Global NEST Journal, Vol 19, No 1, pp 140-159


# New explicit formulations for accurate estimation of aeration-related parameters in steady-state completely mixed activated sludge process

**YETILMEZSOY K.***

Department of Environmental Engineering, Faculty of Civil Engineering, Yildiz Technical University, 34220, Davutpasa, Esenler, Istanbul, Turkey
Received: 13/06/2015, Accepted: 13/03/2017, Available online: 29/03/2017
*to whom all correspondence should be addressed:
e-mail: yetilmez@yildiz.edu.tr ; kyetilmezsoy@gmail.com

**Abstract**

This paper presents new and explicit equations to estimate aeration-related parameters such as standard oxygen requirement, daily energy consumption and total mass transfer coefficient for the diffused aeration. The proposed formulations are derived for the steady-state completely mixed activated sludge process based on the nonlinear regression analysis by using the Richardson's extrapolation method and the Levenberg–Marquardt algorithm. The applicability of the proposed models has been investigated for a wide range of thirteen inputs consisting of the fundamental biological, hydraulic, and physical design variables, and tested against a total of 1500 additional computational scenarios. All estimations are proven to be satisfactory with very high determination coefficients ($R^2$) between 0.961–0.965, 0.967–0.972 and 0.980–0.984, respectively, for the prediction of standard oxygen requirement, daily energy consumption and total mass transfer coefficient for diffused aeration. The proposed models offer sufficiently simple and practical mathematical formulations incorporating routinely obtainable parameters, which are readily available for all activated sludge-based treatment plants. Besides eliminating the need for additional time or computational effort typically performed in the theoretical procedure, the developed equations have simple coefficients to be easily used for manual calculations with a hand-held calculator. The statistical results clearly exhibit that the proposed equations are accurate enough to be used in estimation of the studied aeration parameters based on the practical ranges of the corresponding design variables.

**Keywords:** Completely mixed activated sludge; Energy consumption; Mass transfer coefficient; Nonlinear regression; Oxygen requirement;Statistical analysis

## 1. Introduction

The activated sludge-based process currently represents one of the most widespread and commonly used technology to remove organic pollutants from the wastewater. The aeration tank, where the biological reactions occur and air (or oxygen) is injected in the mixed liquor, constitutes the hearth of this process. Nevertheless, aeration is a critical operation and a major energy consumer in most wastewater treatment plants (Roman and Mureşan, 2014). From the economical point of view, the supply of oxygen accounts for an important part of the running costs of an activated sludge-based wastewater treatment process. For conventional wastewater activated sludge plants, aeration systems are usually the single largest consumer of energy at wastewater treatment installations, typically accounting for 30% (or 45%) to 50% (or 60%) of a treatment facility's total electrical energy use (Bolles, 2006; Casey, 2009). Therefore, an effective control and optimization of the air supply may significantly reduce the operational costs of an activated sludge-based wastewater treatment plant (Makinia and Wells, 1999; Fayolle *et al.,* 2007). Aeration control aims not only at energy savings but will also guarantee that the microorganisms are adequately supplied with oxygen at all times (Roman and Mureşan, 2014). In this regard, it is clear that careful attention is essential to the adequate design, operation and control of aeration equipment, particularly at large wastewater treatment plants.

Diffused (or fine bubble) air systems are broadly classified as coarse or fine bubble systems, depending on the size of bubble generated. Fine bubble diffused air systems are probably superior to all other commercially available systems in terms of their energy efficiency in oxygen transfer (Casey, 2009; Fayolle *et al.,* 2007). However, under activated sludge process conditions, the oxygen transfer efficiency of diffused air systems is influenced by several factors, including geometry of the reactor, temperature, barometric pressure, and the liquid composition (Varolleghem, 1994). Because the characteristics of the aerobic process are time-varying and site-specific, it is necessary to calculate the mass transfer efficiencies under field conditions. Therefore, a number of process-related variables are needed to be taken into account with sufficient accuracy for purposes of optimal aeration control, failure diagnosis and process performance in full operation (Varolleghem, 1994). Considering these complex interactions of variables associated with this process, various aeration strategies can be implemented by means of computational methods for a realistic evaluation of the aeration system (Makinia and Wells, 1999; Makinia, 2010).

The oxygen is transferred to the water by a mass transfer coefficient ($k_La$), which describes how fast the oxygen is transferred to be dissolved in water. It has been stated that for a good aeration performance, the rate of dissolved oxygen supplied to the bioreactor should be equal to the rate of oxygen consumed by the mixed liquor under any given set of circumstances (Roman and Mureşan, 2014). In this regard, Painmanakul *et al.*(2009) have reported that it is frequently necessary to determine this coefficient when designing and evaluating the performance of the aeration systems. They also emphasized that a better forecast of the $k_La$ value would help the optimization of the installations in term of both cost and effectiveness. In another study, Al-Ahmady (2011) conducted a dimensional analysis procedure to evaluate the factors affecting the oxygen mass transfer coefficient ($k_La$).The study concluded that increasing airflow rate, diffusers coverage area and submergence of diffusers increased the value of $k_La$ while increasing Froude number, ratio of the height of water in the tank to the length of the tank, and bubbles diameter showed an adverse influence on this coefficient. Concerning to the estimation of the value of mass transfer coefficient, however, there are almost no papers in the literature proposing a practical equation that takes into account the most common biological, hydraulic, and physical design variables used in the design of activated sludge plants. Many of the proposed correlations are highly theoretical, and not always applicable in practice, since the most of parameters used in the mathematical structure of these models are not readily available or routinely obtainable for all activated sludge-based treatment plants. On the other hand, although some of the previous expressions (Chen*et al.,* 1980; Goto and Andrews, 1985; Holmberg, 1986; Reinius and Hultgren, 1988) seem to have a simple mathematical structure, however, they neglect the effect of several process-related variables, and do not directly reflect the actual behavior the large-scale aeration units. To overcome the limitations and problems associated with the existing models, a more practical approach for the accurate prediction of the value of the mass transfer coefficient could be interesting for engineers and researchers who are concerned with the aeration in activated sludge process.

Mathematical modeling and computer simulation of biological systems are valuable and powerful tools for describing and evaluating their performance under both dynamic and steady-state conditions (Yetilmezsoy, 2010; Kumar, 2011; Yetilmezsoy, 2012; Yetilmezsoy *et al.,* 2015). With the increasing practical experience, continuously developed and improved activated sludge-based models have been utilized to quantify and to evaluate the process performance (Makinia and Wells, 1999; Makinia, 2010; Yetilmezsoy, 2010; Yetilmezsoy, 2016). In recent years, various mechanistic models have been introduced for the assessment of a number of activated sludge-based problems such as modeling of activated sludge thickening in secondary clarifiers (Giokas *et al.,* 2002), modeling of the steady-state biofilm activated sludge reactor under substrate limiting conditions (Fouad and Bhargava, 2005a; Fouad and Bhargava, 2005b), modeling of temperature dynamics for activated sludge systems (Makinia *et al.,* 2005), estimation of completely mixed activated sludge reactor volume (Yetilmezsoy, 2010), prediction of the reduction of biosolids production by ozonation of the return activated sludge (Isazadeh *et al.,* 2014), and prediction of the waste sludge volumetric flow rate (Yetilmezsoy, 2016). Although many other studies (Nuhoglu *et al.,* 2005; Mulas, 2006; Pamukoglu and Kargi, 2007; Szilveszter *et al.,* 2010; Bagheri *et al.,* 2015; Liu and Wang, 2015) have focused on different modeling methodologies in the operation of activated sludge-based treatment plants, however, to date, there are no sound papers in literature regarding the development of explicit mathematical formulations that can be directly used for the prediction of the present aeration-related parameters in the steady-state completely mixed activated sludge process without writing a set of theoretical statements.To the best of the author's knowledge, this work is the first study specifically aimed at investigating new and practical expressions for the direct estimation of aeration-related parameters as a function of the most common biological, hydraulic, and physical design variables used in the design.

In consideration of the foregoing facts, the overall objectives of this study were: (1) to develop simple and explicit equations for practising engineers and researchers, which makes it possible to accurately estimate aeration-related parameters (standard oxygen requirement, daily energy consumption, and total mass transfer coefficient for diffused aeration) for the diffused aeration in the steady-state completely mixed activated sludge process; (2) to verify the model predictions by means of various powerful statistical performance indicators; (3) to assess the predictive capabilities of the developed models by comparing the model outputs with the results of a total of 1500 additional and different computational scenarios; and (4) to validate the applicability of the proposed equations by comparing the consistency of simulation results with the existing literature data.

## 2. Methodology

### *2.1. Representation of input and output variables*

In the planning stage of modeling and simulation-based studies, selection of the most appropriate model components is a crucial factor in order to recognize possible technical faults and to reduce computation time, as well as to develop an accurate modeling methodology for a specific environmental process (Yetilmezsoy *et al.,* 2015; Kanat and Saral, 2009; Yetilmezsoy and Sapci-Zengin, 2009). For the present case, the model variables and their respective ranges were chosen in accordance with the relevant literature (Yetilmezsoy, 2010; Yetilmezsoy, 2016;

Orhon and Artan, 1994; Muslu, 1996a; Muslu, 1996b; Qasim, 1998; Crites and Tchobanoglous, 1998).

For modeling and simulation purposes, thirteen fundamental biological, hydraulic, and physical design variables, which are the most commonly used design parameters, were considered as the following inputs: $X_1$ = $Q$: influent wastewater flow rate (m$^3$/sec), $X_2$ = $\vartheta_c$: mean cell residence time (days), $X_3$ = $Y$: growth yield coefficient (kg MLVSS/kg $BOD_5$), $X_4$ = $S_i$: influent soluble substrate concentration (kg $BOD_5$/m$^3$), $X_5$ = $X$: concentration of the cells (volatile suspended solids) in the reactor (kg MLVSS/m$^3$), $X_6$ = $\varphi$ = $X/X_{TSS}$: volatile suspended solids to total suspended solids ratio (kg MLVSS/kg MLSS), $X_7$ = $S_e$: effluent total substrate concentration (kg $BOD_5$/m$^3$), $X_8$ = $X_e$: effluent total suspended solids concentration (kg MLSS/m$^3$), $X_9$ = $k_d$: microorganism endogenous decay coefficient (day$^{-1}$), $X_{10}$ = $T_a$: ambient air temperature (°C), $X_{11}$ = $T_i$: influent wastewater temperature (°C), $X_{12}$ = $H$: static pressure caused by wastewater depth in the areation basin, measured in head of water (m), and $X_{13}$ = $H_a$: elevation or altitude above sea level (m). These input variables were used for the estimation of three aeration-related outputs: $f_1$= $SOR_d$: standard oxygen requirement for the diffused aeration (kg $O_2$/h), $f_2$= $E_c$: daily energy consumption (kWh/day), and $f_3$= $k_La_d$: total mass transfer coefficient for the diffused aeration (day$^{-1}$).

Detailed definitions of the present model components, which are among the most widely used and monitored parameters in activated sludge-based treatment plants, can be found in several studies (Bolles, 2006; Yetilmezsoy, 2010; Orhon and Artan, 1994; Muslu, 1996a; Muslu, 1996b; Qasim, 1998; Crites and Tchobanoglous, 1998; Eroglu, 1991; Celenza, 1999; Toprak,, 2000; Shammas *et al.,* 2009; Loehr, 2012).

The general schematic of the major parameters used in the proposed models is depicted in Figure 1.

