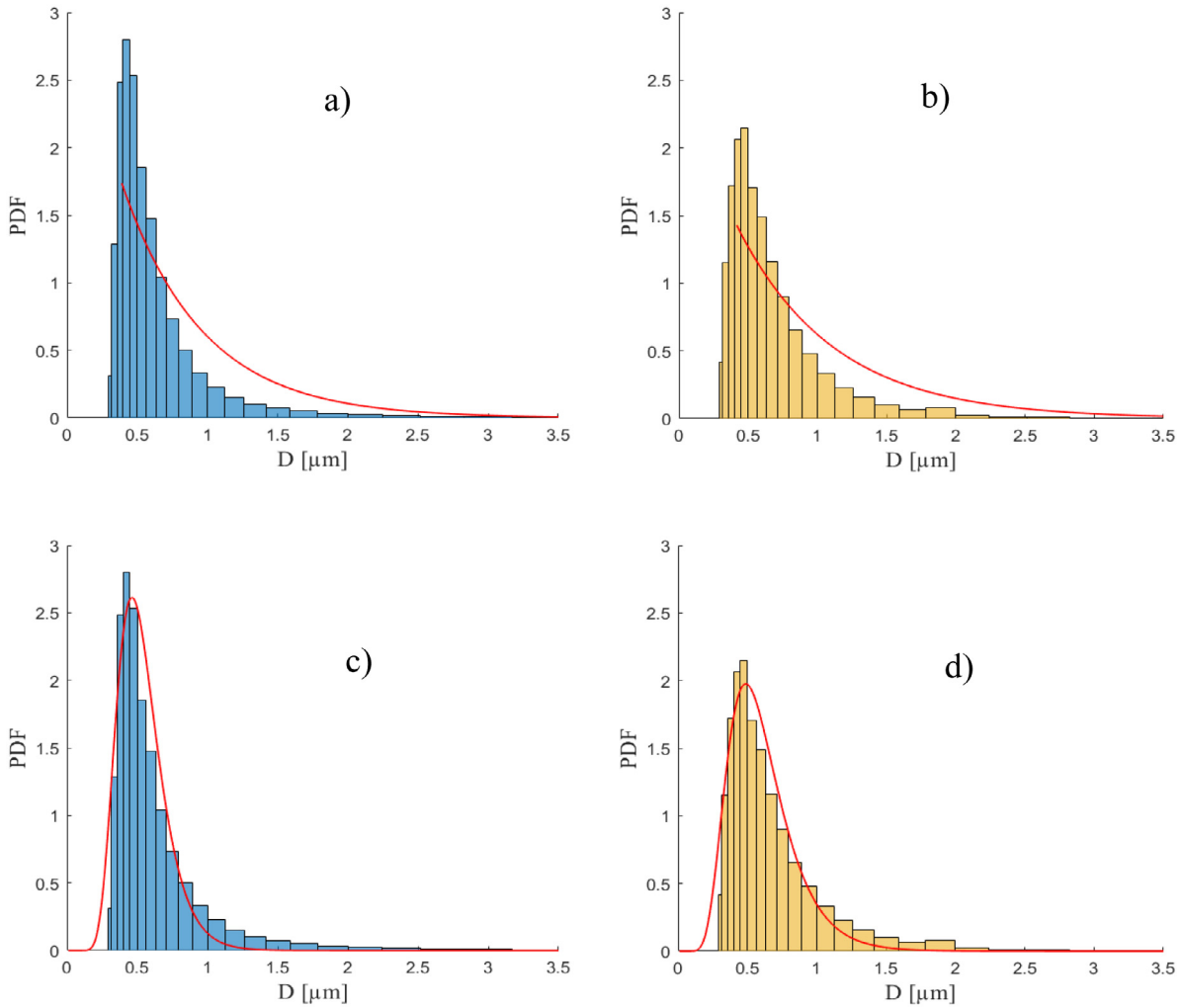

Fig. 2. Least-squares fitting of (a, b) generalized Pareto PDF and (c, d) two-parameter lognormal PDF with the experimental PSD normalized histogram of (a, c) Sample I; (b, d) Sample II.

