

Exploring stochastic differential equation for analyzing uncertainty in wastewater treatment plant-activated sludge modeling

Reza Shahidi Zonouz^a, Vahid Nourani^{a,b,d}, Mina Sayyah-Fard^{a,*}, Huseyin Gokcekus^b and Chang-Qing Ke^c

^a Center of Excellence in Hydroinformatics and Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

^b Faculty of Civil and Environmental Engineering, Near East University, Nicosia, Cyprus

^c School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China

^d College of Engineering, IT & Environment, Charles Darwin University, Darwin, Australia

*Corresponding author. E-mail: m.sayyahfard@tabrizu.ac.ir

ABSTRACT

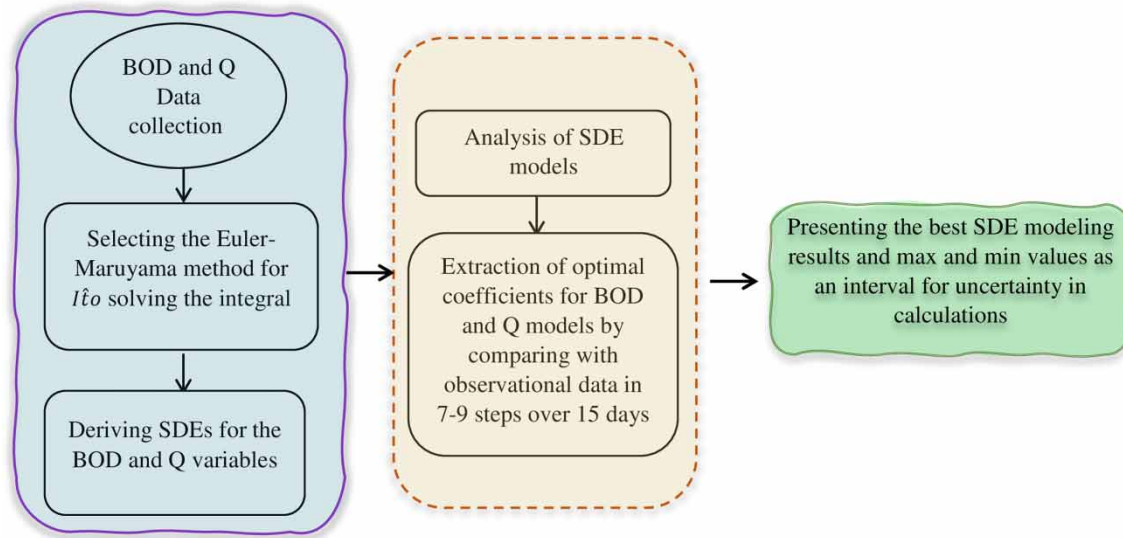
The management of wastewater treatment plant (WWTP) and the assessment of uncertainty in its design are crucial from an environmental engineering perspective. One of the key mechanisms in WWTP operation is activated sludge, which is related to the biological oxygen demand (BOD) parameter. In the modeling of BOD, the conventional approach utilizing ordinary differential equations (ODEs) fails to incorporate the stochastic nature of this parameter, leading to a considerable level of uncertainty in the design of WWTP. To address this issue, this study proposes a stochastic model that utilizes stochastic differential equations (SDEs) instead of ODE to simulate BOD activities of microorganisms and wastewater flow rate (Q). The SDEs and integral \int_0^t are solved using the Euler–Maruyama method for a period of 15 sequential days and the timespan of 2019–2020 for a WWTP in Tabriz City. SDE improves the design of WWTP facilities by identifying uncertainties and enhancing reliability. This, in turn, increases the reliability of the technical structures within the WWTP and improves the performance of its biological system. By considering the randomness of the problem, the proposed method significantly improves the results, with an enhancement of 11.47 and 10.11% for the BOD and Q models, respectively.

Key words: activated sludge unit, stochastic differential equation (SDE), Tabriz City, uncertainty, wastewater treatment plant (WWTP)

HIGHLIGHTS

- This study suggests using stochastic differential equations (SDEs) and the Euler–Maruyama method instead of ordinary differential equations (ODEs) to model BOD and Q in WWTP design, addressing uncertainty.
- A 15-day simulation was conducted to evaluate the variations of BOD and Q pollution parameters in the biological section of WWTP for the first time.

GRAPHICAL ABSTRACT



ABBREVIATIONS

WWTP	Wastewater treatment plant
BOD (or BOD ₅)	Total 5-day biochemical oxygen demand
COD	Total chemical oxygen demand
MLSS	Mixed liquid suspended solids
SDE	Stochastic differential equations
Q	Flow rate of wastewater
ODE	Ordinary differential equations
inf	Sampled at the input of the biological reactor
eff	Sampled at the effluent of the biological reactor
PH	Potential of hydrogen
SS	Suspended solids
TDS	Total dissolved solids
TSS	Total suspended solids
VSS	Volatile suspended solids
EC	Electrical conductivity

SYMBOLS

V	Reactor volume
T	Air temperature
X	Biomass concentration
X_u	The final concentration of biomass
Q_R	Return activated sludge flow rate
Q_w	Activated sludge effluent from the WWTP system
Q	The average flow rate
k_d	Endogenous decay coefficient
S_e	Effluent substrate concentration
S_0	Influent substrate concentration
\bar{S}	The average substrate concentration
y	The cell efficiency factor
θ_c	Mean cell residence time (MCRT)
r	Sampled at returning sludge
W	Wiener process
f	The drift coefficient
g	The diffusion coefficient

a, b and c	Numerical constant coefficients
S_t	An explicit answer of the Black–Scholes equation
μ	Mean of data
σ^2	Variance of data

1. INTRODUCTION

Managing water resources for human societies requires the utilization of urban wastewater as a part of the sustainable water cycle. Wastewater reuse has been acknowledged as a fundamental strategy in safeguarding water resources and providing solutions for global water challenges (Wester *et al.* 2015; Gul *et al.* 2021). To achieve this objective, it is crucial to design close-to-reality wastewater treatment plants (WWTP) acknowledging that the operation and performance of any WWTP are significantly defined by the quality of both the untreated and treated effluents (Ahmadi *et al.* 2017; Nourani *et al.* 2018). Therefore, it is necessary to establish a direct connection (by establishing a more precise scientific link) with the treatment plant's output for industrial, agricultural, public, or sanitary purposes in order to meet environmental standards and indicators (Cresson & Sonner 2018). This will help assess uncertainty regarding physical and chemical parameters and enhance productivity. However, modeling the performance of a WWTP is challenging due to the complexity of treatment processes (Hanbay *et al.* 2008; Nourani *et al.* 2023a, 2023b). In order to enhance the design quality of treated wastewater, it is crucial to monitor and stabilize the physical variables associated with the proper functioning of activated sludge in WWTP. Consequently, applying stochastic mathematical equations to model sewage output, which can be considered a form of pollution control problem, has the potential to minimize uncertainty and enhance the accuracy of the models. Advancing our understanding of wastewater output can contribute significantly to resolving pollution control challenges.

The quality of the effluent of the WWTP undergoes a transformation through the activated sludge process, and this alteration can be assessed by measuring the biological oxygen demand (BOD). Conceptual models with mathematical relationships are commonly used to model BOD in the biological part of WWTP. However, there are two major problems that make modeling difficult when it comes to the mathematical modeling of environmental systems based on the physics of the problem. Firstly, the multitude of variables increases the number of options to be examined and the number of variables to be considered. Secondly, assessing uncertainty becomes challenging due to the intricate connections among the components engaged in stochastic process modeling and the significant level of randomness involved. In order to model BOD and the design of WWTP, ordinary differential equations (ODEs), the equations involving ordinary derivatives of a function, are typically used as mathematical solutions due to their ease of use (e.g., Diehl & Farás 2013; Mauritsson 2013). However, most natural processes are stochastic in nature, and the uncertainty caused by randomness needs to be taken into account when modeling such processes (Nourani *et al.* 2023a, 2023b). The ODE model alone cannot scientifically interpret the true physics of the phenomenon so ignore the uncertainties. Therefore, using the ODE model, uncertainties in the simulated real stochastic systems are disregarded (Compagnoni *et al.* 2023).

Numerical simulations, such as those performed using ODEs, often require accurate measurements or the adjustment of mathematical models based on prototype measurements from material models (Wylie & Streeter 1978). There are three classifications for hydrological/mathematical models: (1) black-box models, (2) conceptual models (sometimes called gray-box models), and (3) physically based models. In the physical approach (sometimes called white-box models or theoretical models), to estimate the output of a real system, it is necessary to pay attention to the inputs and the structure of the system and also its governing physical laws. In such a way, the initial and boundary conditions serve as essential constraints for solving differential equations that govern the modeling. Mathematical models based on a physical approach incorporate uncertainty stemming from simplification and assumptions made during the modeling process. Additionally, they may be susceptible to uncertainty arising from the collected physical data. The simulation of a real system can be achieved by utilizing the numerical model of the governing equations. For instance, in the analysis of activated sludge, the law of conservation of mass is applied to the input and output components of the biological part of a WWTP where the governing equations in WWTP design were of the ODE type (Metcalf 2003; Sin *et al.* 2009; Lin 2014). However, wastewater treatment using an activated sludge process involves a combination of mechanical, chemical, and biological actions, and can be monitored using nonlinear and complete equations with discontinuous solutions (Mauritsson 2013).