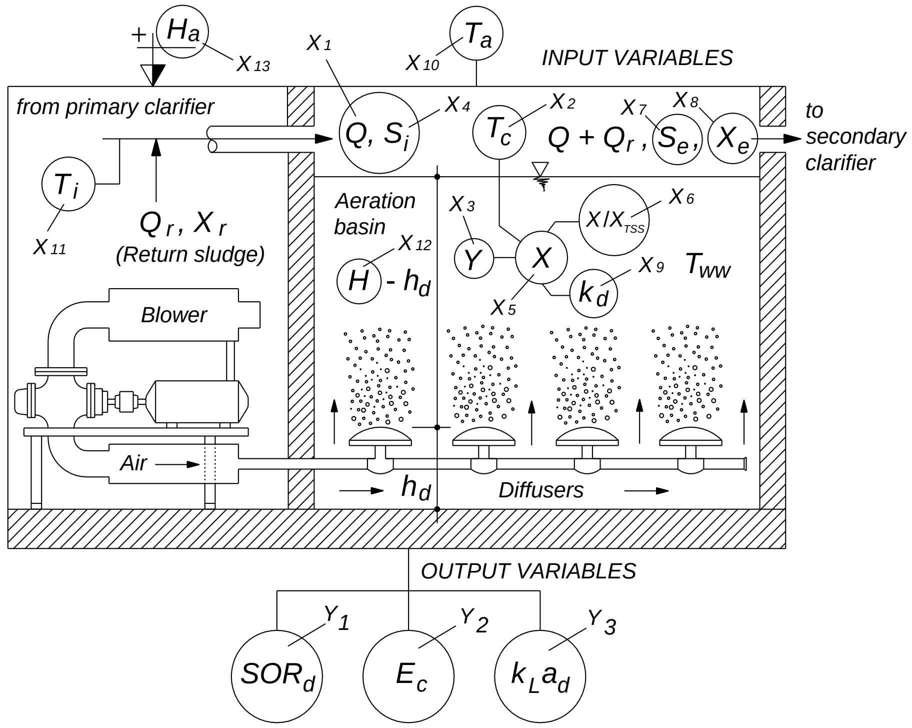


**Figure 1.** General schematic of the major parameters used in the proposed models

*2.2. Steady-state procedure for the completely mixed activated sludge process*

In this study, a theoretical procedure (Table 1) was conducted as the primary step of the nonlinear-based computational analysis to produce a sufficient number of data points to be used in modeling and simulation of standard oxygen requirement for the diffused aeration ($SOR_d$), daily energy consumption ($E_c$) and mass transfer coefficient ($k_La_d$) for diffused aeration.

*2.3. Assumptions*

In this study, the following assumptions are made to describe the specific aspects of the process: (1) the system is assumed to run under a steady-state condition for biomass ($dX/dt$ = 0, $X_0$ = 0) and substrate ($dS/dt$ = 0), (2) the volume used for calculation of mean cell residence time ($\vartheta_c$) accounts volume of the aeration tank only, (3) completely mixed flow regime is maintained in the aeration tank, (4) wastewater was distributed along with return sludge uniformly from one side of the tank to the other, (5) waste stabilization occurs only in the aeration tank, (6) all reactions take place in the aeration basin so that the substrate in the aeration basin is of the same concentration as the substrate in the secondary clarifier and in the effluent, (7) there is no microbial degradation of organic matter and no biomass growth in the secondary clarifier, and (8) waste stabilization is carried out by the microorganisms occurs in the aerator unit. The design parameters and numerical assumptions considered in the present computational analysis are summarized in Table 2.

**Table 1.**Set of steady-state designequations considered in theoretical procedure

| Steady-state design equations | Equation no |
|---|---|

**[1] The effluent soluble substrate concentration (*C*: mg BOD$_5$/l) and the substrate utilization efficiency (or biological treatment efficiency: *$E_b$*) are calculated as follows (Yetilmezsoy, 2010; Yetilmezsoy, 2016; Muslu, 1996a; Muslu, 1996b):**

$$C = S_e - (X_e \cdot f_b \cdot A \cdot \psi) \tag{1a}$$

$$E_b(\%) = \left[(S_i - C)/S_i\right] \times 100 \tag{1b}$$

*where $S_i$ is the influent substrate concentration (mg BOD$_5$/l), $S_e$ is the total effluent substrate concentration as the discharge standard for the receiving water (mg BOD$_5$/l), $X_e$ is the effluent total suspended solids concentration (mg MLSS/l), $f_b$ is the biodegradable fraction of $X_e$, A is the ultimate biochemical oxygen demand (BOD$_u$ ≈ COD) of per kg of bacteria cells (or the oxygen requirement for oxidizing per unit of biomass: g O$_2$ equivalent/g MLVSS), and ψ is the ratio of BOD$_5$ to ultimate BOD$_u$ (commonly ψ = BOD$_5$/BOD$_u$ = 0.68 for the substrate decay coefficient of k = 0.1 day$^{-1}$).*

**[2] The volume of the completely mixed activated sludge reactor (*$V_R$*: m$^3$) is determined by the following steady-state or empirical equations (Yetilmezsoy, 2010; Fouad and Bhargava, 2005a; Fouad and Bhargava, 2005b; Muslu, 1996a; Muslu, 1996b; Toprak, 2000):**

$$V_R = \frac{(Q)(Y)(\theta_c)(S_i - C)}{(X)(1 + k_d \cdot \theta_c)} = \left[\frac{(Q)(\theta_c)(S_i - C)}{X}\right]\left[\frac{Y}{(1 + k_d \cdot \theta_c)}\right] = \frac{(Q)(Y_{obs})(\theta_c)(S_i - C)}{X} \tag{2a}$$

$$V_R = \frac{(0.294)(Q)(Y)(\theta_c)^{0.7}(S_i)^{1.1}}{(X)(k_d)^{0.33}(C)^{0.04}} \tag{2b}$$

*where Q is the influent wastewater flow rate (m$^3$/day), ϑ$_c$ is the mean cell residence time (MCRT) or the sludge age (day) or the solids retention time (SRT), Y is the growth yield coefficient (kg MLVSS/kg BOD$_5$), $Y_{obs}$ is the observed growth yield coefficient (kg MLVSS/kg BOD$_5$), X is the concentration of the cells (volatile suspended solids) in the reactor (kg MLVSS/m$^3$), $k_d$ is the coefficient of endogenous respiration (day$^{-1}$), and others ($S_i$ and C) are defined in previous equations.*

**[3] The food to microorganism (F/M: kg BOD$_5$/kg MLVSS/day) ratio, the volumetric organic loading (*$L_v$*: kg BOD$_5$/m$^3$/day), the hydraulic retention time (*ϑ$_h$*: hour), and the waste (excess) sludge mass flow rate (*$P_x$*: kg MLVSS/day) as a function of *$S_i$*, *Q*, *X*, *$V_R$*, *ϑ$_c$*, *$k_d$*, *Y*, *$E_b$*, substrate utilization rate (*$r_{su}$*: kg BOD$_5$/m$^3$/day) are obtained from the following equations (Yetilmezsoy, 2016; Mulas, 2006; Orhon and Artan, 1994; Muslu, 1996a; Muslu, 1996b; Qasim, 1998; Toprak, 2000; US EPA, 1977):**

$$\frac{1}{\theta_c} = \frac{YQ(S_i - C)}{XV_R} - k_d = \frac{YQ\left(\frac{S_i}{S_i}\right)(S_i - C)}{XV_R} - k_d = \frac{YQ(S_i)\left(\frac{S_i - C}{S_i}\right)}{XV_R} - k_d = \left(\frac{YQS_iE_b}{XV_R}\right) - k_d \tag{3a}$$

$$\frac{1}{\theta_c} = Y\left(\frac{QS_i}{XV_R}\right)E_b - k_d = Y\left(\frac{F}{M}\right)E_b - k_d \rightarrow \frac{F}{M} = \left(\frac{1}{\theta_c} + k_d\right)\frac{1}{(Y)(E_b)} \tag{3b}$$

$$\theta_h = \frac{V_R}{Q} \rightarrow L_v = \frac{QS_i}{V_R} = \frac{S_i}{\theta_h} \rightarrow L_v = \left(\frac{X}{X}\right)\left(\frac{QS_i}{V_R}\right) = X\left(\frac{QS_i}{XV_R}\right) = X\left(\frac{F}{M}\right) \tag{3c}$$

$$P_x = (-Yr_{su} - k_dX)V_R = \left[-Y\left(-\frac{S_i - C}{\theta_h}\right) - k_dX\right]V_R = YQ(S_i - C) - k_dXV_R \tag{3d}$$

*In the computational analysis, the following control ranges were considered for the present activated sludge process (Yetilmezsoy, 2010; Yetilmezsoy, 2016; Muslu, 1996a; Muslu, 1996b; Qasim, 1998): F/M = 0.2–0.6 kg BOD$_5$/kg MLVSS/day, $L_v$ = 0.8–2.0 kg BOD$_5$/m$^3$/day, ϑ$_c$ = 5–15 days, ϑ$_h$ = 3–5 h, and r = $Q_r$/Q from 0.20 (or 0.25) to 0.50 (or 1.0).*

**[4] The wastewater temperature within the aeration basin (*$T_{ww}$*: °C) is calculated by conducting a heat balance around the reactor, resulting the following equation (Muslu, 1996a; Celenza, 1999; Loehr, 2012; Wang *et al.,* 2009):**

$$T_{ww} = \frac{(A_R)(f)(T_a) + (Q)(T_i)}{(A_R)(f) + Q} = \frac{(V_R/H)(f)(T_a) + (Q)(T_i)}{(V_R/H)(f) + Q} \tag{4}$$

*where $A_R$ is the surface area of the aeration basin (m$^2$), f is the proportionality factor (f = 0.5 for Eastern United States and f = 2.5 for Midwestern United States), $T_a$ is the ambient air temperature (°C), $T_i$ is the influent wastewater temperature (°C), H is the average wastewater depth in the areation basin (m), and others (Q and $V_R$) are defined in previous equations.*

**[5] The air pressure (*$p_a$*) and the density of air (*$\rho_a$*) as a function of altitude *$H_a$* (above sea level) are determined by using the following equations (Bugbee and Blonquist, 2006):**

$$p_1 = p_0\left(1-\frac{LH_a}{T_0}\right)^{\frac{gM}{RL}} \tag{5a}$$

$$p_2 = p_0\left[1-\left(\frac{2.25577}{10^5}\times H_a\right)\right]^{5.25588} \tag{5b}$$

$$p_3 = p_0 - p_0\left(1-\left[1-\left(\frac{H_a}{44307.69231}\right)\right]^{5.25328}\right) \tag{5c}$$

$$p_a = \frac{1}{n}\sum_{i=1}^{n=3} p_i = \frac{p_1+p_2+\ldots+p_n}{n} = \frac{p_1+p_2+p_3}{3} \tag{5d}$$

$$\rho_a = \frac{(p_a\times 10^3)(M_a)}{RT} = \frac{(p_a\times 10^3)(M_a)}{R(T_0-LH_a)} \tag{5e}$$

*where $p_1$, $p_2$ and $p_3$ are the air pressure (kPa) at altitude $H_a$ (m), $p_0$ is the sea level standard atmospheric pressure (101.325 kPa), $p_a$ is the average air pressure (kPa) to be used in calculation of the density of air, L is the temperature lapse rate (0.0065 K/m), $T_0$ is the sea level standard temperature (288.15 K), g is the earth-surface gravitational acceleration (9.80665 m/s$^2$), $M_a$ is the molar mass of dry air (0.0289644 kg/mol), R is the ideal (universal) gas constant (8.31447 J/mol/K), and $\rho_a$ is the density of air based on the molar form of the ideal gas law (kg/m$^3$).*

**[6] The effective oxygen percentage as a function of barometric pressure ($p_a$: kPa) from sea level to any elevation is determined as follows (Bugbee and Blonquist, 2006):**

$$O_2(\%) = 20.95 + (100/p_a)(0.2095)(p_a - p_0) \tag{6}$$

**[7] The theoretical oxygen requirement (ThOR: kg $O_2$/day) is calculated by knowing influent and effluent BOD of wastewater, and the amount of organisms wasted from the system (Yetilmezsoy, 2010):**

$$ThOR = \left(\frac{Q(\Delta S)}{\psi}\right) - (A)(P_x) = \left(\frac{Q(S_i - S_e)}{\psi}\right) - (A)\left(\frac{XV_R}{\theta_c}\right) \tag{7}$$

**[8] The dissolved oxygen saturation concentration ($C_s$: mg $O_2$/l) at a given temperature (between $T$ = 4°C – 30°C) and total dissolved solids (TDS) concentration (mg/l) is obtained by one of the following empirical equations (Eroglu, 1991; ASCE, 1997; von Sperling, 2007; Ma *et al.,* 2013):**

$$C_s = \frac{468}{(31.6+T)} - \frac{(0.0036\times \mathrm{TDS})}{(21.2+T)} \tag{8a}$$

$$C_s = (475 - 0.00265\times \mathrm{TDS})/(33.5+T) \tag{8b}$$

$$C_s = 14.652 - (4.1022\times 10^{-1})(T) + (7.9910\times 10^{-3})(T^2) - (7.7774\times 10^{-5})(T^3) \tag{8c}$$

**[9] The comprehensive oxygen solubility correction factor ($F_a$) is calculated from the following equations (Muslu, 1996b; Qasim, 1998; von Sperling, 2007):**

$$F_a = 1 - (H_a/9450) \tag{9a}$$

$$F_a = \frac{1}{2}\left[\frac{(10.33)(p_a/101.325) + H}{10.33} + (1 - E_{O_2})\right] \tag{9b}$$