Table 2
PDF fitting results: minimum square distance, E^2, and correlation coefficient, ρ, for each considered distribution and the two data samples. 2P and 3P stand for two- and three-parameter, respectively.

		Gener. Pareto	2P lognormal	Rosin–Ramm.	Gamma	3P lognormal	Burr
E^2	Sample I	7.03	1.46	4.16	1.86	0.07	0.10
	Sample II	3.67	0.60	1.75	0.85	0.03	0.13
ρ	Sample I	0.833	0.971	0.905	0.963	0.998	0.998
	Sample II	0.874	0.982	0.942	0.974	0.999	0.996

Finally, the consistency between the two specimens of wastewater particles and a given PDF has been assessed by computing and comparing the three statistics described in Section 3.3. To this end, the 10^4 particles picked out from each original data sample have been randomly subdivided into 100 sets of $m = 100$ particles each. For each statistic, the related value is computed for each particle set and, in the end, the 100 obtained values are averaged. In particular, for the chi-squared statistic, the number k of intervals has been set equal to 10. The averages of the statistics for the two specimens are reported in Table 3.

Looking at the results of the least-squares fitting in Figs. 2 and 3 and observing the values of E^2 and ρ in Table 2, it is evident that the generalized Pareto and Rosin–Rammler are PDFs that fail to approximate the histogram, while the two-parameter lognormal and gamma PDFs succeed better in fitting the histogram profile, although significant discrepancies remain, especially in the tails.