To assess uncertainties, investigations in the field of designing and simulating real random phenomena have continued in advanced mathematics on stochastic spaces by handling stochastic differential equations (SDEs) (Dexter *et al.* 2016). SDEs are a better alternative to ODEs for simulating random processes by taking into account uncertainties as well (Kabouris &

Georgakakos 1991). An SDE is a differential equation that is presented along with a random variable, such as white noise. In this way, SDEs are used to describe Brownian Motion and Weiner Process. In SDE, one or more of the terms is a stochastic process, resulting in a solution that is also a stochastic process. A stochastic model through SDEs includes random variables that change over time. These variables can be described by a probability distribution. Stochastic modeling helps estimate variations in a system's state and output (Ouldali *et al.* 1989). SDEs are widely used in complex modeling, such as in the physical modeling of WWTP (e.g., Beck 1987; Browning *et al.* 2020). For instance, Harris (1977) investigated SDEs using Euler equations for WWTP with the activated sludge system, considering daily data over 2 sequential days. His research was innovative, and a stabilizing pool at the influent of the treatment plant was considered to apply a constant flow rate and eliminate stochastic values. To simplify the simulation, most of the parameters and numerical values were considered constant. However, the results showed a significant difference between the calculated and observed BOD. Given the alterations of $100 \text{ m}^3/\text{day}$ for the inflow rate and 240 mg/L for BOD, in the context of a 2-day scale investigation and the presence of the scale effect in these computations, the error in the calculations was also noteworthy.

The current study utilized SDEs to analyze Brownian motion in simulating WWTP in Tabriz City. To solve the stochastic processes of BOD and Q by SDE solution, the Itô integral and Euler–Maruyama method were used. The Itô integral analyses a stochastic process with respect to Brownian motion within a bounded time interval $[0, t]$ (Erfanian *et al.* 2016). Optimal coefficients of SDE were extracted by comparing computational and observational data over 2 weeks (the longest modeling run for BOD and Q variables of WWTP lasted 2 consecutive days, based on prior research by Harris (1977)). The application of this method in the assessment of uncertainty in the activated sludge component is considered an innovation to the present research. A Wiener model using SDE was presented to check real data and measure pollutant biological parameters for short-term changes in BOD and Q , and to assess uncertainties in the biological reactor section.

2. MATERIALS AND METHODS

2.1. Study site and data

The study examined the assessment of uncertainty in simulating BOD and Q agents in WWTP design, with a focus on SDE solutions. Simulations ran from 2019 to 2020 for 15 days (for the first time), for the BOD pollution index and the amount of Q changes in WWTP's biological part in Tabriz City with an activated sludge system.

Tabriz is a developed city in Iran, known for its urbanization and industry. The Tabriz WWTP, a diffuser-aerated activated sludge WWTP, is located 4 km to the west of Tabriz, on the southern side of the Aji-Chay River. This WWTP has been designed and built specifically to treat the urban wastewater of 612,000 people. Its goal is to reduce pollution and provide reusable treated wastewater for agricultural and industrial activities. The WWTP system in Tabriz, shown in Figure 1, consists of two primary stages (physical treatment), followed by secondary treatment (biological treatment), and finally, disinfection. The biological treatment process includes an activated sludge treatment unit, secondary sedimentation, anaerobic digestion tanks, and sludge lagoons.

The WWTP has a designed annual average discharge of $1.5 \text{ m}^3 \text{ s}^{-1}$, but the actual average discharge is $1.15 \text{ m}^3 \text{ s}^{-1}$. Records indicate that the maximum discharge on non-rainy days reaches $2.5 \text{ m}^3 \text{ s}^{-1}$, while on rainy days, it can go up to $3.8 \text{ m}^3 \text{ s}^{-1}$. The influent concentrations of BOD and total chemical oxygen demand (COD) variables are 250 and 320 mg l^{-1} , respectively. The Tabriz City WWTP consists of physical, biological, and chemical treatment stages. The incoming sewage carries suspended materials weighing $32,400$ and $45,360 \text{ kg day}^{-1}$, respectively. Table 1 presents various statistical characteristics of WWTP's pollution indexes. Also, as target parameters in SDE modeling, Figure 2 illustrates the daily fluctuations of Q_{inf} and Q_{eff} to WWTP over a 1-year period (2019–2020).

To prevent critical conditions for the entrance of wastewater into the treatment plant, a controller instrument has been designed and installed in the intake channel. This controller induces changes within a specific range to regulate the wastewater load to the plant. Figure 3 depicts the daily time series of BOD and COD levels in the influent to the treatment plant. Comparing the graph with precipitation data shows a noticeable decrease in pollution indicators. This decrease can be attributed to the dilution of the incoming flow to the WWTP.

2.2. Stochastic differential equation

The probability space associated with a stochastic phenomenon is represented by a continuous path Wiener process within a specific time interval $W(t_i)$, characterized by independent variables. As a result of the presence of Wiener processes in the corresponding stochastic integral, the solution to these equations is not derivative-based concerning time (t), and typically,

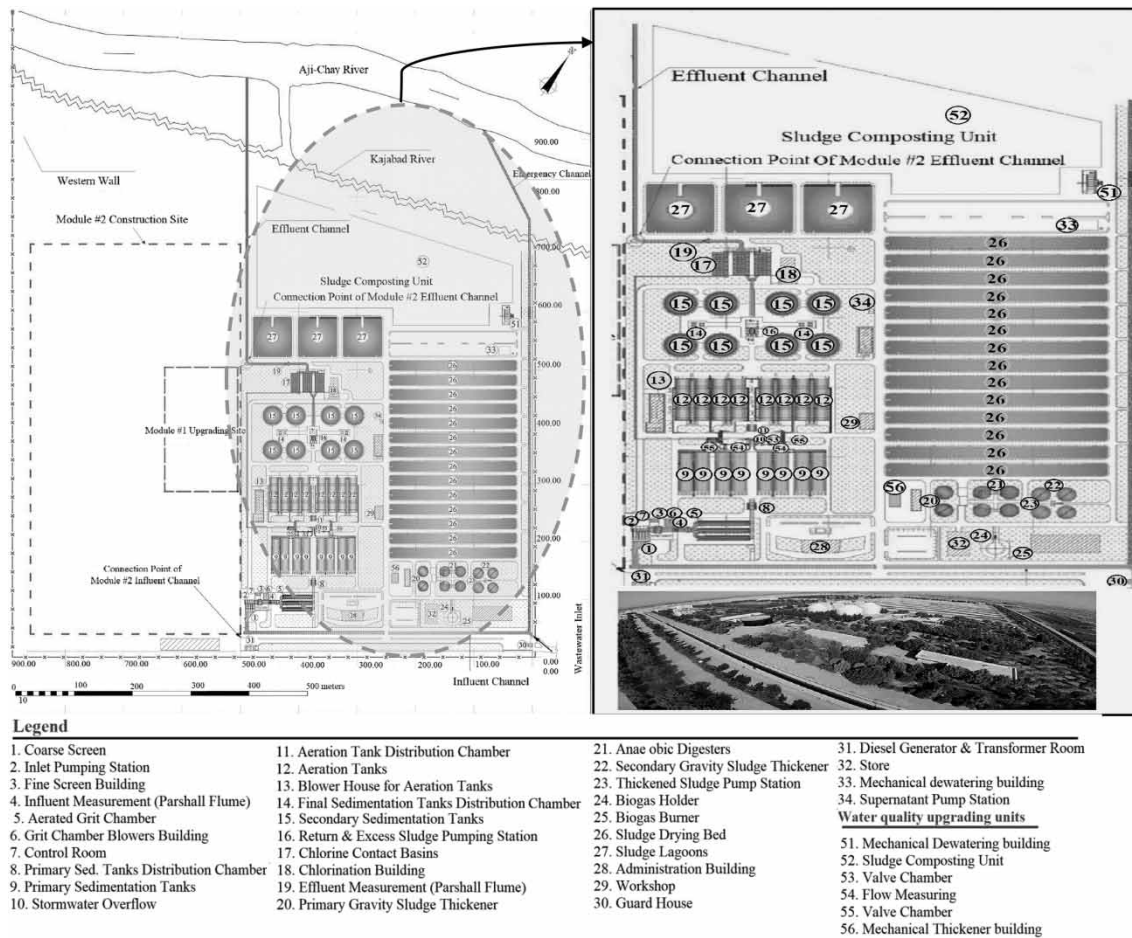


Figure 1 | The schematic diagram of Tabriz WWTP.

such equations are expressed in terms of stochastic integrals $\hat{I}t_0$. The $\hat{I}t_0$ stochastic integral is a concept within stochastic calculus that expands upon the Riemann–Stieltjes integral (which extends the Riemann integral to integrate functions with respect to another function) to encompass stochastic processes. This integral is of great significance in the study of SDEs and finds utility in various disciplines, such as finance, economics, and physics. By accounting for the stochastic nature and volatility of the underlying process, the $\hat{I}t_0$ stochastic integral is defined with respect to a continuous stochastic process. It is employed to solve SDEs and analyze the behavior of stochastic processes (Fima 1999).