*where $E_{O2}$ is the oxygen transfer efficiency of air diffusers (or mass of $O_2$ transferred/mass of $O_2$ supplied, usually $E_{O2}$ = 0.06 – 0.12), 10.33 and 101.325 are the water column (m) and kilopascal (kPa) equivalents of the absolute atmospheric pressure at sea level (1 atm), respectively, and others ($p_a$ and H) are defined in previous equations.*

**[10] Based on the above definitions, the value of standard oxygen requirement (*SOR*: kg $O_2$/h) is calculated from the following equation (Roman and Mureşan, 2014; Makinia, 2010; Muslu, 1996b; Qasim, 1998; Makinia and Wells, 2000):**

$$SOR = \frac{1}{24}\left(\frac{\mathrm{ThOR}}{\left[\left(C_{ww}\beta F_a - C_{\min}\right)/C_{20}\right]\alpha\theta^{(T_{ww}-20)}}\right) \tag{10}$$

*where $C_{ww}$ is the dissolved $O_2$ concentration (mg $O_2$/l) in wastewater at temperature $T_{ww}$, β is the salinity surface tension factor (a ratio of $O_2$ saturation concentration in wastewater to that in clean water) for wastewater (usually β = $C_{20(wastewater)}/C_{20(clean\ water)}$ = 0.70 – 0.98 and typically 0.90 for wastewater), $C_{min}$ is the minimum dissolved $O_2$ concentration (mg $O_2$/l) maintained in the areation basin, $C_{20}$ is the dissolved $O_2$ concentration (mg $O_2$/l) at standard 20 °C, α is the oxygen transfer correction factor (a ratio of $O_2$ transfer in wastewater to that in clean water) for diffused aeration* *(usually $\alpha = k_La_{(wastewater)}/k_La_{(clean\ water)} = 0.40 - 0.80$), $\vartheta$ is the Arrhenius constant or temperature correction coefficient (usually $\vartheta = 1.015 - 1.040$ and typically $\vartheta = 1.024$) for both diffused and mechanical aeration applications, and others (ThOR and $T_{ww}$) are defined in previous equations.*

**[11] The peak factor ($T_1$: h/day) for oxidizing carbonaceous biological matter (or fluctuation factor for $BOD_5$ concentration) is selected depending on the equivalent population ($E_p$), which is computed for a given amount of wastewater ($q$: l/capita/day) discharged per capita and per day as follows (Toprak, 2000):**

$$E_p = \frac{Q}{q} = f(T_1) = \begin{cases} T_1 = 8 & if\ \ E_p \le 5000 \\ T_1 = 12 & if\ \ 5000 < E_p \le 20{,}000 \\ T_1 = 15 & if\ \ 20{,}000 < E_p \le 100{,}000 \\ T_1 = 17 & if\ \ 100{,}000 < E_p \le 300{,}000 \\ T_1 = 21 & if\ \ E_p > 300{,}000 \end{cases} \tag{11}$$

**[12] The total hourly oxygen requirement ($R_o$: kg $O_2$/h) and the standard oxygen requirement ($O_c$: kg $O_2$/h) are also determined as complementary alternatives to Eqs. (7) and (10), respectively, from the following equations (Eroglu, 1991; Toprak, 2000):**

$$R_o = \left(\frac{1}{T_1}\right)(a)\left(\frac{E_b}{100}\right)\left(\frac{QS_i}{10^3}\right) + \left(\frac{1}{T_2}\right)(k_{re})\left(\frac{X/\varphi}{10^3}\right)(V_R) \tag{12a}$$

$$O_c = \left(\frac{R_o}{\sigma}\right)\left(\frac{C_{10}}{C_{ww} - C_{\min}}\right)\sqrt{\frac{D_{10}}{D_{T_{ww}}}} = \left(\frac{R_o}{\sigma}\right)\left(\frac{C_{10}}{C_{ww} - C_{\min}}\right)(1.0188)^{(10 - T_{ww})} \tag{12b}$$

*where $T_1$ is the peak factor for oxidizing carbonaceous biological matter (or fluctuation factor for $BOD_5$ concentration: hour/day), a is the oxygen requirement for oxidizing carbonaceous biological matter (or oxygen requirement for removing per unit of $BOD_5$: kg $O_2$/kg $BOD_5$), $T_2$ is the time factor for endogenous respiration (24 h/day), $k_{re}$ is the respiration rate coefficient (kg $O_2$/kg MLSS/day), σ is a correction factor for dissolved $O_2$ saturation concentration, $C_{10}$ is the dissolved $O_2$ concentration (mg $O_2$/L) at 10°C, $D_{10}$ ve $D_{Tww}$ are the diffusion coefficients ($10^{-9}$ $m^2$/s) at 10°C and temperature $T_{ww}$, respectively, and others ($E_b$, Q, $S_i$, X, φ, $V_R$, $C_{ww}$, and $C_{min}$) are given in previous parts.*

**[13] Based on the data (Eroglu, 1991) showing the effect of the sludge load ($L_s$: kg $BOD_5$/kg MLSS/day) on the respiration rate coefficient ($k_{re}$: kg $O_2$/kg MLSS/day), a third order inverse polynomial empirical function with a very high determination coefficient ($R^2$ = 0.9994, minimum residual = –0.00285, maximum residual = 0.00158) was also derived by the author within the scope of the present study as follows:**

$$L_s = \frac{QS_i}{(X/\varphi)(V_R)} \rightarrow k_{re} = 0.23173 - \left(\frac{0.03705}{L_s}\right) + \left(\frac{4.07896 \times 10^{-3}}{L_s^{\,2}}\right) - \left(\frac{1.39049 \times 10^{-4}}{L_s^{\,3}}\right) \tag{13}$$

**[14] According to Eqs. (10) and (12b), the average standard oxygen requirement for the diffused aeration ($SOR_d$: kg $O_2$/h), which is used as the first output variable ($f_1$) in the subsequent modeling study, is obtained as follows to take into account the effects of different parameters (i.e., $F_a$, $T_1$, $k_{re}$, etc.):**

$$SOR_d = \frac{SOR + O_c}{2} \rightarrow f_1(X_i) = f_1(X_1, X_2, X_3, \ldots, X_n) \tag{14}$$

**[15] Based on the value of standard oxygen requirement obtained from Eq. (14), the theoretical air requirement under field conditions (*ThARfc*: $m^3$/h), the theoretical air requirement (*ThAR*: $m^3$/h), actual air requirement (*AAR*: $m^3$/h), and the air flow rate ($Q_b$) in $m^3$/min are computed from the following expressions:**

$$ThARfc = \frac{SOR_d}{(\rho_a)\left[\mathrm{O_2}(\%)/100\right]} \rightarrow ThAR = \frac{ThARfc}{E_{O_2}} \rightarrow AAR = (ThAR)(\mathrm{SF})_{\mathrm{a}} \rightarrow Q_b = \frac{AAR}{60} \tag{15}$$

*where $(SF)_a$ is the safety factor for air requirement, and others ($SOR_d$, $\rho_a$, $O_2(\%)$, and $E_{O2}$) are defined in previous equations.*

**[16] The daily energy consumption ($E_c$: kWh/day) of the diffused aeration system is computed based on the assumption of adiabatic conditions as follows (Casey, 2009; Muslu, 1996b; Qasim, 1998):**

$$E_c = (P_m)(t) = \left(\frac{(w)(R)(T_0)}{(8.41)(\eta_b)}\left[\left(\frac{p_{out}}{p_{in}}\right)^{0.283} - 1\right](\mathrm{SF})_{\mathrm{b}}\right)(t) \tag{16a}$$

$$E_c = \left( \frac{(Q_b \rho_a)(R)(T_a + 273)}{(8.41)(\eta_b)(60 \text{ s/min})} \left[ \left( \frac{\left(\sum h_L + (H - h_d) + 10.33\right)/10.33}{\left(p_a / 101.325\right)} \right)^{0.283} - 1 \right] (\text{SF})_b \right) (t) \tag{16b}$$

*where $P_m$ is the total power requirement (kW) for blowers, t is the effective operating time (i.e., t = 24 h/day if the power is supplied at full power all day), w is the air mass flow rate (kg/s), 8.41 is the constant (kg/kmol), $\eta_b$ is the blower efficiency for the diffused aeration system, $T_0$ is the inlet or ambient air temperature ($T_0 = T_a + 273$ [=] K), $p_{out}$ is the absolute outlet or air supply pressure (atm), $p_{in}$ is the absolute inlet or ambient barometric pressure (atm), $\sum h_L$ is the total head losses (see Table 1) in air piping system and diffusers (m), $h_d$ is the distance between the bottom of the areation basin and the top of submerged diffusers, $(SF)_b$ is the safety factor for blower power, and others ($Q_b$, $\rho_a$, R, $T_a$, $p_a$, and H) are explained in previous steps.*

**[17] According to Eq. (16b), the daily energy consumption ($E_c$: kWh/day) of the diffused aeration system, which is considered as the second output variable ($f_2$) in the nonlinear regression analysis, is defined as follows:**

$$E_c = (P_m)(t) \rightarrow f_2(X_i) = f_2(X_1, X_2, X_3, ..., X_n) \tag{17}$$

**[18] Finally, the total mass transfer coefficient for the diffused aeration ($k_L a_d$: day$^{-1}$) is computed from the following equation (Henderson, 2002):**

$$k_L a_d = \frac{(\alpha_f)(R_{air})(H - h_d)(Q_b)(60 \text{ min/h})(24 \text{ h/day})}{(C_{20})(V_R)} \tag{18}$$

*where $\alpha_f$ is the admixtures correction factor, $R_{air}$ is the $O_2$ flow per volume and immersion depth in clean water at $S_0 = 0$ and a specified temperature T = 20°C (g/m$^3$/m), and others (H, $h_d$, $C_{20}$, and $V_R$) are expressed in previous equations.*

**[19] Based on the gas/liquid mass transfer (film or penetration) theory, $k_L a_d$ describes the oxygen transfer coefficient for the diffused aeration, which has the dimension of the inverse of time, $a_d$ is defined as $A_G/V_L$ (where $A_G$: total gas surface [=] m$^2$, and $V_L$: liquid volume [=] m$^3$) and describes the interfacial area (m$^2$/m$^3$) in diffused aeration, while $k_L$ is the liquid film coefficient (m/s) and describes the velocity of the transport (Henkel, 2010).**

**According to Eq. (19), the total mass transfer coefficient for the diffused aeration ($k_L a_d$: day$^{-1}$), which is considered as the third output variable ($f_3$) in the nonlinear regression analysis, is given as follows:**

$$k_L a_d = f(\alpha_f, R_{air}, H, h_d, Q_b, C_{20}, V_R) \rightarrow f_3(X_i) = f_3(X_1, X_2, X_3, ..., X_n) \tag{19}$$