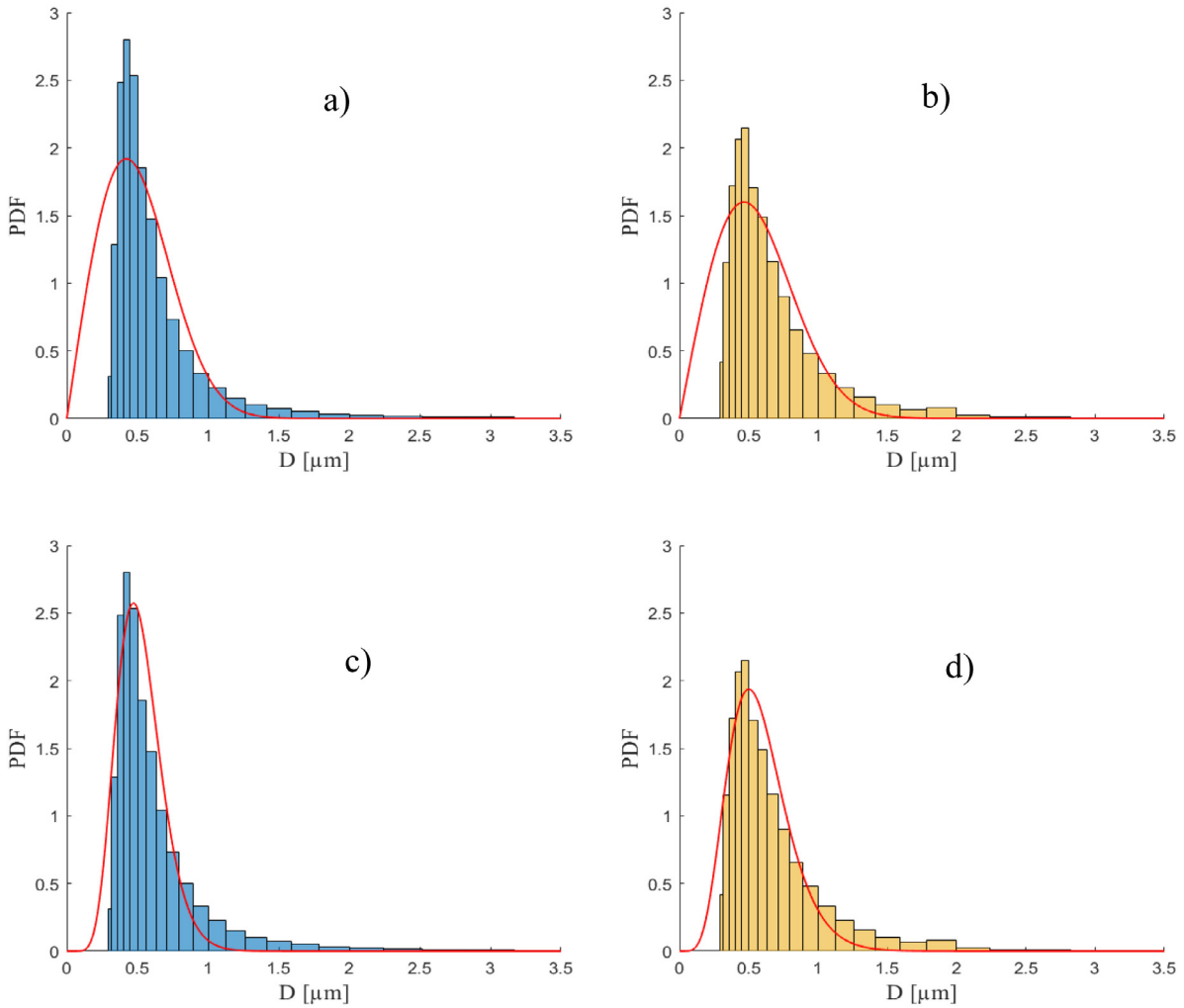

Fig. 3. Least-squares fitting of the (a, b) Rosin–Rammler PDF and (c, d) gamma PDF with the experimental PSD normalized histogram of (a, c) Sample I; (b, d) Sample II.

Table 3
Consistency between wastewater particles and the considered PDFs: averages of the Kolmogorov–Smirnov (S_{KS}), Anderson–Darling (S_{AD}) and chi-squared statistics (S_{C2}). For each data sample, 100 sets of 100 randomly chosen particles each are used; 10 intervals are used in the chi-squared statistic. 2P and 3P stand for two- and three-parameter, respectively.

		Gener. Pareto	2P lognormal	Rosin–Ramm.	Gamma	3P lognormal	Burr
S_{KS}	Sample I	0.37	0.18	0.24	0.20	0.13	0.13
	Sample II	0.36	0.16	0.20	0.18	0.12	0.12
S_{AD}	Sample I	14.45	7.50	7.21	8.18	1.89	1.54
	Sample II	17.50	5.63	6.94	7.86	1.42	1.59
S_{C2}	Sample I	50.07	29.90	10.33	72.19	5.34	4.79
	Sample II	54.41	20.99	30.29	60.10	6.37	7.60

Fig. 4 shows that the least-squares fitting of the three-parameter lognormal and Burr PDFs provides a very close matching between the normalized histogram and the density functions, at both head and tails. Embracing both data samples, the E^2 values for these two PDFs are equal to or less than 0.13 (i.e., at least 5 times lower than those for the other PDFs) and the correlation coefficient exceeds 0.99 for both these PDFs.

The results of fitting between histogram and density functions are confirmed by the values of the three considered statistics. For Sample I, in the case of the Kolmogorov–Smirnov distance, both the three-parameter lognormal and Burr PDFs reported a statistic of 0.13, while the best of the remaining PDFs is the two-parameter lognormal with a statistic of