The general form of an SDE is presented in the following equation as shown by Erfanian *et al.* (2016):

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t \tag{1}$$

where in a d -dimensional field, let X_t be a set of variables, f be the drift coefficient, g be the diffusion coefficient of the $(d \times m)$ -dimensional matrix, and W_t be the m -dimensional Wiener process. To define the $\hat{I}t_0$ integral with a certain time step and the condition $(t_0 \leq t_1 \leq \dots \leq t_n = T)$ a part of the range $[t_0, T]$, and $\lambda_n = \max(t_i, t_{i+1})$ are considered. In this case, the $\hat{I}t_0$ integral can be defined as shown in the following equation:

$$Y_t(t) = \int_{t_0}^T g(s, X_s)dW_s = \lim_{\lambda_n \rightarrow 0} \sum_{i=1}^n g(t_{i-1}^*, X_{i-1})(W_{t_i} - W_{t_{i-1}}); t_{i-1}^* = t_{i-1} \tag{2}$$

If g is a measurable function and the integral of $\int_{t_0}^T E(|g(s, X_s)|^2)ds \leq \infty$ is finite, then the limit of Equation (2) exists as a result of the $\hat{I}t_0$ integral. The Euler–Maruyama method is used for the numerical solution of SDE in this study. Geometric

Table 1 | Pollution index statistics for Tabriz WWTP from 2015 to 2020

Variable	Measurement stage	Unit	Minimum	Maximum	Mean	Std
PH _{inf}	Influent to WWTP	–	7.29	8.61	8.05	0.19
BOD _{inf}	Influent to WWTP	mg L ⁻¹	220.00	315.00	290.12	13.05
COD _{inf}	Influent to WWTP	mg L ⁻¹	362.17	516.67	479.16	22.01
SS _{inf}	Influent to WWTP	mg L ⁻¹	0.00	1.50	0.61	0.151
TSS _{inf}	Influent to WWTP	mg L ⁻¹	234.00	336.33	296.40	14.64
TSS _r	Return activated sludge	mg L ⁻¹	3,048.67	3,892.00	3,271.21	89.83
VSS _{inf}	Influent to WWTP	mg L ⁻¹	140.50	235.33	191.13	15.12
TDS _{inf}	Influent to WWTP	ppm	1,067.17	1,366.83	1,202.01	51.19
T _{inf}	Influent to WWTP	°C	14.30	26.78	21.04	2.98
T _{air}	Air temperature	°C	-9.08	31.42	12.42	10.84
BOD _{eff}	Effluent from WWTP	mg L ⁻¹	17.33	26.33	22.72	1.28
BOD _{eff}	Effluent from primary settling	mg L ⁻¹	140.00	197.50	181.37	8.48
BOD _{eff}	Secondary sedimentation effluent	mg L ⁻¹	17.50	27.33	23.66	1.35
COD _{eff}	Effluent from primary settling	mg L ⁻¹	235.33	320.50	297.25	13.33
COD _{eff}	Effluent from WWTP	mg L ⁻¹	30.00	40.83	37.98	1.87
TDS _{eff}	Effluent from WWTP	ppm	779.83	1,188.50	936.34	64.76
VSS _{eff}	Effluent from WWTP	mg L ⁻¹	6.33	13.13	9.68	1.16
T _{eff}	Effluent from WWTP	°C	13.97	27.73	21.05	3.34
TSS _{eff}	Effluent from primary settling	mg L ⁻¹	96.67	132.33	119.46	5.55
TSS _{eff}	Secondary sedimentation effluent	mg L ⁻¹	14.80	26.45	22.30	2.09
TSS _{eff}	Effluent from WWTP	mg L ⁻¹	14.47	25.98	21.92	2.043
COD _{eff}	Effluent from primary settling	mg L ⁻¹	235.33	320.50	297.25	13.33
PH _{eff}	Effluent from primary settling	–	7.09	8.16	7.70	0.128
PH _r	Return activated sludge	–	6.72	7.95	7.43	0.156
MLSS	Active sludge	mg L ⁻¹	1,138.00	2,870.00	2,273.48	147.50
EC _{inf}	Active sludge	μSiemens cm ⁻¹	824.00	2,320.00	1,550.23	280.76

Brownian motion is often used in technical analysis, which includes positive and negative variables (Erfanian *et al.* 2016). When modeling the parameters of the treatment plant, the activated sludge section, and the aeration lagoons, it is important to consider that the system is progressive and the measured variables will not be negative. Therefore, Brownian motion with log-normal distribution, which yields only non-negative values, was used. The changes in dS_t in this system follow the recursive process of the following equation.

$$dS_t = a(b - S_t)dt + c\sqrt{S_t}dW_t \quad (3)$$

where a , b , and c are numerical constant coefficients that follow the Hull and White risk natural stochastic volatility model (Higham 2001). dW_t represents changes in Brownian motion, and S_t is an explicit answer of the Black-Scholes equation, which is a more general form of Equation (1). The value of S_t can be determined using the following equation:

$$S_t = S_0 e^{(\mu - (1/2)\sigma^2)t + \sigma W_t} \quad (4)$$

Symbols μ and σ^2 represent the mean and variance of data, respectively. σ denotes standard deviation and W_t represents Wiener probability process at each time step of δt .

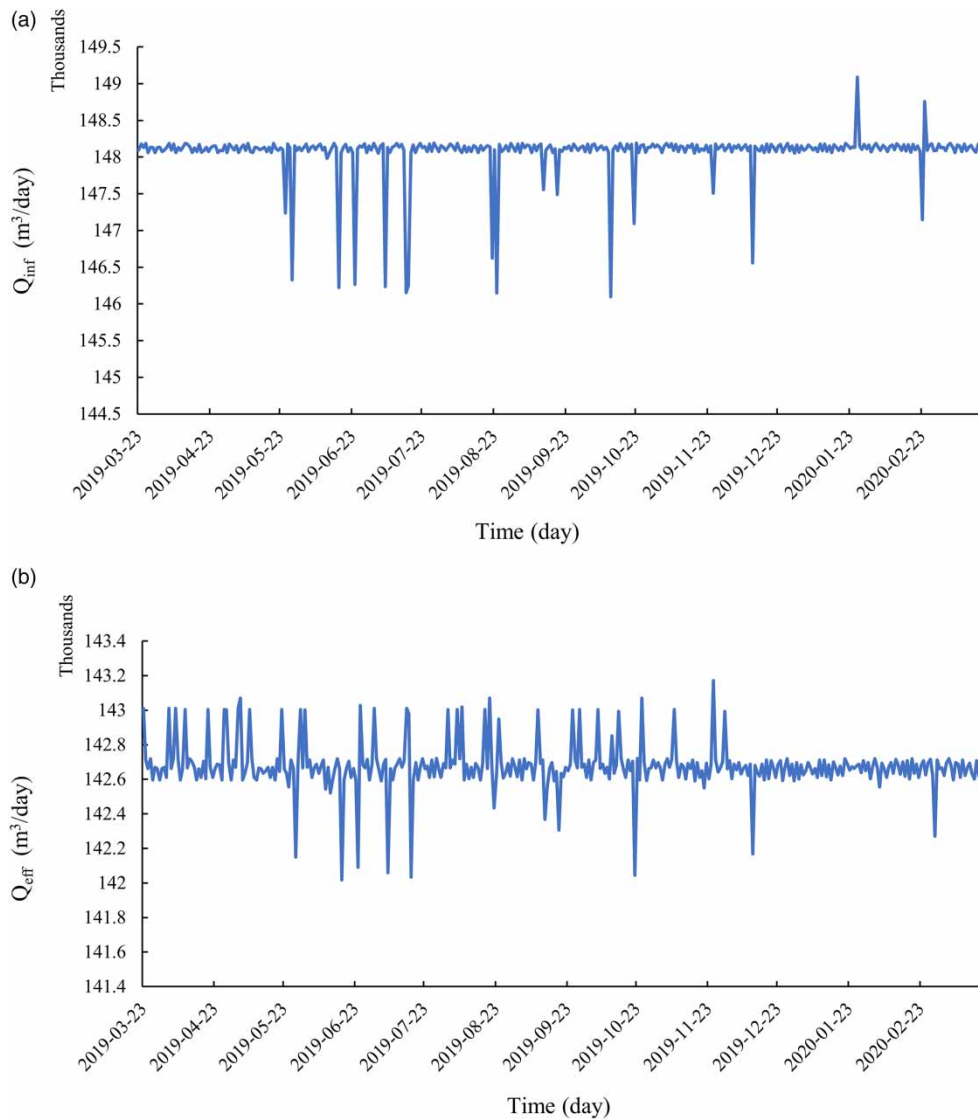


Figure 2 | Daily variations in wastewater discharge at the treatment plant for (a) Q_{inf} and (b) Q_{eff} .

2.2.1. SDE for activated sludge process: the concept and algorithm

In general, the behavior of the activated sludge process can be analyzed using ODE as demonstrated by [Kabouris & Georgakakos \(1991\)](#). However, to address the uncertainty associated with pollution parameters in the biological treatment stage, SDEs are employed in order to overcome the limitations of ODE modeling ([Pérez-Aros et al. 2022](#)). Therefore, in this study, SDE was utilized to mitigate the impact of uncertainties related to the activated sludge stage and the biological parameters of the effluent or influent system. The activated sludge process is formed in the secondary sedimentation section ([Metcalf 2003](#)). The variables involved in the formation of the activated sludge process are depicted in [Figure 4](#). Within this cycle, Q represents the secondary influent flow rate ($\text{m}^3 \text{day}^{-1}$), Q_R denotes the return activated sludge flow rate ($\text{m}^3 \text{day}^{-1}$) with concentration X_u , and Q_w signifies the activated sludge effluent from the WWTP system ($\text{m}^3 \text{day}^{-1}$). V represents the volume of the tank or reactor (m^3), X represents the biomass concentration (mg L^{-1}), X_u represents the final concentration of biomass (mg L^{-1}), S_0 represents the influent substrate concentration at time t_0 , which is equivalent to BOD_{inf} , and S_e represents the effluent substrate concentration at time t , which is equivalent to $\text{BOD}_{5_{eff}}$ (mg L^{-1}) as provided by the environmental organization. The primary objective of this unit is to separate a large volume of mixed liquor suspended