**Table 2.** Design parameters and assumptions to be used in the computational analysis

| Constituents | Values used in the computational analysis |
|---|---|
| Empirical formula of bacteria cells | $C_5H_7NO_2$ (MW = 113 g/mol) (Yetilmezsoy, 2010; Comeau, 2008) |
| Ultimate biochemical oxygen demand ($BOD_u \approx COD$) of per kg of bacteria cells (or oxygen requirement for oxidizing per unit of biomass) | *A* = 1.42 g $O_2$ equivalent/g cell (MLVSS) (Yetilmezsoy, 2010; Martinez *et al.,* 2004) |
| Biodegradable fraction of $X_e$(in terms of effluent total suspended solids concentration) | $f_b$ = 65% (Muslu, 1996b; Toprak, 2000) |
| Substrate decay coefficient | *k* = 0.1 day$^{-1}$ (Yetilmezsoy, 2010) |
| Ratio of $BOD_5$ to ultimate BOD ($BOD_u$) | $\psi = BOD_5/BOD_u = 0.68$ (Celenza, 1999) |
| Propotionality factor (or overall heat transfer coefficient) | *f* = 0.50 (Muslu, 1996b; Celenza, 1999) |
| Oxygen transfer efficiency of air diffusers (or mass of $O_2$ transferred/mass of $O_2$ supplied) ($E_{O2}$ = 0.06 – 0.12) | 8% or $E_{O2}$ = 0.08 (Qasim, 1998) |
| Salinity surface tension factor (a ratio of $O_2$ saturation concentration in wastewater to that in tap water) for wastewater ($\beta$ = 0.70 – 0.98) | $\beta$ = 0.90 (Roman and Mureşan, 2014; Qasim, 1998; Toprak, 2000) |
| Minimum dissolved $O_2$ concentration maintained in the areation basin ($C_{min}$= 1.0 or 1.5 – 2.0 mg/l) | $C_{min}$= 1.5 mg/l (Bolles, 2006; Toprak, 2000; Ghangrekar and Kharagpur, 2014) |
| Oxygen transfer correction factor (a ratio of $O_2$ transfer in wastewater to that in tap water) for diffused aeration ($\alpha$ = 0.4 – 0.8) | $\alpha$ = 0.75 (Roman and Mureşan, 2014; Qasim, 1998) |
| Arrhenius constant or temperature correction coefficient ($\vartheta$ = 1.015 – 1.040) for diffused aeration application | $\vartheta$ = 1.024 (Qasim, 1998; Toprak, 2000) |
| Safety factors for air requirement and blower power | $(SF)_a$ = 1.50 (Bolles, 2006; von Sperling, 2007) and $(SF)_b$ = 1.20, respectively. |
| Blower efficiency for diffused aeration systems ($\eta_b$ = 0.70 – 0.90) | 75% or $\eta_b$ = 0.75 (Bolles, 2006) |
| Amount of wastewater discharged per capita and per day | *q* = 200 l/capita/day |
| Oxygen requirement for oxidizing carbonaceous biological matter (or oxygen requirement for removing per unit of $BOD_5$) (*a* = 0.45 – 0.65 kg $O_2$/kg $BOD_5$) | *a* = 0.5 kg $O_2$/kg $BOD_5$ (Eroglu, 1991; Toprak, 2000) |

| | |
|---|---|
| Respiration rate coefficientused for computation of $R_o$. | $k_{re}$= 0.010 – 0.20 kg $O_2$/kg MLSS/day depending on the $L_s$ (Eroglu, 1991) |
| Peak factor for oxidizing carbonaceous biological matter (or fluctuation factor for $BOD_5$ concentration) ($T_1$= 8 – 21 hour/day) | $T_1$= 8 – 21 h/day depending on the equivalent population (Toprak, 2000) |
| Time factor for endogenous respiration | $T_2$= 24 h/day (Toprak, 2000) |
| A correction factor for dissolved $O_2$ saturation concentration($\sigma$ = 0.60 – 0.90 for wasterwater, and $\sigma$ = 1.0 for pure water) | $\sigma$ = 0.80 (Eroglu, 1991; Toprak, 2000) |
| Oxygen flow per volume and immersion depth in clean water ($S_0$ = 0, $T$ = 20°C) | $R_{air}$ = 16 g/m$^3$/m (Makinia and Wells, 2000) |
| Total head losses in air pipings = 5-25 cm | 25 cm (Muslu, 1996b; Qasim, 1998) |
| Air filter losses for the centrifugal blower = 13-76 mm | 50 mm (Muslu, 1996b; Qasim, 1998) |
| Silencer losses for the centrifugal blower = 13-38 mm | 30 mm (Muslu, 1996b; Qasim, 1998) |
| Check valve and fittings losses = 20-203 mm | 200 mm (Muslu, 1996b; Qasim, 1998) |
| Diffuser losses for fine bubble aeration = 40-50 cm | 400 mm (Muslu, 1996b; Qasim, 1998) |
| Allowance for clogging of diffusers and miscellaneous head losses under emergency conditions | 100 cm (Muslu, 1996b; Qasim, 1998) |
| Depth of submergence (water depth above the diffusers) | ($H$ – $h_d$) = ($H$ – 0.30) m (Muslu, 1996b; Qasim, 1998) |

*2.4. Computational procedure*

In this study, MATLAB® R2009b (V7.9.0.529, The MathWorks, Inc., Natick, MA, US) and DataFit® (V8.1.69, Oakdale Engineering, PA, US) softwares running on a CPU N280 (Intel® Atom™ Processor 1.66 GHz, 0.99 GB of RAM) PC, were used for simulation and modeling purposes, respectively. The main structure of the present calculation protocol and the corresponding computational pathways are depicted in Figure 2.

In the first step of the computational analysis, A MATLAB®-based algorithm (Figure 2) was constructed for the processing of a series of compulsory design calculations (i.e., several biological conversions, determination of reactor volume, definition of a heat mass balance around the aeration basin, calculation of air pressure as a function of the elevation factor, determination of the saturation concentration of the oxygen at operating temperature, determination of oxygen content and density of air as a function of absolute atmospheric pressure at elevation above sea level, formulation of equivalent population as a function of peak factor (or fluctuation factor for $BOD_5$), computation of required blower power, etc.) based on the theoretical procedure (see *Section 2.2*).

In the second step of the computational analysis, a total of 500 different operating scenarios were implemented at random for simulating a wide range of thirteen inputs consisting of thirteen fundamental biological, hydraulic, and physical variables (see *Section 2.1*). Then the simulated data was stored as a [500×16] matrix (for $X_i$ and $Y_j$, $i$ = 1,2,..13 and $j$ = 1,2,3) in the workspace of MATLAB® software for the further nonlinear regression-based analysis (see *Section 2.5*). Data statistics of the simulated operating variables in the present computational analysis are summarized in Table 3.

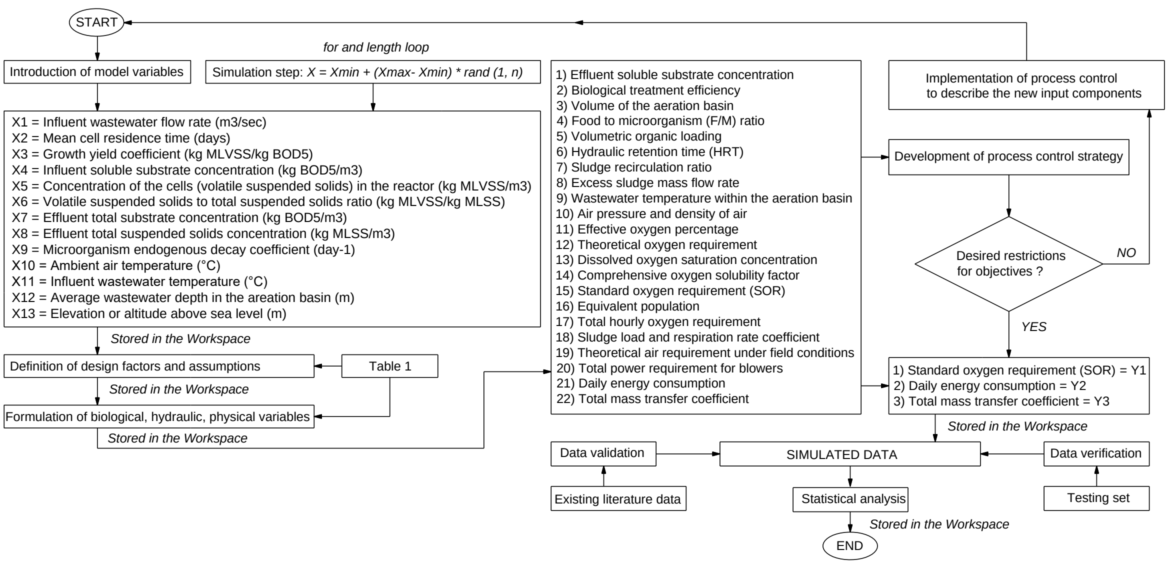


**Figure 2.** Flowchart of the computational prediction process in MATLAB® environment

**Table 3.** Data statistics of the simulated operating variables considered in the present computational analysis ($n$ = 1000 for each variable)

| Component | Units | Minimum | Maximum | Average | SD[a] |
|---|---|---|---|---|---|
| **Input variables ($X_i$)** | | | | | |
| $X_1 = Q$ | [$m^3$/sec] | 0.047 | 0.579 | 0.323 | 0.150 |
| $X_2 = \vartheta_c$ | [days] | 5.03 | 14.97 | 10.27 | 2.87 |
| $X_3 = Y$ | [kg MLVSS/kg $BOD_5$] | 0.401 | 0.799 | 0.606 | 0.115 |
| $X_4 = S_i$ | [kg $BOD_5/m^3$] | 0.20 | 0.40 | 0.30 | 0.059 |
| $X_5 = X$ | [kg MLVSS/$m^3$] | 2.0 | 5.48 | 3.77 | 1.05 |
| $X_6 = \varphi$ | [kg MLVSS/kg MLSS] | 0.600 | 0.899 | 0.744 | 0.091 |
| $X_7 = S_e$ | [kg $BOD_5/m^3$] | 0.020 | 0.045 | 0.033 | 0.007 |
| $X_8 = X_e$ | [kg MLSS/$m^3$] | 0.005 | 0.020 | 0.013 | 0.004 |
| $X_9 = k_d$ | [$day^{-1}$] | 0.040 | 0.075 | 0.057 | 0.010 |
| $X_{10} = T_a$ | [°C] | –5.0 | 30.0 | 12.7 | 10.6 |
| $X_{11} = T_i$ | [°C] | 5.0 | 24.9 | 15.2 | 5.8 |
| $X_{12} = H$ | [m] | 3.0 | 5.0 | 3.98 | 0.59 |
| $X_{13} = H_a$ | [m] | 3 | 2998 | 1511.5 | 857.6 |
| ***Output variables ($Y_i$)*** | | | | | |
| $Y_1 = SOR_d$ | [kg $O_2$/h] | 60.70 | 1676.5 | 608.8 | 319.7 |
| $Y_2 = E_c$ | [kWh/day] | 3441.3 | 149,630 | 41,031 | 25,811 |
| $Y_3 = k_La_d$ | [$day^{-1}$] | 159.73 | 2324.8 | 709.79 | 333.07 |

[a]*Standard deviation.*

*2.5. Nonlinear regression analysis-based modeling*

The simulated data was imported from the MATLAB® workspace used as an open database connectivity data source, and then the nonlinear regression analysis was conducted within the framework of DataFit® software package containing 298 two-dimensional (2D) and 242 three-dimensional (3D) nonlinear regression models. The nonlinear convergence criteria were selected for the following values of the solution preferences: regression tolerance = $1\times10^{-10}$, maximum number of iterations = 1000, and diverging nonlinear iteration limit = 10. When performing the nonlinear regression, the Richardson's extrapolation method was used to calculate numerical derivatives for the solution of the models (see *Section 3.1*). The nonlinear regression analysis was conducted based on the Levenberg–Marquardt method with double precision. Each independent operating variable ($X_i$) was assumed to be equally important ($I_i$ = 1.0), and no particular safety precautions were considered in the construction of the models, as similarly conducted in the previous studies of the author (Yetilmezsoy, 2010; Yetilmezsoy, 2012; Yetilmezsoy, 2016; Yetilmezsoy and Sapci-Zengin, 2009; Yetilmezsoy, 2005; Yetilmezsoy, 2006; Yetilmezsoy, 2007; Yetilmezsoy and Sakar, 2008; Yetilmezsoy, 2012; Yetilmezsoy and Abdul-Wahab, 2014).

In the computational analysis, the stepwise selection procedure (SSP) was performed as the combination of the forward selection and backward elimination procedures for variable selection process within the framework of DataFit® software. The SSP begins with a forward step (with no variables in the model). After the forward step, the $p$ values of the variable coefficients are re-examined and any now insignificant variables are removed in a backward step. This process continues until no variables are either added or removed from the model. The SSP is more generally popular than either the forward or backward procedures.

As the nonlinear regression models were solved on the DataFit® numeric computing environment, they were automatically sorted according to the goodness-of-fit criteria into a graphical interface. Additionally, regression variables (i.e.,$\beta_0$ and $\beta_1, \beta_2, \beta_3, \ldots, \beta_{13}$), standard error of the estimate (SEE), coefficient of multiple determination ($R^2$), adjusted coefficient of multiple determination ($R_a^2$), number of nonlinear iterations (NNI) were computed to evaluate the performance of the regression models. Moreover, $t$-ratios and the corresponding $p$-values were also calculated for the appraisal of the significance of the regression coefficients. An alpha ($\alpha$) level of 0.05 (or 95% confidence) was used to determine the statistical significance of the model components.

*2.6. Verification of simulation results*

Finally, the predictive capabilities of the developed models were tested against the results of a total of 1500 additional and different computational scenarios. For this purpose, the overall testing set (a total of 24,000 different data points composed of an input matrix of [1500×13] and an output matrix of [1500×3]) was randomly allocated into three sub-testing groups (as three [500×16] matrices including output variables) for each model (i.e.,$n_1$ = 500 for $SOR_d$, $n_2$ = 500 for $E_c$ and $n_3$ = 500 for $k_La_d$, where $n_1, n_2, n_3$ are the number of randomly generated testing scenarios) to evaluate the predictive capability of the proposed formulations.