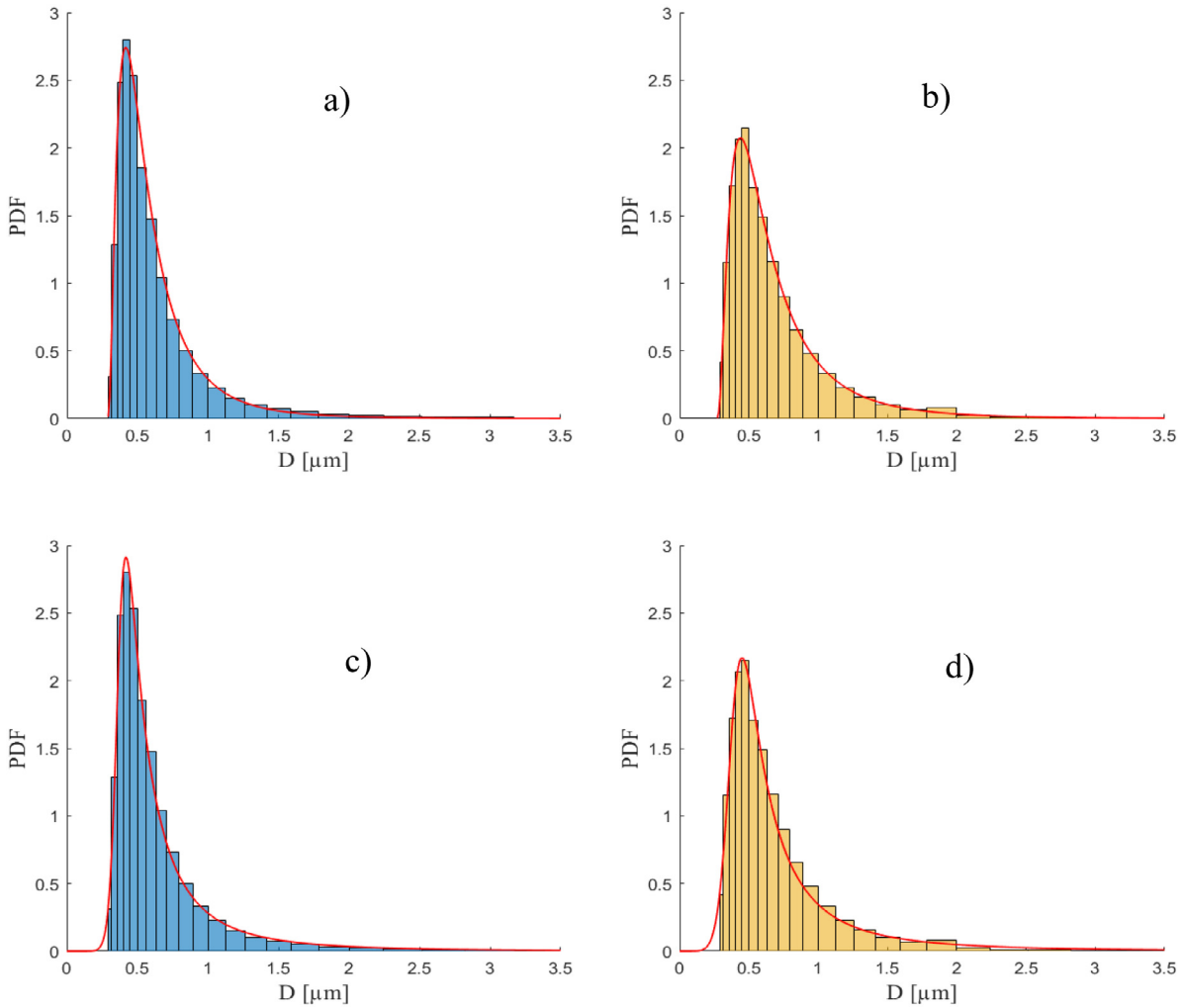

Fig. 4. Least-squares fitting of the (a, b) three-parameter lognormal PDF and (c, d) Burr PDF with the experimental PSD normalized histogram of (a, c) Sample I; (b, d) Sample II.

0.18. These values become 0.12 and 0.16, respectively, when Sample II is considered. In the case of the Anderson–Darling statistic, which pays more attention to function tails, the superiority of the three-parameter lognormal and Burr PDFs is more pronounced for both data samples: their statistics are at least 3.5 times lower than that of the best among the remaining PDFs (the Rosin–Rammler for Sample I and the two-parameter lognormal for Sample II). Also for the chi-squared statistics, when the least-squares fitting is applied the Rosin–Rammler (for Sample I) and the two-parameter lognormal (for Sample II) are the best PDFs among the four classical functions considered. Like in the case of the other two functionals, the statistic values for the three-parameter lognormal and Burr PDFs are close to each other and, in this case, at least two times lower than those of the classical PDFs.