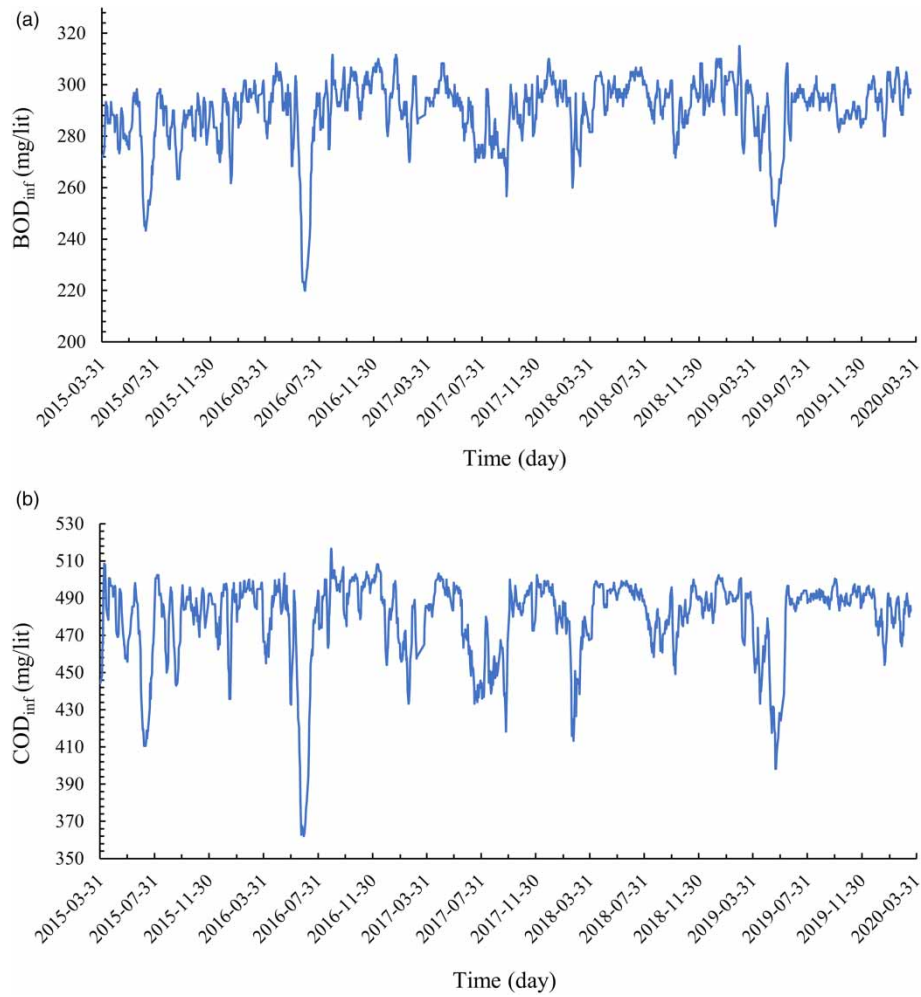


Figure 3 | Daily variations in wastewater treatment from 2015 to 2020 for (a) BOD_{inf} and (b) COD_{inf} .

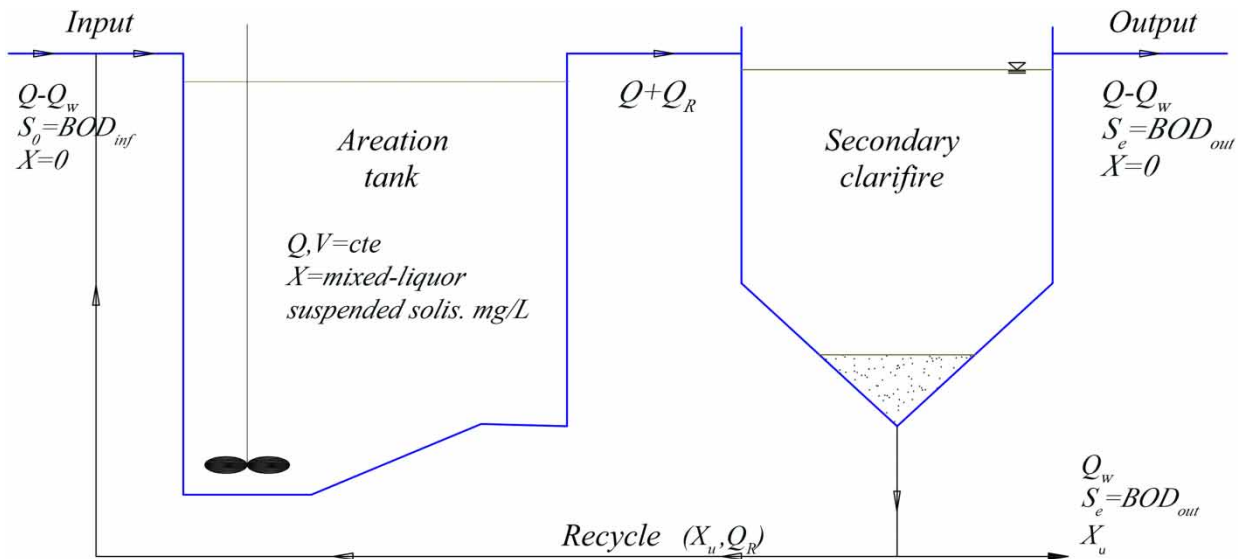


Figure 4 | A diagram of the biological section of the WWTP utilizing the activated sludge method (Metcalf 2003).

solids (MLSS) from the aeration tank and produce a clear and stable stream with low total suspended solids (TSS) and BOD concentration. This cycle is illustrated schematically in Figure 4.

The initial endeavor to address SDEs and provide analytical descriptions of the probabilistic phases was initiated by Serano, who attempted to solve the SDEs by employing the Euler–Maruyama discretization method and the $\dot{I}\acute{o}$ process, considering uncertain values such as Wiener and white noise. Subsequently, Särkkä (2006) introduced a novel solution for modeling through SDEs, while Richardson (2009) and Evans (2012) proposed a generalized analytical method of $\dot{I}\acute{o}$ for solving first-order linear stochastic equations. These advancements enabled the numerical modeling of the Wiener process in various time intervals and dispersion, specifically for one-dimensional and Brownian motion, as applied to the activated sludge process. Higham (2001) introduced a probabilistic numerical algorithm that obtained solutions for the Euler–Maruyama and Millstein approximations by utilizing Euler’s stochastic chain method and considering both weak and strong convergences. This algorithm was then employed to solve simple nonlinear differential equations. However, it should be noted that the presented models may not adequately address most engineering problems, as the majority of the differential equations governing such problems are nonlinear and partial equations, as highlighted by Xiu & Hesthaven (2005). To summarize the content, the calculation steps are depicted schematically in Figure 5 and are followed here to model random effective parameters in the quality of wastewater treatment plant (WWTP) effluent.

2.2.2. SDE for activated sludge process: math relationships development

In a solution based on SDEs, there are two main components which followed in this study. The first involves converting activated sludge equations into the SDE format using treatment plant design principles. The second component involves solving these equations numerically using the $\dot{I}\acute{o}$ integral and the Euler–Maruyama method. Activated sludge calculations rely on

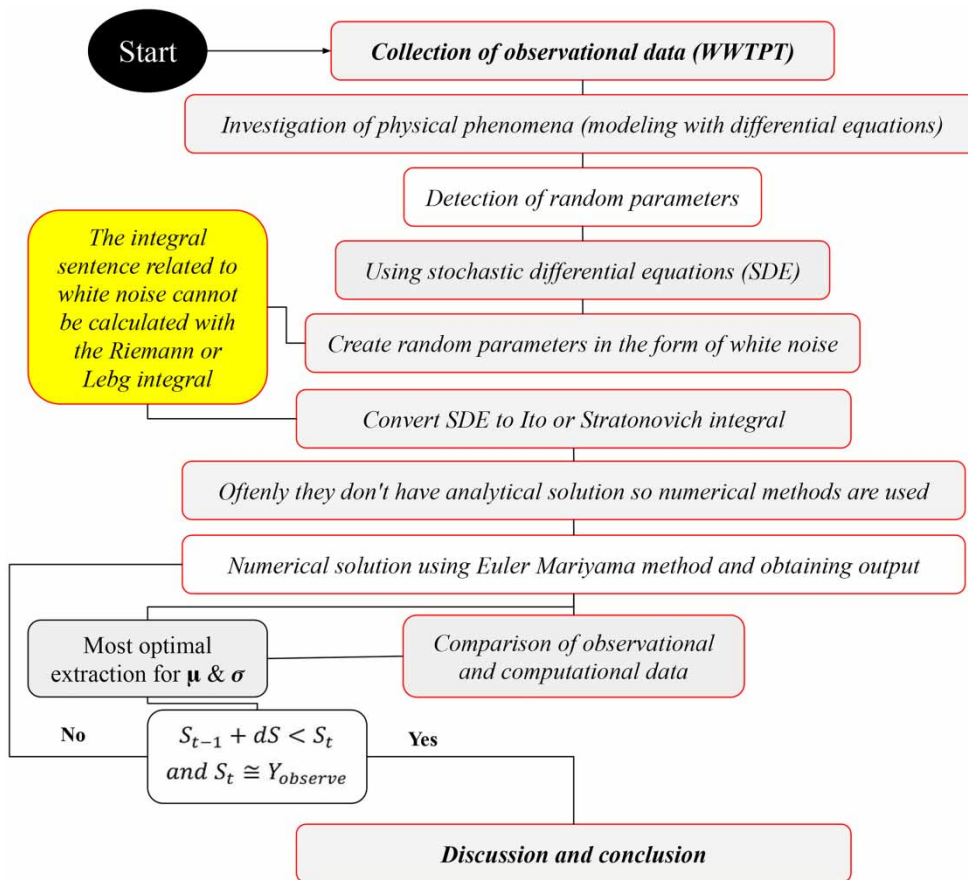


Figure 5 | The Euler–Maruyama solution method’s diagram to compute the $\dot{I}\acute{o}$ integral of SDEs in modeling BOD and Q, as there is no analytical solution available for it.

four fundamental Equations (5) through (8), as outlined by Metcalf (2003):

$$V \frac{dX}{dt} = Qy(S_0 - S_e) + Q_w X_u + k_d V X \tag{5}$$

where y represents the cell efficiency factor, which depends on substrate characteristics, activated sludge type, and temperature. k_d represents the endogenous decay coefficient (T^{-1}). In the second step, the treatment unit’s capacity is determined using the following equation:

$$\text{if } Q = \text{cte} \Rightarrow V = t \times Q \tag{6}$$

where t is the hydraulic retention time (HRT), and V is the tank volume (m^3). θ_c represents the mean cell residence time (MCRT), a crucial design factor for activities of microorganisms in the biological reactor, obtainable from the following equation.

$$\theta_c = \frac{X V}{Q_w X_u} \tag{7}$$

the value of θ_c must exceed the duration of the highest bacterial population growth in the system in order to avoid the release of biomass or biological clots with the effluent. Assuming a constant flow rate, the equation of continuity in the biological reactor section yields the following equation:

$$(Q + Q_R)X = (Q_w + Q_R)X_u \tag{8}$$

the term $(Q + Q_R)X$ represents the influential biological flocculation, while the term $(Q_w + Q_R) X_u$ represents the effluent biological flocculation ($mg L^{-1}$). The inputs and outputs for these equations are listed in Table 2.