*2.7. Statistical analysis*

In order to describe the overall performance of the proposed equations, quantifying the goodness of the estimate should be implemented as a crucial part of the model development (Yetilmezsoy and Abdul-Wahab, 2014). For this purpose, various descriptive statistical indicators such as coefficient of determination ($R^2$), mean absolute error (MAE), root mean square error (RMSE), systematic and unsystematic RMSE ($RMSE_S$ and $RMSE_U$, respectively), index of agreement (IA), the factor of two (FA2), fractional variance (FV), proportion of systematic error (PSE), coefficient of variation (CV) and Durbin–Watson statistic (DW) were computed as helpful mathematical tools and robust statistics to assess the model's prediction performance and determine the residual error of the estimation.

The obtained output data ($SOR_d$, $E_c$, and $k_La_d$) were statistically evaluated by means of several appropriate parametric (unpaired or two-sample and matched-pair *t* tests) or non-parametric tests (the Mann-Whitney U (or the Wilcoxon rank-sum) test and the Kruskal-Wallis test with the Dwass-Steel-Chritchlow-Fligner method). Prior to applying parametric tests, the Shapiro-Wilk W and the Levene's tests were consecutively implemented as preconditions to ensure that the considered subsets (theoretically calculated values and predicted values) had a normal or non-normal distribution, and variances (or standard deviations) of the paired groups were homogeneous or unequal. When the output values were not normally distributed, non-parametric tests were implemented. Results were assessed with two-tailed *p* values to reflect the statistical significance between paired groups ($\alpha$ = 0.05 or 95% confidence). The parametric or non-parametric tests were conducted by using a licensed statistical software package (StatsDirect, V2.7.2, StatsDirect Ltd., Altrincham, Cheshire, UK).

Based on the other descriptive statistics (i.e., minimum, lower quartile ($Q_1$), median ($Q_2$), upper quartile ($Q_3$), maximum) of independent samples (theoretical data and testing outputs), box-and-whisker plots were also drawn by writing a solution script in the M-file Editor within the framework of MATLAB® software to appraise the statistical results in a pictorial manner. The built-in functions `boxplot([x1,x2],'Param1',val1, 'Param2',val2,...)` (Statistics Toolbox) and `subplot(m,n,p)` (MATLAB® Function Reference) were implemented in MATLAB® for creating these types of display.

## 3. Results

### *3.1. Proposed formulations and estimation of $SOR_d$, $E_c$ and $k_La_d$*

The computational analysis including a total of 8000 different data points (or 500 random operating scenarios in the form of a [500×16] matrix for $X_i$ and $Y_j$, where $i$ = 1,2,..,13 and $j$ = 1,2,3) was carried out for thirteen input and three output variables. The results of the nonlinear regression analysis give the final form of the proposed models as a new function of the selected biological, hydraulic, and physical design variables in Eqs. (20)–(21). These are shown below.

$$\begin{aligned} f_1 &= (2.93)Q + (0.024)\theta_c - (0.077)Y + (3.64)S_i + (0.0097)X - (0.281)\varphi \\ &- (3.28)S_e + (2.58)X_e + (0.227)k_d + (0.0005)T_a + (0.0048)T_i - (0.0162)H \\ &+ (0.00003)H_a + 4.19 \rightarrow SOR_d = \exp(f_1) \end{aligned} \tag{20}$$

$$\begin{aligned} f_2 &= (2.92)Q + (0.024)\theta_c - (0.101)Y + (3.72)S_i + (0.013)X - (0.257)\varphi \\ &- (3.2)S_e + (2.01)X_e + (0.403)k_d + (0.0036)T_a + (0.005)T_i + (0.088)H \\ &+ (0.00045)H_a + 7.23 \rightarrow E_c = \exp(f_2) \end{aligned} \tag{21}$$

$$\begin{aligned} f_3 &= (-0.177)Q - (0.048)\theta_c - (1.88)Y + (0.063)S_i + (0.271)X - (0.246)\varphi \\ &+ (0.153)S_e - X_e + (6.42)k_d + (0.0001)T_a + (0.0038)T_i + (0.253)H \\ &+ (0.00032)H_a + 5.34 \rightarrow k_La_d = \exp(f_3) \end{aligned} \tag{22}$$

where $Q$ is the influent wastewater flow rate ($m^3$/sec), $Y$ is the growth yield coefficient (kg MLVSS/kg $BOD_5$), $S_i$ is the influent soluble substrate concentration (kg $BOD_5/m^3$), $\vartheta_c$ is the mean cell residence time (days), $S_e$ is the total effluent substrate concentration (kg $BOD_5/m^3$), $X_e$ is the effluent total suspended solids concentration (kg MLSS/$m^3$), and $k_d$ is the microorganism endogenous decay coefficient ($day^{-1}$),$X$ is the concentration of the cells in the reactor (kg MLVSS/$m^3$), $\varphi = X/X_{TSS}$ is the volatile suspended solids to total suspended solids ratio (kg MLVSS/kg MLSS), $T_a$ is the ambient air temperature (°C), $T_i$ is the influent wastewater temperature (°C), $H_a$ is the elevation or altitude above sea level (m), and $H$ is the average wastewater depth in the areation basin (m).

The relationship between the proposed models given in Eqs. (20)–(21), and the steady-state theoretical data is outlined in Figure 3. The determination coefficients ($R^2$ = 0.9704 for $SOR_d$, $R^2$ = 0.9795 for $E_c$, and $R^2$ = 0.9824 for $k_La_d$) demonstrated that only 2.95%, 2.05%, and 1.77% of the total variations were unexplained by the proposed $SOR_d$, $E_c$, and $k_La_d$ models respectively. As seen from Figure 3, the predictions of Eqs. (20)–(21) correspond very well with the theoretical values, implying that the proposed equations satisfactorily account for the prediction of the aeration-related parameters in a wide range of the studied variables.

The analysis of variance (ANOVA) showed that the proposed $SOR_d$ model was highly significant, as was evident from the Fisher's *F*-test ($(MS)_{\text{model-1}}/(MS)_{\text{error-1}} = F_{\text{model-1}}$ = 1189.3725) with a very low probability value ($\text{Prob}(F)_{\text{model-1}} \approx 0.0000$). Additionally, the calculated *F* value ($S_r^2/S_e^2 = F_{\text{cal-1}}$ = 1189.3725) was found to be greater than the critical (or tabulated) *F* value ($F_{\alpha,df,n-(df+1)}) = F_{0.05,13,486} = S_r^2/S_e^2 = F_{cr}$ =

1.7403) at the 5% level, indicating that the computed Fisher's variance ratio at this level was large enough to justify a very high degree of adequacy of the $SOR_d$ model. According to the Fisher's $F$-test, the ANOVA indicated that the proposed $E_c$ model was also highly significant ($F_{\text{model-2}}$ = 1765.5427 >>$F_{\text{cr}}$ = 1.7403, Prob($F$)$_{\text{model-2}}$ ≈ 0.0000). Furthermore, the Fisher's $F$-test concluded with 95% certainty that the proposed $k_La_d$ model explained a significant amount of the variation in the dependent variable ($F_{\text{model-2}}$ = 2085.9952 >>$F_{\text{cr}}$ = 1.7403, Prob($F$)$_{\text{model-3}}$ ≈ 0.0000).

In the literature, it has been reported that the $t$-ratio represents the ratio of the estimated parameter effect to the estimated parameter standard deviation. Moreover, the $p$-value is used as a useful tool to check the significance of each of the coefficients. The variable with the larger $t$-ratio and with the smaller $p$-value is considered as the more significant parameter in the regression model (Yetilmezsoy and Sapci-Zengin, 2009; Yetilmezsoy and Sakar, 2008; Yetilmezsoy *et al.,* 2009; Yetilmezsoy and Abdul-Wahab, 2012). It is also noted that values that yield Prob($t$) factors (or $p$-values) of greater than 0.9 may be neglected until all remaining factors are calculated at once (Fingas and Fieldhouse, 2009). In other words, the Prob($t$) or $p$-value is the probability that input can be dropped without affecting the regression or goodness-of-fit (Yetilmezsoy *et al.,* 2011). The regression variable results including standard errors, $t$-statistics (determined by Student's $t$-test) and the corresponding $p$-values for Eqs. (20)–(21) are summarized in Tables 4–6.

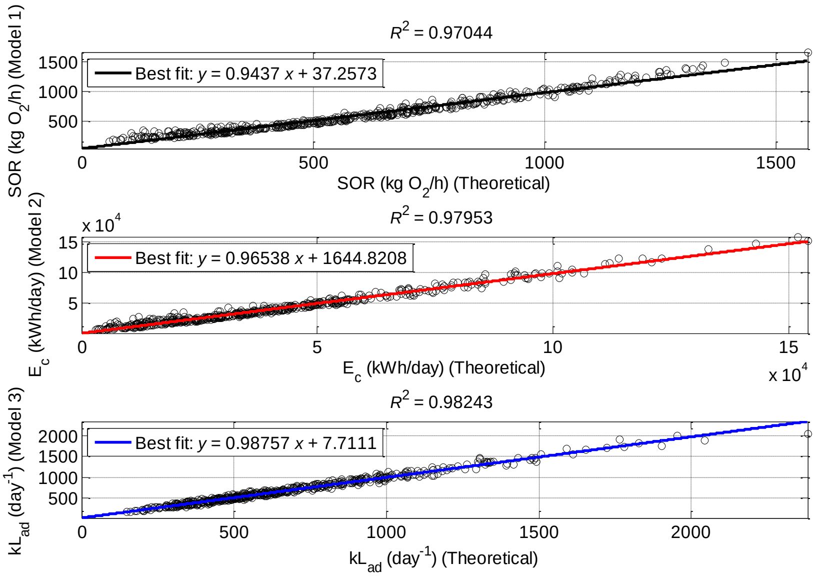


**Figure 3.** Relationships between the proposed models and the steady-state theoretical data

According to absolute $t$-ratios and $p$-values given in Table 4, the influent wastewater flow rate ($X_1 = Q$), the influent soluble substrate concentration ($X_4 = S_i$), and the mean cell residence time ($X_2 = \vartheta_c$) have more importance than other variables for the derived exponential model in prediction of $SOR_d$.

**Table 4.** Nonlinear regression results and significance of model variables in estimation of standard oxygen requirement ($SOR_d$)

$$SOR_d = \exp\left[\beta_1(Q)+\beta_2(\theta_c)+\beta_3(Y)+\beta_4(S_i)+\beta_5(X)+\beta_6(\varphi)+\beta_7(S_e)+\beta_8(X_e)+\beta_9(k_d)+\beta_{10}(T_a)+\beta_{11}(T_i)+\beta_{12}(H)+\beta_{13}(H_a)+\beta_0\right]$$

| Variables[a] | Coefficients | Values | Standard error | $t$-ratio | $p$-value[b] |
|---|---|---|---|---|---|
| $X_1 = Q$ | $\beta_1$ | 2.93 | 0.0311 | 94.0644 | 0.00000 |
| $X_2 = \vartheta_c$ | $\beta_2$ | 0.024 | 0.0013 | 18.4747 | 0.00000 |
| $X_3 = Y$ | $\beta_3$ | –0.077 | 0.0334 | –2.3055 | 0.02156 |
| $X_4 = S_i$ | $\beta_4$ | 3.64 | 0.0671 | 54.2803 | 0.00000 |
| $X_5 = X$ | $\beta_5$ | 0.0097 | 0.0035 | 2.7213 | 0.00674 |
| $X_6 = \varphi$ | $\beta_6$ | –0.281 | 0.0407 | –6.8864 | 0.00000 |
| $X_7 = S_e$ | $\beta_7$ | –3.28 | 0.5373 | –6.0973 | 0.00000 |
| $X_8 = X_e$ | $\beta_8$ | 2.58 | 0.9048 | 2.8509 | 0.00454 |
| $X_9 = k_d$ | $\beta_9$ | 0.227 | 0.3665 | 0.6186 | 0.53646 |
| $X_{10} = T_a$ | $\beta_{10}$ | 0.0005 | 0.000353 | 1.4125 | 0.15845 |
| $X_{11} = T_i$ | $\beta_{11}$ | 0.0048 | 0.000622 | 7.7476 | 0.00000 |
| $X_{12} = H$ | $\beta_{12}$ | –0.0162 | 0.006390 | –2.5349 | 0.01156 |
| $X_{13} = H_a$ | $\beta_{13}$ | 0.00003 | 0.000004 | 6.9558 | 0.00000 |
| Constant | $\beta_0$ | 4.19 | 0.0645 | 64.9898 | 0.00000 |

*[a]Units of variables are previously defined in Eqs. (20)–(21); [b]p values < 0.05 were considered to be significant.*

As seen from Table 5, the influent wastewater flow rate ($X_1 = Q$), the elevation or altitude above sea level ($X_{13} = H_a$), and the influent soluble substrate concentration ($X_4 = S_i$) showed more importance compared to the others for the derived exponential model in prediction of $E_c$.