On one hand, the results of S_{KS}, S_{AD}, and S_{C2}, which are computed on the test set with regard to the four classical PDFs, are in partial disagreement with those of E^2 and ρ, which are calculated on the samples used to fit the models. On the other hand, all such measures agree in identifying the three-parameter lognormal and Burr PDFs as the best to model, joined with the least-squares fitting, the samples under consideration. This conclusion holds for both Sample I and Sample II.

5. Conclusions

The final aim of this work is to identify and evaluate the best options for the mathematical modeling of the PSD of municipal wastewaters. The statistical modeling of the wastewater particles by the PDFs already used for this purpose and by the other PDFs widely adopted for modeling the PSDs of different materials is investigated using the trust-region-reflective algorithm to solve the least-squares fitting problem. The outcome is that the generalized Pareto (which includes

the power-law distribution), lognormal, Rosin–Rammler, and gamma PDFs are not suited to model the two specimens of wastewater particles analyzed in this paper. However, the introduction of a third parameter in the lognormal PDF, aimed at translating the function along the x-axis, allows the least-squares optimization to provide a very accurate fitting of this density with the normalized histograms of the particle size. Among the many PDFs considered, the Burr density also obtained a fitting whose error and correlation values are comparable with those of the three-parameter lognormal. In addition, the Kolmogorov–Smirnov, Anderson–Darling, and chi-squared statistics showed that the three-parameter lognormal and Burr PDFs, fitted to the experimental PSD data through the least-squares procedure, are significantly more consistent with a random set of wastewater particles than the four classical PDFs mentioned above. This makes it possible to conclude that the three-parameter lognormal and Burr PDFs can be successfully fitted to the considered data and, then, used to model the size of particles of both the wastewater specimens considered. The identified PDFs can be applied in conjunction with classical membrane models, such as those proposed by Foley et al. (1995), and Broeckmann et al. (2006), in order of achieving accurate modeling of membrane treatment of municipal wastewaters. The development of these membrane plants, aiming at water recuperation, is now at the cutting edge of research. An interesting application is under study in the naval sector, for cruise ships in particular, which are currently responsible for discharging 4 billion liters of black and gray waters, not treated or only partially treated, in the oceans (Rothkopf, 2014). Reliable models can be appropriately exploited in the investigation of the most suitable membrane characteristics, the design of the plant configuration, and in choosing the operating parameters.

CRediT authorship contribution statement

Matteo Cornacchia: Investigation, Data curation, Writing – original draft. **Gabriele Moser:** Formal analysis, Validation. **Ezio Saturno:** Resources. **Andrea Trucco:** Methodology, Visualization. **Paola Costamagna:** Conceptualization, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

M. C. acknowledges financial support from ‘Programma Operativo Por FSE Regione Liguria 2014-2020, Italy, assegno di ricerca “RLOF18ASSRIC/31/1” under the supervision of P.C. and E.S.

References

Adin, A., Alon, G., 1993. The role of particle characterization in advanced wastewater treatment. Water Sci. Technol. 27, 131–139. http://dx.doi.org/10.2166/wst.1993.0219.

Alon, G., Adin, A., 1994. Mathematical modeling of particle size distribution in secondary effluent filtration. Water Environ. Res. 66, 836–841. http://dx.doi.org/10.2175/WER.66.6.11.

Anderson, T.W., Darling, D.A., 1954. A test of goodness of fit. J. Am. Stat. Assoc. 49, 765–769. http://dx.doi.org/10.1080/01621459.1954.10501232.

Anon, 2020a. Clima Magliolo [WWW Document]. https://it.climate-data.org/europa/italia/liguria/magliolo-114140/. (Accessed 17 May 2022).

Anon, 2020b. Open innovation city Hackathon blue [WWW Document]. http://www.hackathongenova.it/. (Accessed 13 September 2021).

Arimi, M.M., 2018. Particle size distribution as an emerging tool for the analysis of wastewater. Environ. Technol. Rev. 7, 274–290. http://dx.doi.org/10.1080/21622515.2018.1540666.