Hence, the initial objective of the study was to derive the Wiener form of Equation (5) for Q_{eff} and BOD_{eff} in the first phase, and subsequently transform them into stochastic form. It means that the equations pertaining to activated sludge were derived, taking into account the uncertainties associated with the stochastic mode and the Euler–Maruyama form. In the second phase, the relevant analytical codes were randomly executed over a period of 15 days, spanning from 2019 to 2020, and the outcomes were documented. The data used for the model were selected from a 1-year period. Since it was possible to analyze the data using SDE, it was divided into periods of up to 15 days. It was not possible to increase the time period for more than 15 days, in order to apply the divergence in the calculations and account for the scale effect. The calculations are divided into 2-week periods for 1 year. After various modeling and optimal value extraction, the best model was chosen. The calculations incorporated variations in instantaneous time steps and daily data collection, owing to the inherent uncertainty in the physics of the phenomenon. The stochastic variations can be analyzed from three different perspectives:

- Q_{inf} and BOD_{inf} uncertainties, and their influence on X , impacted by climate change, precipitation, and flow volume, can decrease wastewater BOD. Consumption patterns during seasons can also affect these parameters.
- The uncertainty surrounding the parameters that govern the behavior of microorganisms, specifically k_d and y , as well as the impact of past data on these values, is a matter of concern. The daily fluctuations in the birth and death rates of the population have a direct bearing on the aforementioned parameters, and their investigation is imperative in light of the uncertainties involved.

Table 2 | Inputs and outputs which play a crucial role in the design equations of activated sludge systems

Variable	Inputs							Outputs		
	V	k_d	y	X_u	X	S_0	Q_w	Q_R	Q	S_e
Units	m^3	day^{-1}	-	$mg L^{-1}$	$mg L^{-1}$	$mg L^{-1}$	$m^3 day^{-1}$	$m^3 day^{-1}$	$m^3 day^{-1}$	$mg L^{-1}$

- Flow rates fluctuate in certain hours of the day to their maximum and minimum values over a period of 1 week to 1 month, but are treated as constant in calculations and significantly affect activated sludge parameters.

To address the uncertainty surrounding the output of WWTP and the system itself, two key assumptions are made. The first assumption posits that the volume of sludge is directly proportional to the quantity of sludge increase.

$$\left(\frac{dx}{dt} \propto X\right) \Rightarrow \frac{dx}{dt} = k_d X \Rightarrow dx = X_{t+1} - X_t \tag{9}$$

The variable dx represents sludge generated from food consumption. The second hypothesis relates to changes in microorganisms' food source, causing fluctuations in BOD value. This implies that:

$$\left(\frac{dS}{dt} \propto S\right) \Rightarrow \frac{dS}{dt} = -qx \tag{10}$$

The change in sludge concentration is determined by dS . The following equation connects the ratio of dx/dt to dS/dt .

$$dx = y \times ds \tag{11}$$

The following equations are derived based on the principle of conservation of mass and underlying assumptions.

$$V \frac{ds(t)}{dt} = QS_0(t) - QS_e(t) - Vx(t)q \tag{12}$$

$$V \frac{dx(t)}{dt} = QX_0(t) - Q_w X_u - k_d Vx \tag{13}$$

In this context, $QS_0(t)$ and $QX_0(t)$ represent mass entering the system, while $QS_e(t)$ represents mass exiting the system. X_0 is zero at the influent stage of the WWTP. The effluent mass from the system is denoted as $Q_w X_u$, which has a negative connotation. The parameter $k_d Vx$ is associated with the mass of the reactor. In Equations (12) and (13), the values of $S_e(t)$ and $x(t)$ are unknown. $x(t)$ represents the quantity of sludge, while $S_e(t)$ represents the quantity of BOD_{eff} . S_e is a limited value determined according to environmental organization standards.

The design adjusts parameters based on S_e , allowing for modification until the desired outcome is achieved. X_u is influenced by sludge concentration. The final step is stochastic equations based on Q and S_e (or BOD). A sequence of variables Q and S_e produces data to derive Equations (14) and (15) representing stochastic equations for flow rate and sludge concentration. The Euler–Maruyama technique with geometric Brownian motion and a log-normal distribution is used to derive these equations.

$$dQ = a_1(\bar{Q} - Q_t)dt + c_1 \sigma_Q \sqrt{Q_t} dW_1 \tag{14}$$

$$dS_0 = a_2(\bar{S} - S_e)dt + c_2 \sigma_{S_0} \sqrt{S_e} dW_2 \tag{15}$$

\bar{Q} is the average flow rate, \bar{S} is the average BOD based on observational data, Q_t and S_e are specific values at a given moment, dW_i are stochastic independent variables in the form of $\sqrt{\delta t}N(0, 1)$ following a Brownian motion with a log-normal distribution, $a_1, a_2, c_1,$ and c_2 are numerical constants determined through trial and error, and σ is the standard deviation of the observational data.

Due to the comparison of SDE data with the design criteria and relying on the derived equations, it is impossible to calculate the error. Equations (16) and (17) can be taken to determine the percentage improvement in the estimation of BOD and

Q variables using the proposed model (Scott *et al.* 2003).

$$r1 = \frac{|x - y|}{\frac{(x+y)}{2}} \times 100 \tag{16}$$

$$r2 = \frac{|z - y|}{\frac{(z+y)}{2}} \times 100 \tag{17}$$

where *x* represents the ODE design of the WWTP (Metcalf 2003), *y* stands for the average SDE simulation result, and *z* denotes the average observational data. The disparity between *r1* and *r2*, i.e., Δ (*r1*, *r2*), determines the improvement percentage.

3. RESULTS

3.1. SDE modeling results for flow rate (Q)

The assessment of uncertainty took into account the fluctuations in influent flow caused by varying consumption patterns. To prevent turbulent flows during floods or rains, a flow stabilization criterion was implemented, ensuring a specific range of fluctuations. This characteristic results in the imposition of a specific level of influent sewage load on the treatment facility. However, uncertainties arising from surface runoff or other factors could not be measured at specific time intervals for excessive design loads. The coefficients in Equation (14) were determined to match the observational data over a 2-week computational time series. The Wiener process (*W_i*) is employed to model uncertainties, with the vector function *a_i* and matrix function *c_i* being influenced by several factors such as the flow rate and volume of sludge, as well as the overall kinetic and stoichiometric biological constants found in wastewater. The study focuses on a real model of a WWTP and Table 3 presents the modeling results, comparing the optimal coefficients of stochastic equations with the observational data over a period of 15 days.

A total of 15 days are chosen at random from the 365-day period, and calculations were carried out in 15-day intervals throughout the year (2019–2020). The optimal coefficients were estimated through trial and error, resulting in the lowest comparative error. Comparative charts in Figure 6 illustrate an example of the results. Using the Euler–Maruyama method,

Table 3 | Optimal coefficients of stochastic equations for WWTP effluent flow

Day	Observation data (m ³ day ⁻¹)	Random number	$\bar{Q} - Q_t$	$\sqrt{Q_t}$	<i>dw_i</i>	Prediction data (m ³ day ⁻¹)	Constant coefficients
1	142,719	0.188	57.87	377.78	0.26	142,719.00	<i>a</i> ₁ = 0.0035
2	142,691	0.910	-85.57	377.78	-0.42	1,427,760.87	<i>c</i> ₁ = 0.0028
3	142,653	0.484	89.70	377.78	0.41	142,689.29	<i>dt</i> = 1 × 10 ⁻³ [s]
4	142,663	0.532	10.36	377.80	0.04	142,779.00	
5	142,677	0.711	-66.98	377.84	-0.32	142,789.35	
6	142,688	0.694	-15.87	377.83	-0.08	142,722.38	
7	142,639	0.119	-143.50	377.87	-0.68	142,706.51	
8	142,649	0.037	89.52	377.78	0.41	142,563.00	
9	142,604	0.681	-26.27	377.80	-0.13	142,652.53	
10	142,673	0.322	63.94	377.76	0.29	142,626.25	
11	142,719	0.367	27.88	377.65	0.12	142,690.19	
12	142,667	0.931	-22.33	377.62	-0.11	142,718.08	
13	142,555	0.390	-2.83	377.64	-0.02	142,695.74	
14	142,694	0.908	-2.83	377.68	0.28	142,692.91	
15	142,611	0.019	-10.33	377.75	-0.06	142,754.81	
16	142,673	0.261	65.27	377.79	0.29	142,744.48	