The concentration of the cells (volatile suspended solids) in the reactor ($X_5 = X$), the elevation or altitude above sea level ($X_{13} = H_a$), and the growth yield coefficient ($X_3 = Y$) were found to be more significant parameters than others for the proposed $k_La_d$ model (Table 6).

Descriptive statistics of the residuals errors in estimation of $SOR_d$, $E_c$, and $k_La_d$ are listed in Table 7.

Scatter plots of $SOR_d$, $E_c$, and $k_La_d$ as a function of each of the predictor variables are illustrated in Figures 4–6.

**Table 5.** Nonlinear regression results and significance of model variables in estimation of daily energy consumption ($E_c$)

$$E_c = \exp\left[\beta_1(Q)+\beta_2(\theta_c)+\beta_3(Y)+\beta_4(S_i)+\beta_5(X)+\beta_6(\varphi)+\beta_7(S_e)+\beta_8(X_e)+\beta_9(k_d)+\beta_{10}(T_a)+\beta_{11}(T_i)+\beta_{12}(H)+\beta_{13}(H_a)+\beta_0\right]$$

| Variables[a] | Coefficients | Values | Standard error | *t*-ratio | *p*-value[b] |
|---|---|---|---|---|---|
| $X_1 = Q$ | $\beta_1$ | 2.92 | 0.0306 | 95.2579 | 0.00000 |
| $X_2 = \vartheta_c$ | $\beta_2$ | 0.024 | 0.0013 | 18.5309 | 0.00000 |
| $X_3 = Y$ | $\beta_3$ | –0.101 | 0.0330 | –3.0440 | 0.00246 |
| $X_4 = S_i$ | $\beta_4$ | 3.72 | 0.0671 | 55.5219 | 0.00000 |
| $X_5 = X$ | $\beta_5$ | 0.013 | 0.0035 | 3.7774 | 0.00018 |
| $X_6 = \varphi$ | $\beta_6$ | –0.257 | 0.0397 | –6.4766 | 0.00000 |
| $X_7 = S_e$ | $\beta_7$ | –3.20 | 0.5196 | –6.1675 | 0.00000 |
| $X_8 = X_e$ | $\beta_8$ | 2.01 | 0.8859 | 2.2662 | 0.02388 |
| $X_9 = k_d$ | $\beta_9$ | 0.403 | 0.3542 | 1.1375 | 0.25589 |
| $X_{10} = T_a$ | $\beta_{10}$ | 0.0036 | 0.000348 | 10.2297 | 0.00000 |
| $X_{11} = T_i$ | $\beta_{11}$ | 0.0050 | 0.000619 | 8.0777 | 0.00000 |
| $X_{12} = H$ | $\beta_{12}$ | 0.088 | 0.006218 | 14.0900 | 0.00000 |
| $X_{13} = H_a$ | $\beta_{13}$ | 0.00045 | 0.000005 | 85.4257 | 0.00000 |
| Constant | $\beta_0$ | 7.23 | 0.0663 | 109.0137 | 0.00000 |

*[a]Units of variables are previously defined in Eqs. (20)–(21); [b]p values < 0.05 were considered to be significant.*

**Table 6.** Nonlinear regression results and significance of model variables in estimation of mass transfer coefficient for diffused aeration ($k_La_d$)

$$k_La_d = \exp\left[\beta_1(Q)+\beta_2(\theta_c)+\beta_3(Y)+\beta_4(S_i)+\beta_5(X)+\beta_6(\varphi)+\beta_7(S_e)+\beta_8(X_e)+\beta_9(k_d)+\beta_{10}(T_a)+\beta_{11}(T_i)+\beta_{12}(H)+\beta_{13}(H_a)+\beta_0\right]$$

| Variables[a] | Coefficients | Values | Standard error | *t*-ratio | *p*-value[b] |
|---|---|---|---|---|---|
| $X_1 = Q$ | $\beta_1$ | –0.177 | 0.0185 | –9.5215 | 0.00000 |
| $X_2 = \vartheta_c$ | $\beta_2$ | –0.048 | 0.0010 | –50.3091 | 0.00000 |
| $X_3 = Y$ | $\beta_3$ | –1.88 | 0.0232 | –81.0003 | 0.00000 |
| $X_4 = S_i$ | $\beta_4$ | 0.063 | 0.0486 | 1.2867 | 0.19882 |
| $X_5 = X$ | $\beta_5$ | 0.271 | 0.0028 | 95.1186 | 0.00000 |
| $X_6 = \varphi$ | $\beta_6$ | –0.246 | 0.0305 | –8.0714 | 0.00000 |
| $X_7 = S_e$ | $\beta_7$ | 0.153 | 0.3805 | 0.4034 | 0.68683 |
| $X_8 = X_e$ | $\beta_8$ | –1.00 | 0.6400 | –1.5609 | 0.11919 |
| $X_9 = k_d$ | $\beta_9$ | 6.42 | 0.2753 | 23.3185 | 0.00000 |
| $X_{10} = T_a$ | $\beta_{10}$ | 0.0001 | 0.000261 | 0.1918 | 0.84800 |
| $X_{11} = T_i$ | $\beta_{11}$ | 0.0038 | 0.000475 | 8.0784 | 0.00000 |
| $X_{12} = H$ | $\beta_{12}$ | 0.253 | 0.004696 | 53.8576 | 0.00000 |
| $X_{13} = H_a$ | $\beta_{13}$ | 0.00032 | 0.000003 | 94.5400 | 0.00000 |
| Constant | $\beta_0$ | 5.34 | 0.0498 | 107.1153 | 0.00000 |

*[a]Units of variables are previously defined in Eqs. (20)–(21); [b]p values < 0.05 were considered to be significant.*

**Table 7.** Descriptive statistics of the residuals errors for the derived nonlinear regression models

| **Residual statistics** | **Calculation**[a] | **Regression results**[b] |
|---|---|---|
| Sum of residuals | $SR=\sum_{i=1}^{n}(Y_a-Y_p)$ | –2053.647 |
| | | –119,806.225 |
| | | 429.199 |
| Average residual | $AR=\frac{1}{n}\sum_{i=1}^{n}(Y_a-Y_p)$ | –4.107 |
| | | –239.612 |
| | | 0.858 |
| Standard error of the estimate | $SEE=\sqrt{\frac{\sum_{i=1}^{n}(Y_a-Y_p)^2}{n-p}}=\sqrt{\frac{SSE}{n-p}}$ | 54.177 |
| | | 3877.351 |
| | | 45.516 |
| Adjusted coefficient of multiple determination | $R_a^2=\left[(n-1)R^2-k\right]/(n-1-k)$ | 0.969 |
| | | 0.979 |
| | | 0.982 |
| Durbin–Watson (DW) statistic ($e$ = $Y_a - Y_p$) | $DW=\sum_{i=2}^{n}(e_i-e_{i-1})^2 \Big/ \sum_{i=1}^{n}e_i^2$ | 1.969 |
| | | 1.892 |
| | | 1.943 |

*[a]$Y_a$ is the actual data point, $Y_p$ is thepredicted values, n is the number of data points or observations, p is the number of parameters or variables in the regression model, $R^2$ is the determination coefficient, and k is the number of regression parameters in the model; [b]Results have been given in vertical order for $SOR_d$, $E_c$, $k_La_d$ models, respectively.*

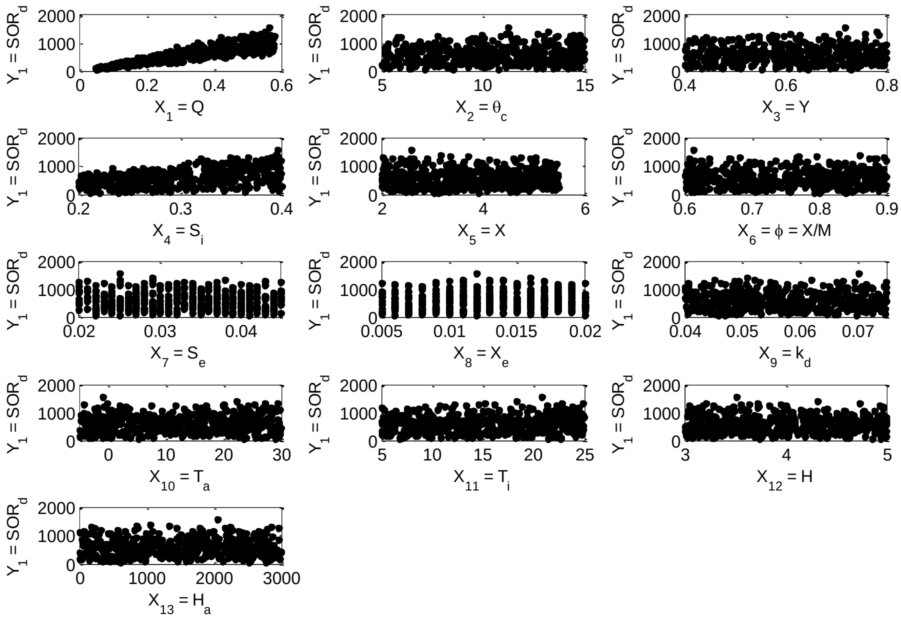


**Figure 4.** Scatter plots of the standard oxygen requirement ($Y_1 = SOR_d$) as a function of each of the predictor variables ($X_i$ for $i$ = 1,2,..,13)

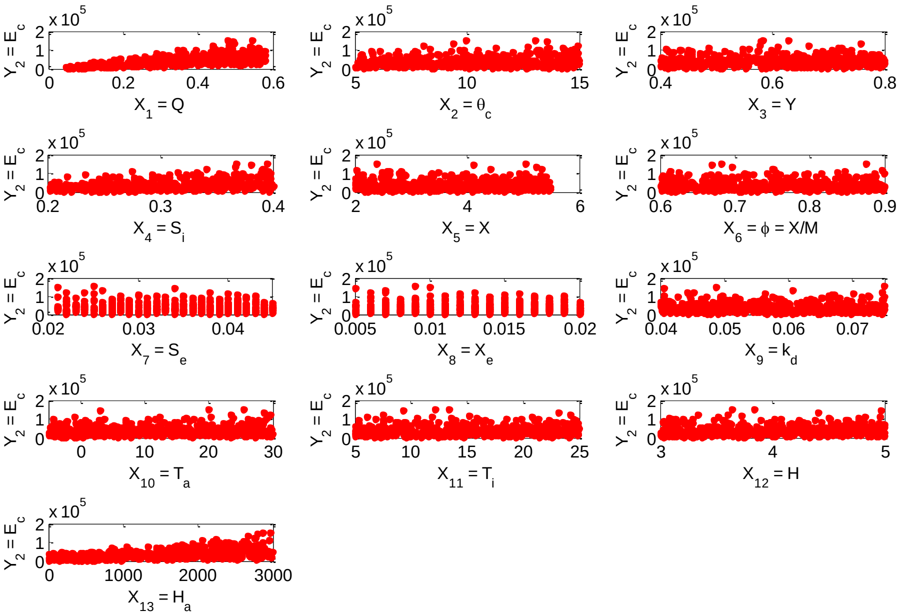

**Figure 5.** Scatter plots of the daily energy consumption ($Y_2 = E_c$) as a function of each of the predictor variables ($X_i$ for $i$ = 1,2,..,13)

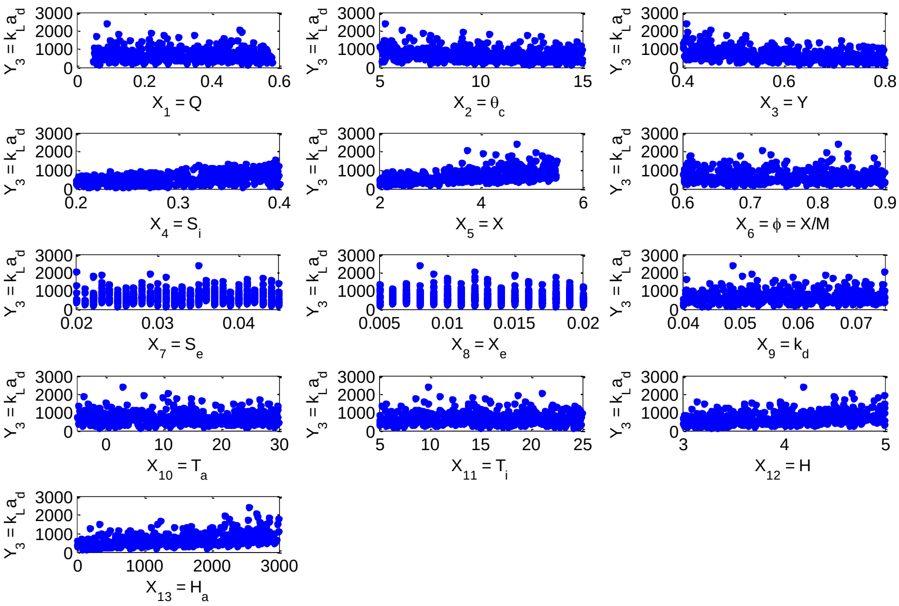

**Figure 6.** Scatter plots of the total mass transfer coefficient ($Y_3 = k_La_d$) as a function of each of the predictor variables ($X_i$ for $i$ = 1,2,..,13)

Based on the above-noted facts, the resulting regression models (Tables 4–6) showed that all parameters contributed to the final result, and that none of them could be cut without affecting the outcome of the models. Furthermore, considering the density of clusters in specific ranges (Figures 4–6), all variables showed a certain importance, indicating that they should not be eliminated from the models.