Azema, N., Pouet, M.-F., Berho, C., Thomas, O., 2002. Wastewater suspended solids study by optical methods. Colloids Surf. A 204, 131–140. http://dx.doi.org/10.1016/S0927-7757(02)00006-7.

Bayat, H., Rastgo, M., Mansouri Zadeh, M., Vereecken, H., 2015. Particle size distribution models, their characteristics and fitting capability. J. Hydrol. 529, 872–889. http://dx.doi.org/10.1016/j.jhydrol.2015.08.067.

Broeckmann, A., Busch, J., Wintgens, T., Marquardt, W., 2006. Modeling of pore blocking and cake layer formation in membrane filtration for wastewater treatment. Desalination 189, 97–109. http://dx.doi.org/10.1016/j.desal.2005.06.018.

Chavez, A., Jimenez, B., Maya, C., 2004. Particle size distribution as a useful tool for microbial detection. Water Sci. Technol. 50, 179–186. http://dx.doi.org/10.2166/wst.2004.0119.

Chavez, A., Maya, C., Jimenez, B., 2006. Particle size distribution to design and operate an APT process for agricultural wastewater reuse. Water Sci. Technol. 53, 43–49. http://dx.doi.org/10.2166/wst.2006.206.

Coleman, T.F., Li, Y., 1996. An Interior Trust Region approach for nonlinear minimization subject to bounds. SIAM J. Optim. 6, 418–445. http://dx.doi.org/10.1137/0806023.

Dapkunas, S.J., Jillavenkatesa, A., 2001. NIST Recommended Practice Guide. Gaithersburg, MD. http://dx.doi.org/10.6028/NBS.SP.960-1.

Devore, J.L., Berk, K.N., 2012. Modern Mathematical Statistics with Applications, the American Statistician, Springer Texts in Statistics. Springer New York, New York, NY, http://dx.doi.org/10.1007/978-1-4614-0391-3.

Dulekgurgen, E., Doğruel, S., Karahan, Ö., Orhon, D., 2006. Size distribution of wastewater COD fractions as an index for biodegradability. Water Res. 40, 273–282. http://dx.doi.org/10.1016/j.watres.2005.10.032.

Foley, G., Malone, D.M., MacLoughlin, F., 1995. Modelling the effects of particle polydispersity in crossflow filtration. J. Membr. Sci. 99, 77–88. http://dx.doi.org/10.1016/0376-7388(94)00207-F.

García-Mesa, Juan José, Poyatos, J.M., Delgado, F., Hontoria, E., 2010a. The influence of biofilm treatment systems on particle size distribution in three wastewater treatment plants. Water Air Soil Pollut. 212, 37–49. http://dx.doi.org/10.1007/s11270-009-0320-5.