α_Q = 42.98

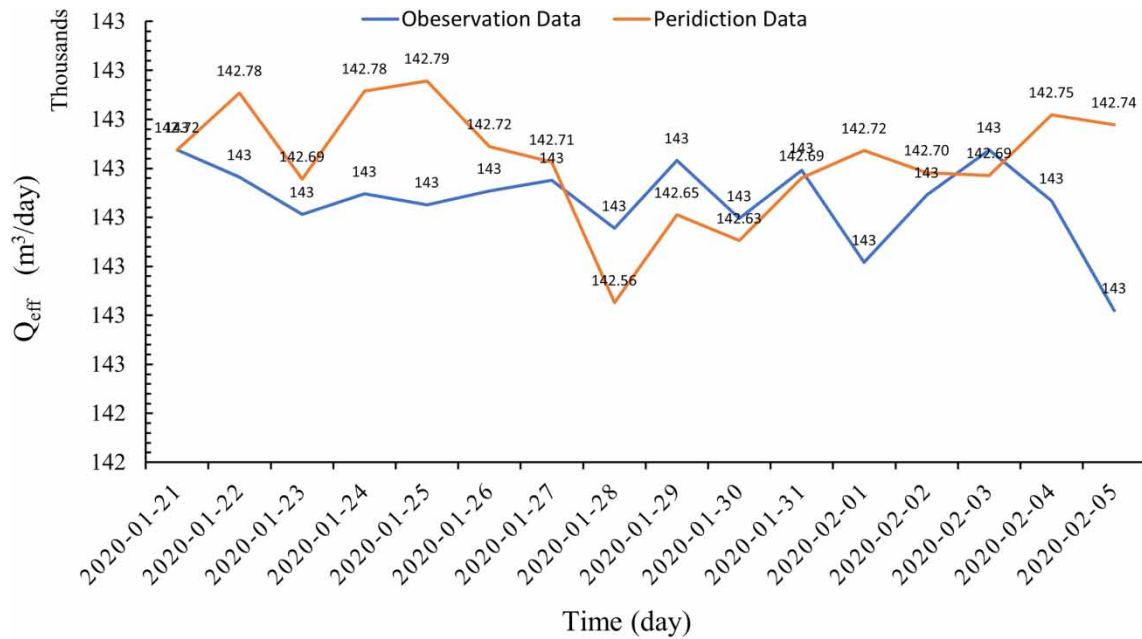


Figure 6 | Comparison of observed and estimated Q_{eff} for WWTP in Tabriz City using the SDE model.

relevant analyses were performed to select predicted values with the least difference compared to the observed data. However, due to the development of stochastic values in the likelihood space and the nature of Brownian motion, the possibility of error increases with each progression of time steps in calculations. The coefficients obtained for SDE were implemented in the best optimal mode for predicted values in the calculations. The obtained results were generated as random numbers in each calculation time step. The input figures for Q or BOD have a wide range of changes, and the scale effect has a significant impact on this analysis as well. As a result, abnormal jumps and extreme fluctuations were observed in the results within the 15-day time period, with considerable intervals between them. It means that the convergence of answers obtained for long time intervals led to the calculations were time-consuming. In order to determine the optimal coefficients of the SDEs, a significant amount of trial and error has been employed. These limitations are important to note when considering the method presented in this research.

As time progresses during calculations, the possibility of error increases due to the development of stochastic values in the probability space and the nature of Brownian motion. However, the predicted values in the calculations were obtained by applying the coefficients obtained for SDE in the most optimal mode. The fluctuation of analysis figures over time for each data is apparent. The extended duration required for computation due to the convergence of answers over prolonged periods has emerged as a constraint of the method examined in this study. Subsequently, a set of answer categories was determined for each day within the 15-day period. This particular category of answers was chosen based on the highest degree of overlap, by creating a loop within the MATLAB software, between the maximum, minimum, and average values, and involved seven calculation steps which were determined through trial and error. It is important to note that the data information for the first day is already known and remains fixed throughout the calculations. Finally, the changes in Q_{eff} for each day, encompassing both the maximum and minimum values, are illustrated in Figure 7 based on the results.

3.2. SDE modeling results for BOD

The BOD variable underwent modeling in three distinct phases spanning the time period of 2019–2020. These phases included the input of the treatment plant, the output of primary sedimentation, and the output of secondary sedimentation. The fluctuating range of this variable during each phase is depicted in Figure 8, which serves to determine the biological performance of the WWTP located in Tabriz City. The numerical coefficients were determined using Equation (15) and a process of trial and error in order to establish a meaningful relationship between the input and calculated data from the Brownian motion at each time step, from t_i to t_{i+1} . The findings were subsequently presented in Table 4.

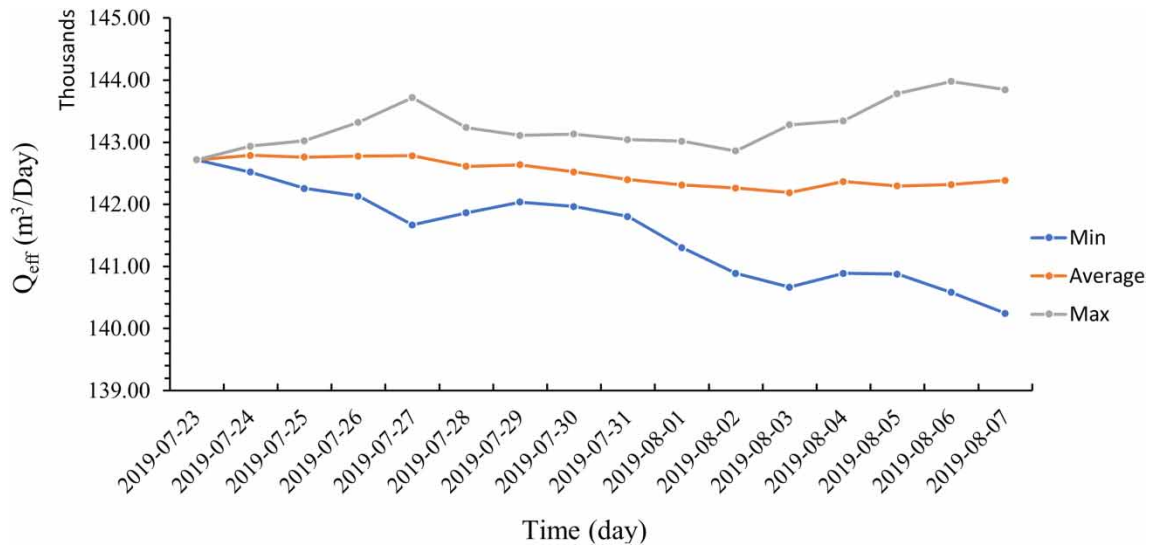


Figure 7 | Estimated the maximum, minimum, and average values of Q_{eff} utilizing the SDE model across a duration of 15 days and 7 steps.

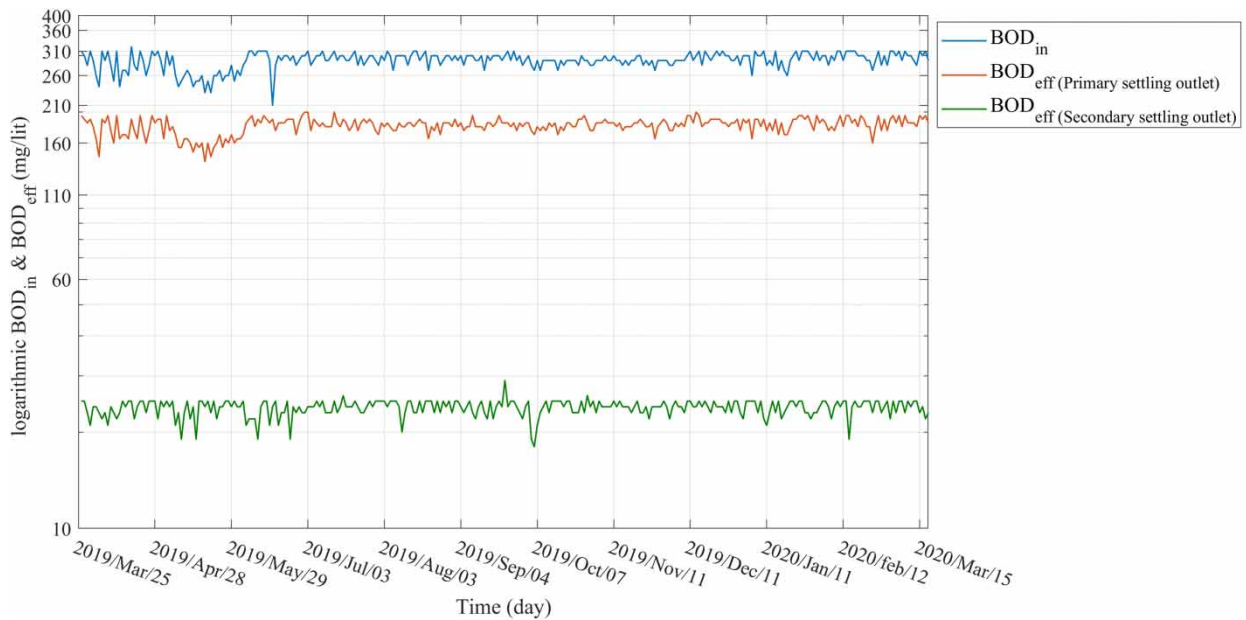


Figure 8 | Fluctuations of BOD variable in three different phases of the WWTP in Tabriz City. To improve graphical visualization, the logarithmic behavior of the variables was depicted.