*3.2. Measuring of the goodness of the estimate*

After a considerable consistency was obtained between the computed results and the theoretical data, varying inputs were randomly applied to observe the prediction stability of the proposed model under various operating conditions. For this purpose, a total of 24,000 new data points (or 1500 new random operating scenarios) were introduced to the MATLAB® algorithm (see *Section 2.2*) and used as the testing data. The whole testing set was arbitrarily divided into three sub-testing sets (as three [500×16] matrices) labeled testing sets 1, 2 and 3 for each model. This was conducted to evaluate the predictive performance of the proposed model on various input scenarios.

In order to verify the predictive capability of the proposed models ($SOR_d$, $E_c$, and $k_La_d$), the goodness of the estimates was appraised by calculating various statistical indicators for each sub-testing set. The obtained results for descriptive performance indicators are presented in Table 8. Applying a linear regression analysis between each of testing outputs and the corresponding theoretical results indicated that Eqs. (20)–(21) can be reliably utilized to predict the aeration-related parameters such as standard oxygen requirement, daily energy consumption and total mass transfer coefficient for the diffused aeration. As seen in Table 8, the values of the determination coefficients ($R^2$) were computed between $R^2$ = 0.9614–0.9646, $R^2$ = 0.9672–0.9717, and $R^2$ = 0.9803–0.9844 for $SOR_d$, $E_c$ and $k_La_d$ models, respectively, The $R^2$ values demonstrated that unexplained variations were calculated only between 3.54–3.86%, 2.83–3.28%,1.56–1.97% of all the variations in prediction of the $SOR_d$, $E_c$, and $k_La_d$, respectively, revealing that the proposed formulations satisfactorily estimated the expected targets with very small deviations.

**Table 8.** Descriptive performance indicators used for testing the goodness of the estimate for the proposed formulations (given in vertical order for each testing set)

| Performance indice | Calculation[a] | Testing sets 1, 2 and 3 | | |
|---|---|---|---|---|
| | | $SOR_d$ | $E_c$ | $k_La_d$ |
| | | 1,2,3 | 1,2,3 | 1,2,3 |
| Determination coefficient ($R^2$) | $R^2=\dfrac{\left(\sum_{i=1}^{n}(O_i-O_m)(P_i-P_m)\right)^2}{\sum_{i=1}^{n}(O_i-O_m)^2\sum_{i=1}^{n}(P_i-P_m)^2}$ | 0.9614<br>0.9646<br>0.9620 | 0.9672<br>0.9711<br>0.9717 | 0.9844<br>0.9803<br>0.9818 |
| Mean absolute error (MAE) | $MAE=\frac{1}{n}\sum_{i=1}^{n}\lvert P_i-O_i\rvert$ | 53.3285<br>50.0345<br>50.2999 | 3800.8<br>3334.0<br>3500.8 | 35.0215<br>34.5944<br>35.3596 |
| Root mean squared error (RMSE) | $RMSE=\left(\frac{1}{n}\sum_{i=1}^{n}[P_i-O_i]^2\right)^{0.5}$ | 66.1589<br>64.2315<br>63.2041 | 5047.3<br>4535.6<br>4715.1 | 48.8225<br>47.0747<br>49.3155 |
| Systematicroot mean squared error (RMSEs) | $RMSE_S=\left(\frac{1}{n}\sum_{i=1}^{n}\left[(P_i)_{reg}-O_i\right]^2\right)^{0.5}$ | 27.3782<br>29.6355<br>28.2099 | 1409.5<br>1449.0<br>1666.5 | 11.9493<br>9.4401<br>11.1685 |
| Unsystematicroot mean squared error (RMSEu) | $RMSE_U=\left(\frac{1}{n}\sum_{i=1}^{n}\left[(P_i)_{reg}-P_i\right]^2\right)^{0.5}$ | 60.2281<br>56.9862<br>56.5594 | 4846.5<br>4297.9<br>4410.7 | 47.3376<br>46.1185<br>48.0342 |
| Proportion of systematic error (PSE) | $PSE=(RMSE_S)^2/(RMSE_U)^2$ | 0.2066<br>0.2704<br>0.2488 | 0.0846<br>0.1137<br>0.1428 | 0.0637<br>0.0419<br>0.0541 |
| Index of agreement (IA) | $IA=1-\left(\dfrac{\sum_{i=1}^{n}(P_i-O_i)^2}{\sum_{i=1}^{n}(\lvert P_i-O_m\rvert+\lvert O_i-O_m\rvert)^2}\right)$ | 0.9890<br>0.9895<br>0.9889 | 0.9913<br>0.9922<br>0.9922 | 0.9959<br>0.9949<br>0.9952 |
| Fractional variance (FV) | $FV=2(\sigma_o-\sigma_p)/(\sigma_o+\sigma_p)$ | 0.0641<br>0.0744<br>0.0738 | 0.0294<br>0.0352<br>0.0453 | 0.0163<br>0.0044<br>0.0085 |
| Factor of two (FA2) | $0.5\le FA2=(1/n)\sum_{i=1}^{n}(O_i/P_i)\le 2.0$ | 0.9458<br>0.9455<br>0.9591 | 0.9438<br>0.9442<br>0.9575 | 1.0121<br>1.0150<br>1.0149 |
| Coefficient of variation (CV, %) | $CV=(RMSE/O_m)\times 100$ | 11.6070<br>11.1222<br>10.8790 | 12.6107<br>11.8569<br>11.6362 | 6.4877<br>6.5136<br>6.6331 |
| Durbin–Watson (DW) statistic ($e_i = O_i - P_i$) | $DW=\sum_{i=2}^{n}(e_i-e_{i-1})^2\Big/\sum_{i=1}^{n}e_i^2$ | 2.1218<br>1.7793<br>1.8741 | 2.1700<br>1.7806<br>1.8224 | 1.9839<br>1.9426<br>2.1060 |

[a]*O, P, σ, and the subscripts m and reg indicate the observed, predicted, standard deviation, mean and regression, respectively.*

The determined IA (0.9889–0.9895, 0.9913–0.9922, and 0.9949–0.9959 for $SOR_d$, $E_c$, and $k_La_d$, respectively) and FA2 (0.9455–0.9591, 0.9438–0.9575, and 1.0121–1.0150 for $SOR_d$, $E_c$, and $k_La_d$, respectively) values were computed to be very close to 1, indicating that very good agreements were achieved between the theoretical values and the outputs of the models. The low values of the coefficient of variation (CV = 10.88–11.12%, CV = 11.64–12.61%, and CV = 6.49–6.63% for the testing data sets of $SOR_d$, $E_c$, and $k_La_d$, respectively) demonstrated a high degree of precision and a good deal of the reliability of the proposed equation, as similarly reported in previous works (Yetilmezsoy, 2016; Yetilmezsoy *et al.,* 2009). Other descriptive performance indices such as PSE and FV also revealed that the proposed empirical models gave very small residuals and demonstrated a noticeable estimation performance on forecasting of the studied the aeration-related parameters. Furthermore, the DW statistics ($DW_{model\text{-}1}$ = 1.969, $DW_{model\text{-}}$

$_2$ = 1.892, and DW$_{model-3}$ = 1.943) were determined to be very close to 2, indicating the goodness of fit of the derived models (Yetilmezsoy, 2016; Yetilmezsoy *et al.,* 2009; Yetilmezsoy and Abdul-Wahab, 2012; Hewings *et al.,* 2002).

### *3.3. Non-parametric tests and box-and-whisker plots*

For the proposed $SOR_d$ model (first model), both the Mann–Whitney U test and the Kruskal-Wallis test (with the Dwass-Steel-Chritchlow-Fligner method) showed that there was no statistically significant difference between the outputs of the testing set 1 (a matrix of [500×2]) and the corresponding theoretical data ($p_{MW} = p_{KW} = 0.648$). In this case, because the $p$ value was higher than the chosen $\alpha$ level of 0.05 (or 95% confidence), the null hypothesis ($H_0$) was not rejected in favor of the alternative hypothesis ($H_a$). Similarly, for a total of 1000 observations ($n_1$ = 500 for the first model + $n_2$ = 500 for testing set 1), the same non-parametric tests concluded that there was insufficient evidence for a significant difference between the predictions of the testing set 2 and the theoretical responses ($p_{MW} = p_{KW} = 0.721$). Moreover, the non-parametric tests revealed that no sufficient evidence was found for a significant difference between the estimated data (testing set 3) and the respective theoretical values ($p_{MW} = p_{KW} = 0.712$).

For the proposed $E_c$ model (second model), the non-parametric tests indicated that no sufficient evidence was generated for a significant difference between $E_c$ values obtained from both the outputs of the testing set 1 and the corresponding theoretical data ($p_{MW} = p_{KW} = 0.446$). Once again, $H_a$ was rejected in favor of $H_0$, since no statistically significant difference was recorded between the outputs of the testing set 2 and the corresponding theoretical data ($p_{MW} = p_{KW} = 0.491$). The non-parametric tests indicated that no sufficient evidence was produced to prove that there was a significant difference between $E_c$ values obtained from the testing set 3 and the respective theoretical values ($p_{MW} = p_{KW} = 0.462$).

For the proposed $k_La_d$ model (third model), both the Mann–Whitney U test and the Kruskal-Wallis test demonstrated that no sufficient evidence was offered to support a significant difference between the forecasted $k_La_d$ values (testing set 1) and the respective theoretical data set ($p_{MW} = p_{KW} = 0.772$). Likewise, the same non-parametric tests concluded that there was insufficient evidence for a significant difference between the predictions of the testing set 2 and the theoretical responses ($p_{MW} = p_{KW} = 0.681$). Furthermore, $H_0$ was not rejected in favor of the $H_a$, since no statistically significant difference was found between the outputs of the testing set 3 and the corresponding theoretical data ($p_{MW} = p_{KW} = 0.651$).

The above statistical findings obtained from the non-parametric analysis confirmed with 95% certainty that the proposed models satisfactorily described the behavior of the present aeration-related parameters (standard oxygen requirement, daily energy consumption and total mass transfer coefficient for diffused aeration) even in a widely varying input regime.

Finally, all local differences between the testing outputs and the theoretical data sets have been described graphically by means of the box-and-whisker-plots, which are shown in Figure 7. These diagrams summarize each variable by four components as follows: (1) a central line in each box is the sample median to indicate central tendency or location, (2) a box to indicate variability around this central tendency (the edges of the box are the 25th and 75th percentiles), (3) whiskers around the box to indicate the range of the variable, and (4) observations beyond the whisker length are marked as outliers displayed with a plus (+) sign where its value is more than 1.5 times the interquartile range (IQR = $Q_3 - Q_1$) away from the top or bottom of the box (Singh *et al.,* 2010; Nasr *et al.,* 2012; Sharma *et al.,* 2014).