García-Mesa, Juan J., Poyatos, J.M., Delgado-Ramos, F., Muñio, M.M., Osorio, F., Hontoria, E., 2010b. Water quality characterization in real biofilm wastewater treatment systems by particle size distribution. Bioresour. Technol. 101, 8038–8045. http://dx.doi.org/10.1016/j.biortech.2010.05.071.
Hankins, N.P., Singh, R., 2016. Introduction to membrane processes for water treatment. In: Emerging Membrane Technology for Sustainable Water Treatment. Elsevier, pp. 15–52. http://dx.doi.org/10.1016/C2012-0-07949-5.
Herdan, G., Smith, M.L., 1960. The distributions of particle size and their averages. In: Small Particles Statics: An Account of Statistical Methods for the Investigation of Finely Divided Materials. Butterworths Scientific Publications, pp. 29–40.
Hinds, W.C., 1999. Particle size statics. In: Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles. John Wiley & Sons, pp. 75–110.
Hocaoglu, S.M., Orhon, D., 2013. Particle size distribution analysis of chemical oxygen demand fractions with different biodegradation characteristics in black water and gray water. CLEAN - Soil, Air, Water 41, 1044–1051. http://dx.doi.org/10.1002/clen.201100467.
Jillavenkatesa, A., Dapkunas, S.J., Lum, L.-S.H., 2001. Particle Size Characterization, Special Pu. ed. National Institute of Standards and Technology.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, vol. 1, second ed. John Wiley & Sons.
Kuśnierz, M., Wiercik, P., 2016. Analysis of particle size and fractal dimensions of suspensions contained in raw sewage, treated sewage and activated sludge. Arch. Environ. Prot. 42, 67–76. http://dx.doi.org/10.1515/aep-2016-0031.
Lawler, D.F., 1997. Particle size distributions in treatment processes: Theory and practice. Water Sci. Technol. 36, 15–23. http://dx.doi.org/10.1016/S0273-1223(97)00414-9.
Levine, A.D., Tchobanoglous, G., Asano, T., 1985. Characterization of the size distribution of contaminants in wastewater: Treatment and reuse implications. J. Water Pollut. Control Fed. 57, 805–816.
Liu, Y., Sansalone, J.J., 2020. Physically-based particle size distribution models of urban water particulate matter. Water Air Soil Pollut. 231, 555. http://dx.doi.org/10.1007/s11270-020-04925-z.
Oliveira, S.C., Souki, I., von Sperling, M., 2012. Lognormal behaviour of untreated and treated wastewater constituents. Water Sci. Technol. 65, 596–603. http://dx.doi.org/10.2166/wst.2012.899.
Ott, R.L., Longnecker, M., 2016. An Introduction to Statistical Methods & Data Analysis, seventh ed. Cengage Learning.
Papoulis, A., Pillai, S.U., 2015. Probability, Random Variables and Stochastic Processes. McGraw-Hill.
Piro, P., Carbone, M., Garofalo, G., Sansalone, J., 2010. Size distribution of wet weather and dry weather particulate matter entrained in combined flows from an urbanizing sewershed. Water Air Soil Pollut. 206, 83–94. http://dx.doi.org/10.1007/s11270-009-0088-7.
Raabe, O.G., 1971. Particle size analysis utilizing grouped data and the log-normal distribution. J. Aerosol Sci. 2, 289–303. http://dx.doi.org/10.1016/0021-8502(71)90054-1.
Rickert, D.A., Hunter, J.V., 1971. General nature of soluble and particulate organics in sewage and secondary effluent. Water Res. 5, 421–436. http://dx.doi.org/10.1016/0043-1354(71)90005-4.
Rothkopf, J., 2014. Cruise ships dumped over 1 billion gallons of untreated waste into the oceans [WWW document]. https://www.salon.com/2014/12/05/cruise_ships_dumped_over_1_billion_gallons_of_untreated_waste_into_the_oceans_this_year/. (Accessed 10 February 2022).
Ryzak, M., Bieganowski, A., 2011. Methodological aspects of determining soil particle-size distribution using the laser diffraction method. J. Plant Nutr. Soil Sci. 174, 624–633. http://dx.doi.org/10.1002/jpln.201000255.
Sanchis, M.I.A., Sáez, J., Lloréns, M., Soler, A., Ortuño, J.F., 2003. Particle size distribution in slaughterhouse wastewater before and after coagulation-flocculation. Environ. Prog. 22, 183–188. http://dx.doi.org/10.1002/ep.670220316.
Sansalone, J., Ying, G., 2008. Partitioning and granulometric distribution of metal leachate from urban traffic dry deposition particulate matter subject to acidic rainfall and runoff retention. Water Res. 42, 4146–4162. http://dx.doi.org/10.1016/j.watres.2008.06.013.
Selbig, W.R., Fienen, M.N., 2012. Regression modeling of particle size distributions in urban storm water: Advancements through improved sample collection methods. J. Environ. Eng. 138, 1186–1193. http://dx.doi.org/10.1061/(ASCE)EE.1943-7870.0000612.
Sochan, A., Polakowski, C., Łagód, G., 2014. Impact of optical indices on particle size distribution of activated sludge measured by laser diffraction method. Ecol. Chem. Eng. S 21, 137–145. http://dx.doi.org/10.2478/eces-2014-0012.