Equation (7) was utilized to analyze the randomly selected BOD_{eff} data for a consecutive period of 15 days, in conjunction with Q_{eff} , spanning from 2019 to 2020. The outcomes of the best model are depicted in Figure 9, which illustrates a comparison between the computational and observational data. The group of responses for BOD_{eff} was determined similar to Q_{eff} calculations. The calculation process involved a time period of 15 days and 10 calculation steps, which were determined through trial and error. The first day's data were a fixed number from observation data. Figure 10 depicts a variation of the BOD_{eff} during SDE modeling, including the highest, mean, and lowest values obtained for each day.

4. DISCUSSION

To design WWTP, using linear analyses overlooks important factors of fluctuations in wastewater phenomena such as critical rainfall events. Insufficient understanding of the uncertainties involved in the processes often leads to significant errors in the

Table 4 | Optimal coefficients of stochastic equations for BOD_{eff} in WWTP of Tabriz City

Day	Observation data (mg L ⁻¹)	Random number	$\bar{S} - S_t$	$\sqrt{S_t}$	dw_t	Prediction data (mg L ⁻¹)	Constant coefficients
1	24	0.511	-0.9375	4.8990	-0.338	24.000	$a_1 = 0.005$
2	23	0.307	-0.5002	4.8541	0.000	20.893	$c_1 = 0.7$
3	24	0.618	-0.0346	4.8059	0.110	21.014	$dt = 1 \times 10^{-3}$ [s]
4	25	0.584	-0.1654	4.8195	-0.343	22.058	
5	22	0.350	-0.2729	4.8307	0.267	19.161	
6	23	0.577	-0.8680	4.8919	-0.300	21.316	
7	21	0.761	0.3462	4.7662	-0.029	18.882	
8	23	0.966	0.1595	4.7857	0.677	18.776	
9	22	0.481	0.7090	7.7279	0.090	23.968	
10	23	0.655	0.7498	7.7236	0.213	24.969	
11	24	0.534	0.9054	4.7071	-0.020	27.238	
12	25	0.315	1.0441	4.6924	0.077	27.179	
13	22	0.653	0.5534	4.7444	-0.504	28.171	
14	23	0.998	1.7057	4.6213	-0.199	22.653	
15	21	0.394	2.1998	4.5676	0.356	20.980	
16	24	0.740	1.8310	4.6078	-0.383	24.089	

$\sigma_Q = 1.24$

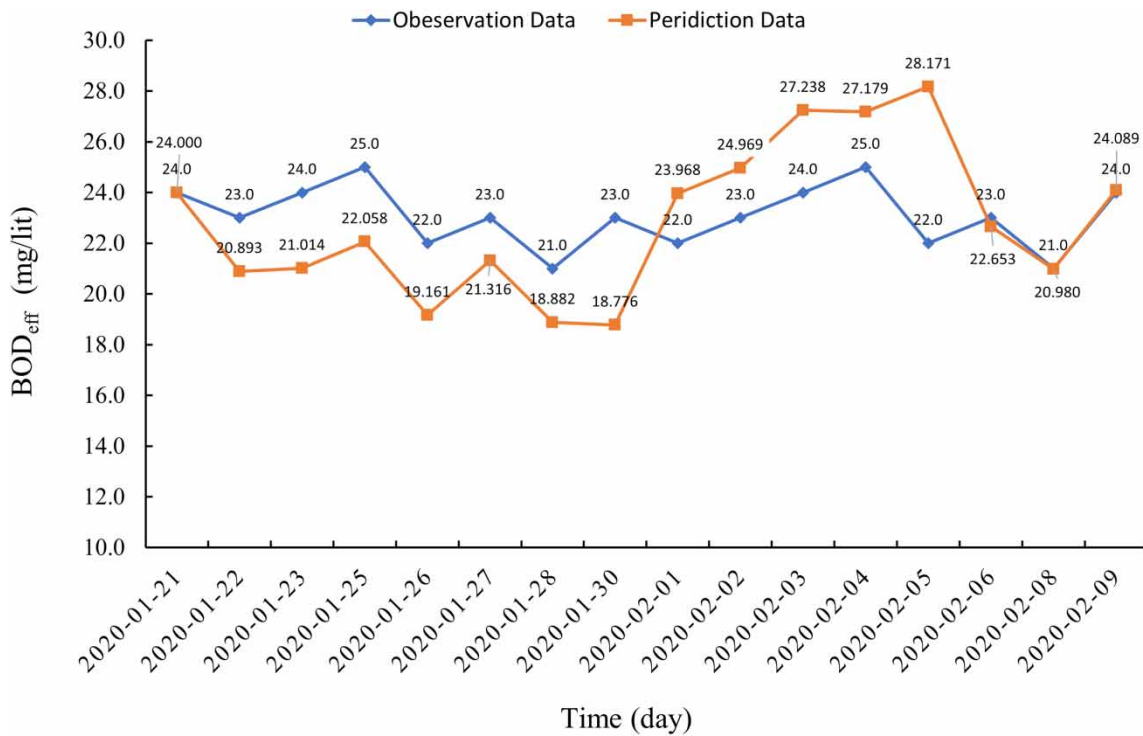


Figure 9 | Comparison of observed and estimated BOD_{eff} for WWTP in Tabriz City using the SDE model.

calculation factors and implementation volumes of WWTP projects, resulting in increased financial burdens. For instance, failure to accurately calculate the minimum coefficient Q_{inf} to design with ODE in critical situations can lead to prolonged hydraulic retention times in sedimentation ponds, aeration, and the time required for settled sludge to reach the nitrification

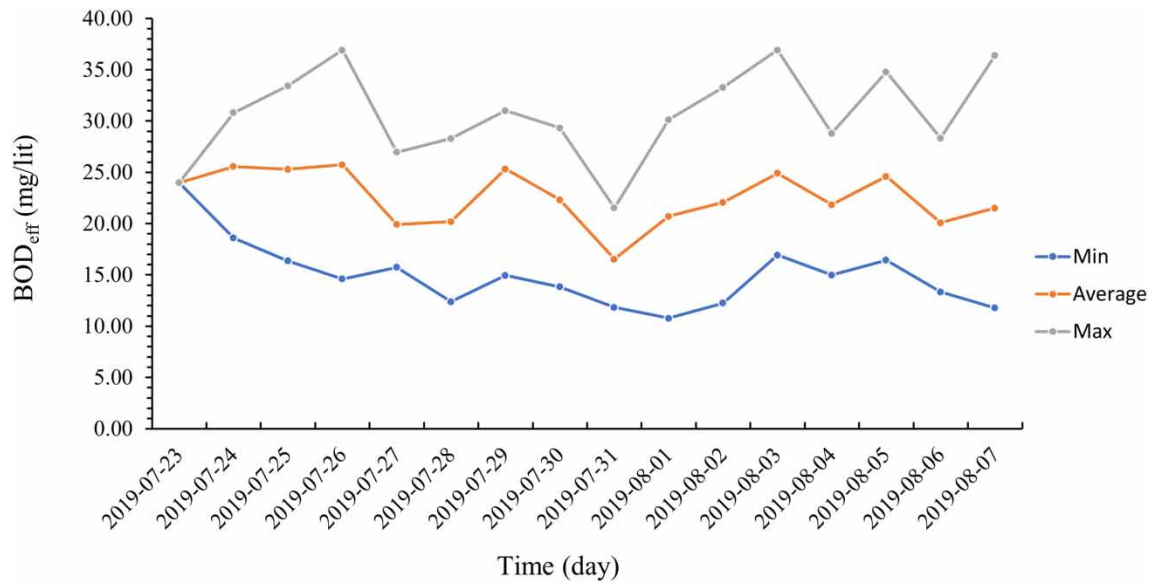


Figure 10 | Estimated the maximum, minimum, and average values of BOD_{eff} utilizing the SDE model across a duration of 15 days and 10 steps.

stage. This, in turn, causes an increase in the sedimentation level of compressed particles, resulting in a decline in the quality of treated wastewater. In this study, the foundation of the work was established upon models for stochastic processes, which have the potential to make a significant contribution to the development of models based on linear first-order ODEs, utilized for the analysis of equations in the activated sludge section of the WWTP. To achieve this objective, the stochastic form of the equations for two output parameters, namely Q_{eff} and BOD_{eff} , was extracted from the series of linear equations of the activated sludge section. Afterwards, the constant coefficients (a_i and c_i) were modified for both models. The optimal values for a_i and c_i for Q_{eff} were 0.0035 and 0.0028, respectively, while for BOD_{eff} , they were obtained as 0.005 and 0.7, respectively. To ensure maximum convergence in results, δt was set to 0.001. However, one of the limitations of this method is its inability to accurately compute for long-term time scales. The presence of computation errors in each time step leads to the accumulation of unacceptable errors in modeling. Nevertheless, this method demonstrated a significant ability to monitor the fluctuations and activities of microorganisms and control the changes in short time scales. It is noteworthy that the used method can handle uncertainty in the activated sludge process for short time scales, while the results for long scales may not be reliable.

Due to the absence of analytical solutions, numerical methods have been necessary for the computational simulation of most SDEs, especially nonlinear ones. In the modeling process, the results at each step, t_i compared to t_{i+1} , exhibit variations due to the presence of Brownian motion and W_i values. The objective is to achieve a numerical approximation that minimizes the error between observational and computational data, thereby obtaining optimal constant coefficients through such a solution. In this regard, analytical results demonstrate the convergence of the Euler-Maruyama method for small-time step sizes. The typical fluctuations observed in evaluating SDEs are practically absent in the activated sludge process, which is considered a macro physical phenomenon that eliminates Wiener process noises. The models developed based on SDE theory, as integrated noisy models, serve two fundamental purposes. Firstly, it is crucial to ascertain that the solution process exhibits a Markovian characteristic, wherein the value of S_{i+1} is contingent upon S_i . Secondly, the set of obtained results must belong to a broad class of martingales, guaranteeing that the solution process exhibits the characteristics of Brownian motion or Wiener process.

A comprehensive comparison between the final performance of ODE and SDE analysis with observational data is presented in Table 5. In this table, the data collected from the design using ODEs are related to the design of the sewage treatment plant system. The modeling was not based on ODE values. Instead, the ODE values were derived from figures provided by the World Health Organization (WHO) for WWTP. On the other hand, the observational data refer to the data gathered within a 15-day period from the actual performance of the sewage treatment plant. It is evident that the utilization of the SDE method has led to a reduction in model output uncertainties. Considering the utilization of ODEs in the design

Table 5 | Comparison of ODE and SDE data analysis results with observational data (population to design = 612,000 person for module1 of WWTP)

SDE (average) in 15 days	ODE (designing value)	Observation (average) in 15 days	Variable
22.834	20.00	23.0625	BOD _{eff} (mg L ⁻¹)
142,707.5	129,600	142,660.9	Q _{eff} (m ³ day ⁻¹)

section to obtain a solitary outcome for each variable, average observational data and SDE over a span of 15 days were used to compare the results. The findings unequivocally demonstrate a good agreement between the results obtained through the SDE method and the observational data. The proposed method significantly improves the results (derived from Equations (16) and (17)) by considering the randomness of the problem, resulting in an enhancement of 11.47 and 10.11% for the BOD and Q models, respectively.

5. CONCLUSIONS

Numerous physical and engineering phenomena exhibit a random nature, and neglecting the stochastic terms when modeling such processes with ODEs amplifies the uncertainty in the resulting outputs. Considering the assessment of uncertainty in modeling, SDEs offer a superior alternative to ODEs. This study aimed to derive the governing SDEs for the activated sludge process in the WWTP in Tabriz City, considering variations in the pollution parameters Q_{eff} and BOD_{eff} as biological operations. To address the fluctuations in the activated sludge process and numerically solve the Itô integral from the SDE, the Euler–Maruyama method was utilized. The findings suggest that the biological aspect of WWTP can be explored through SDE. However, in order to avoid accumulating errors and account for the stochastic nature of the process, a 15-day interval was deemed optimal for determining the constant coefficients and variables. The results indicate that the use of SDE over extended periods of time with close-time data, due to fluctuations in constant coefficients, may not yield satisfactory outcomes and is not recommended for WWTP design. The SDE technique has demonstrated its reliability in monitoring the variations in Q_{eff} and BOD_{eff} within short time periods. This solution effectively managed the uncertainty associated with the dynamic conditions of death and birth, active microorganisms, (i.e., regulating the optimal food concentration for these bacteria in the secondary sedimentation unit). It is important to emphasize that the utilization of these microorganisms in the WWTP system leads to substantial overhead expenses for the beneficiaries. In contrast to black-box methods, the SDE method offers the ability to model stochastic processes with minimal data, making it suitable for monitoring the performance of the activated sludge sector with limited observational data and within a short timeframe. Future studies should explore alternative methods for solving SDEs and modeling activated sludge, comparing them with the results presented in this paper. Subsequently, the outcomes from these models can be combined to enhance the accuracy of the output.

AUTHOR CONTRIBUTIONS

R. S. Z. rendered support in project administration, formal analysis, investigation, resources, data curation, and writing the original draft; V. N. conceptualized the whole article, supervised the article, and devised the methodology; M. S.-F. investigated the article, brought the resources, wrote the original draft, and wrote the review and edited the article; H. G. supervised the article and developed the methodology; C.-Q. K. supervised the article and devised the methodology.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Ahmadi, M. M., Mahdavi-rad, H. & Bakhtiari, B. 2017 Multi-criteria analysis of site selection for groundwater recharge with treated municipal wastewater. *Water Science and Technology* **76** (4), 909–919.

- Beck, M. B. 1987 *Water quality modeling: A review of the analysis of uncertainty*. *Water Resources Research* **23**, 1393–1442.
- Browning, A. P., Warne, D. J., Burrage, K., Baker, R. E. & Simpson, M. J. 2020 *Identifiability analysis for stochastic differential equation models in systems biology*. *Journal of the Royal Society Interface* **17** (173), 1–23.
- Compagnoni, E. M., Biggio, L., Orvieto, A., Proske, F. N., Kersting, H. & Lucchi, A. 2023 *An SDE for modeling SAM: Theory and insights*. In: *International Conference on Machine Learning. Proceedings of Machine Learning Research*, pp. 25209–25253.
- Cresson, J. & Sonner, S. 2018 *A note on a derivation method for SDE models: Applications in biology and viability criteria*. *Stochastic Analysis and Applications* **36**, 224–239.
- Dexter, N. C., Webster, C. G. & Zhang, G. 2016 *Explicit cost bounds of stochastic Galerkin approximations for parameterized PDEs with random coefficients*. *Computers & Mathematics with Applications* **71**, 2231–2256.
- Diehl, S. & Faràs, S. 2013 *A reduced-order ODE-PDE model for the activated sludge process in wastewater treatment: Classification and stability of steady states*. *Mathematical Models and Methods in Applied Sciences* **23**, 369–405.
- Erfanian, H. R., Hajimohammadi, M. & Abdi, M. J. 2016 *Using the Euler-Maruyama method for finding a solution to stochastic financial problems*. *International Journal of Intelligent Systems and Applications* **6**, 48–55.
- Evans, L. C. 2012 *An Introduction to Stochastic Differential Equations*. American Mathematical Society, Berkeley, CA.
- Fima, C. K. 1999 *Introduction to Stochastic Calculus with Applications*. Imperial College Press, London.
- Gul, S., Gani, K. M., Govender, I. & Bux, F. 2021 *Reclaimed wastewater as an ally to global freshwater sources: A PESTEL evaluation of the barriers*. *AQUA – Water Infrastructure, Ecosystems and Society* **70** (2), 123–137.
- Hanbay, D., Turkoglu, I. & Demir, Y. 2008 *Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks*. *Expert Systems with Applications* **34**, 1038–1043.
- Harris, C. 1977 *Modelling, simulation and control of stochastic systems with applications in wastewater treatment*. *International Journal of Systems Science* **8**, 393–411.
- Higham, D. J. 2001 *An algorithmic introduction to numerical simulation of stochastic differential equations*. *Society for Industrial and Applied Mathematics Review* **43**, 525–546.
- Kabouris, J. C. & Georgakakos, A. P. 1991 *Stochastic control of the activated sludge process*. *Water Science and Technology* **24**, 249–255.
- Lin, S. D. 2014 *Water and Wastewater Calculations Manual*. McGraw-Hill Education, New York.
- Mauritsson, G. 2013 *Simulation of Wastewater Treatment Plants Modeled by a System of Nonlinear Ordinary and Partial Differential Equations*. *Master's Theses in Mathematical Sciences*, Lund University.
- Metcalf, L. 2003 *Wastewater Engineering: Treatment and Reuse*, 4th edn. McGraw-Hill, New York.
- Nourani, V., Elkiran, G. & Abba, S. I. 2018 *Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach*. *Water Science Technology* **78** (10), 2064–2076.
- Nourani, V., Sayyah-Fard, M., Kantoush, S. A., Bhambe, K. P., Sumi, T. & Saber, M. 2023a *Optimization-based prediction uncertainty qualification of climatic parameters*. *Journal of Hydrometeorology* **24** (10), 1679–1697.
- Nourani, V., Shahidi Zonouz, R. & Dini, M. 2023b *Estimation of prediction intervals for uncertainty assessment of artificial neural network based wastewater treatment plant effluent modeling*. *Journal of Water Process Engineering* **55**, 104145.
- Ouldali, S., Leduc, R. & Nguyen, V. T. V. 1989 *Uncertainty modeling of facultative aerated lagoon systems*. *Water Research* **23** (4), 451–459.
- Pérez-aros, P., Quiñinao, C. & Tejo, M. 2022 *Control in probability for SDE models of growth population*. *Applied Mathematics & Optimization* **86**, 44.
- Richardson, M. 2009 *Numerical Methods for Option Pricing*. University of Oxford, Special Topic, Oxford.
- Särkkä, S. 2006 *Recursive Bayesian Inference on Stochastic Differential Equations*. *A Dissertation*, Helsinki University of Technology, pp. 1–249.
- Scott, L. E., Galpin, J. S. & Glencross, D. K. 2003 *Multiple method comparison: Statistical model using percentage similarity*. *Cytometry Part B: Clinical Cytometry: The Journal of the International Society for Analytical Cytology* **54**, 46–53.
- Sin, G., Gernaey, K. V., Neumann, M. B., Van Loosdrecht, M. C. & Gujer, W. 2009 *Uncertainty analysis in WWTP model applications: A critical discussion using an example from design*. *Water Research* **43**, 2894–2906.
- Wester, J., Timpano, K. R., Çek, D., Lieberman, D., Fieldstone, S. C. & Broad, K. 2015 *Psychological and social factors associated with wastewater reuse emotional discomfort*. *Journal of Environmental Psychology* **42**, 16–23.
- Wylie, E. B. & Streeter, V. L. 1978 *Fluid Transients*. McGraw-Hill International Book, New York.
- Xiu, D. & Hesthaven, J. S. 2005 *High-order collocation methods for differential equations with random inputs*. *SIAM Journal on Scientific Computing* **27**, 1118–1139.

First received 30 October 2023; accepted in revised form 30 January 2024. Available online 10 February 2024