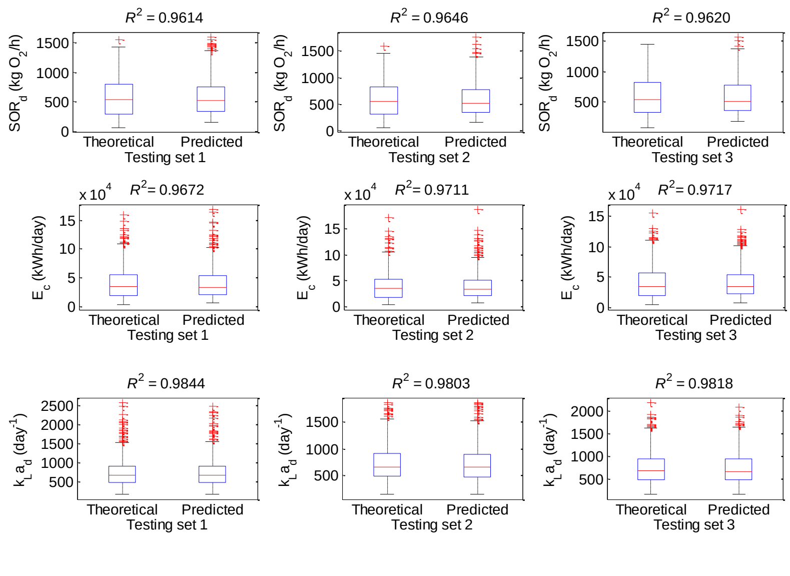


**Figure 7.** Box-and-whisker plots of the theoretical data sets and the outputs of the proposed models for testing sets 1, 2 and 3 (top line: the maximum level, the bottom line: the minimum level, top line of the box: 75th percentile (or upper quartile), the bottom line of the box: 25th percentile (or lower quartile), middle line of the box: 50th percentile or the median)

As seen in Figure 7, the box-and-whisker plots indicate that the proposed models ($SOR_d$, $E_c$ and $k_La_d$) produce quantatively similar results compared to the theoretical data sets. According to the minimum, lower quartile ($Q_1$), median ($Q_2$), upper quartile ($Q_3$) and the maximum categories, the box-and-whisker plots suggest that the distributions of independent samples (theoretical data sets and testing outputs) are close enough to be comparable for statistical purposes. The box-and-whisker plots readily convey that the shape of the predicted data sets (box plots on the right for each testing set) come from an exponential distribution (Martinez and Martinez, 2001; Martinez *et al.*, 2004). The outliers on the boxes represent extreme points that arise randomly according to the range of simulations conducted in the computational analysis (Yetilmezsoy, 2016).

On the basis the overall results, the general applicability and universality of the proposed models were further investigated by a validation study presented in the next section.

### *3.4. Validation of models*

Finally, a validation study was implemented to assess the consistency of simulation results with the results calculated based on the existing literature data (Kumar, 2011; Muslu, 1996b; Toprak, 2000; von Sperling, 2007; Ong, 2005; NPTEL, 2015). The results are summarized in Table 9. As similarly conducted for the sub-testing sets, (see *Section 3.3*), both the Mann–Whitney U test and the Kruskal-Wallis test (with the Dwass-Steel-Chritchlow-Fligner method) demonstrated that there was no statistically significant difference between the outputs of the proposed models and the corresponding validation data set given in Table 9 ($p_{MW}$ = 0.8182 and $p_{KW}$ = 0.7488 for $SOR_d$ model, $p_{MW}$ = 0.9372 and $p_{KW}$ = 0.8728 for $E_c$ model, and $p_{MW}$ = 0.3939 and $p_{KW}$ = 0.3367 for $k_La_d$ model). Therefore, the null hypothesis ($H_0$) was not rejected in favor of the alternative hypothesis ($H_a$), since the $p$ values were higher than the chosen α level of 0.05 (or 95% confidence) for all cases. Furthermore, the determination coefficients ($R^2$ = 0.9734 for $SOR_d$ model, $R^2$ = 0.9847 for $E_c$ model, and $R^2$ = 0.9652 for $k_La_d$ model) indicated that only 2.66%, 1.53%, and 3.48% of the total variations were unexplained by the proposed $SOR_d$, $E_c$ and $k_La_d$ models, respectively.

**Table 9.** Comparison of the model outputs with the results calculated based on the existing literature data

| Input variables | | | | | | | | | | | | | Reference | Output variables[a] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | $X_{11}$ | $X_{12}$ | $X_{13}$ | | Calculated results | Proposed models |
| $Q$ | $\vartheta_c$ | $Y$ | $S_i$ | $X$ | $\varphi$ | $S_e$ | $X_e$ | $k_d$ | $T_a$ | $T_i$ | $H$ | $H_a$ | | | |
| 0.451 | 10 | 0.5 | 0.2 | 3 | 0.8 | 0.02 | 0.02 | 0.06 | 30 | 24 | 4.5 | 450 | Muslu (1996a) | 553.1954<br>26,514<br>508.1834 | 553.1615<br>25,354<br>499.6626 |
| 0.282 | 10 | 0.5 | 0.175 | 3.5 | 0.8 | 0.045 | 0.022 | 0.06 | 25 | 20 | 3.85 | 75.78 | Toprak (2000) | 261.2219<br>9702.1<br>466.0227 | 280.0851<br>10,109<br>436.8962 |
| 0.250 | 10 | 0.5 | 0.250 | 3.5 | 0.8 | 0.02 | 0.022 | 0.06 | 20 | 20 | 6 | 287 | Ong (2005) | 371.9594<br>19,264<br>742.7382 | 352.6020<br>17,205<br>810.2153 |
| 0.222[b] | 6 | 0.6 | 0.239 | 3 | 0.77 | 0.02 | 0.03 | 0.08 | 20 | 20 | 4 | 800 | von Sperling (2007) | 323.3133<br>16,107<br>600.8974 | 303.7211<br>14,829<br>575.9397 |
| 0.266 | 5 | 0.6 | 0.224 | 2.56 | 0.8 | 0.02 | 0.03 | 0.06 | 30 | 30 | 5 | 6 | Kumar (2011) | 319.0877<br>13,776<br>503.1022 | 318.0384<br>12,686<br>482.1089 |
| 0.139[c] | 5 | 0.6 | 0.252 | 4 | 0.8 | 0.023 | 0.02 | 0.07 | 18 | 18 | 3.1 | 6 | NPTEL (2015) | 195.6992<br>6067.4<br>526.5417 | 230.3869<br>7363.1<br>463.9663 |
| Units of input variables | | | | | | | | | | | | | | Units of output variables | |
| $Q$ [=] m³/sec, $\vartheta_c$ [=] days, $Y$ [=] kg MLVSS/kg $BOD_5$, $S_i$ [=] kg $BOD_5$/m³, $X$ [=] kg MLVSS/m³, $\varphi$ [=] [kg MLVSS/kg MLSS], $S_e$ [=] kg $BOD_5$/m³, $X_e$ [=] kg MLSS/m³, $k_d$ [=] day$^{-1}$, $T_a$ [=] °C, $T_i$ [=] °C, $H$ [=] m, $H_a$ [=] m | | | | | | | | | | | | | | $SOR_d$ [=] kg $O_2$/h, $E_c$ [=] kWh/day, $k_La_d$ [=] day$^{-1}$ | |

*[a]Results have been given in vertical order for $SOR_d$, $E_c$, $k_La_d$ models, respectively; [b]Maximum influent flow rate is considered for the population equivalent of 67,000 inhabitants; [c]Q value is computed for q = 200 l/capita/day (see Eq. (11) in Section 2.2, and Table 2).*

Consequently, the validation results corroborated that the proposed models satisfactorily accounted for the behavior of the present aeration-related parameters. The statistical results clearly support the general applicability of the proposed formulations for any wastewater treatment plant within the proposed limits of the relevant input variables, to various operating conditions, as well as their predictive capability and accuracy for practical design purposes.

## 4. Discussion

Activated sludge-based models have been widely studied by design engineers in a variety of applications, and are likely to continue to beused in the future. Nevertheless, many of the proposed expressions are highly theoretical, and not always practicable, since the most of variables used in the mathematical structure of these models are not readily available or routinely obtainable for all activated sludge-based treatment plants. On the contrary, although the simplicity ofsome of other previous expressionshas been articulated by a number of authors, however, they neglect the effect of several important process-related variables, and do not directly reflect the actual behavior the large-scale aeration units.To overcome the limitations and problems associated with the existing deterministic models, some authors have also focused on solving the problems of activated sludge process by using stochastic methods or specific computational programs. Despite the extensive use of personal computers, practising engineers frequently need to perform quick, but approximate calculations in a simple manner. At this point, “back-of-the-envelope” type solutions providing simplified assumptions are absolutely helpful in such conditions.Although iterative solution of the relevant equations using a computer program does not seem to behard,sufficiently effortless and accurate explicit expressionsare always given priority over a software-dependent solution. For this reason, considering the practical needs, economic reasons and time constraints in engineering, the present study attempts to address this gap by implementingimplementing entirely new explicit equations as functions of thirteen fundamental biological, hydraulic, and physical variables, which are the most widely used design parameters in activated sludge-based treatment plants.

The proposed explicit equations satisfactorily accounted for the accurate quantitative estimation of aeration-related parameters ($SOR_d$, $E_c$, and $k_La_d$) in the steady-state completely mixed activated sludge process. The only requirements are limited to input variables, which are readily incorporated into the routine analyses performed in almost all activated sludge-based treatment plants. The empirical equations that are described in this study have the ability to provide realistic results, and it is therefore believed that these modelscan be used as alternative mathematical formulations to the real-world activated sludge-based problems.

Engineers, designers and researches may not have enough time to compute all time-consuming and task-intensive calculations in environmental engineering practice. Hence, a number of attempts in developing representative equations will provide a new scientific contribution in modeling of aeration-related parameters in the completely mixed activated sludge process. In this regards, the formulations derived in the scope of this study have obviated the need fora number of consecutive and laborious design calculations performed in the conventionalsolution procedure.

The regression coefficients of the model variables have been statistically rounded in an acceptable way (without changing the original determination coefficients), and simplified to be used for practical computations with a hand-held calculator. Therefore, compared with the conventional calculation procedure, the equations proposed herein are considerably simpler and more practical in form. The simple character mathematical formulations eliminated the various variable interactions and many unit conversions carried out in the theoretical approach.

Consequently, from the engineering point of view, it is believed that the proposed equations can be utilized as a practical tool within comparatively shorter computation time to evaluate the values of standard oxygen requirement, daily energy consumption and total mass transfer coefficient for the diffused aerationin the steady-state completely mixed activated sludge process.

## 5. Conclusions

The problem of predicting the aeration-related parameters, such as standard oxygen requirement, daily energy consumption and total mass transfer coefficient for the diffused aeration, has been studied. Complex biological and hydraulic interactions in the completely mixed activated sludge process have been attempted to be described by three simple mechanistic models. The proposed models offer sufficiently practical mathematical formulations incorporating the most common biological, hydraulic, and physical design parameters, which are readily available and routinely obtainable for almost all activated sludge-based treatment plants.

The results have been validated by comparison of calculated and estimated data in terms of the present aeration-related parameters. The statistical results corroborate that the proposed equations can produce theoretically meaningful outputs and represented the aeration data very accurately. It is also noteworthy that the new equations may remedy shortcomings of several existing cumbersome correlations, since they have a number of motivations such as simple and explicit forms, wider operating ranges of both the input and output variables, high accuracy, relatively small computation time, and a fundamental basis.

In accordance with the present results, it is believed that the developed formulations can facilitate the computation of the present aeration-related parameters and will be of interest to practising engineers and researchers who are concerned with the design of the activated sludge process. The proposed equations can be safely used for any wastewater treatment plant which has activated sludge system as biological process within the proposed limits of the relevant input data.

**Acknowledgements**

The author would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the manuscript.

**Declaration of conflicting interests**

There is no conflict of interest declared by the author.

## References

Al-Ahmady K.K. (2011), Mathematical model for calculating oxygen mass transfer coefficient in diffused air systems, *Al-Rafadain EngineeringJournal,* **19**(4), 43–54.

ASCE (American Society of Civil Engineers) (1997), Standard Guidelines for In-Process Oxygen Transfer Testing, ASCE 18-96, New York, NY.

Bagheri M., Mirbagheri S.A., Bagheri Z. and Kamarkhani A.M. (2015), Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach, *Process Safety and Environmental Protection,* **95,** 12–25.

Bolles S.A. (2006), Modeling wastewater aeration systems to discover energy savings opportunities, Process Energy Services LLC, pp.11, http://www.processenergy.com/Aeration%20Paper.pdf. Accessed in November 2014.

Bugbee B. and Blonquist M. (2006), Absolute and relative gas concentration: Understanding Oxygen in Air, pp.9, http://www.apogeeinstruments.com/content/o2s_correcting.pdf. Accessed in July 2015.

Casey T.J. (2009), Diffused air aeration systems for the activated sludge process, Design performance testing, Aquavarra Research RD Publications, Water Engineering Papers, Paper 4, County Dublin, Ireland, 1– 4.

Celenza G.J. (1999), Industrial waste treatment process engineering: Biological Processes, Volume 2, CRC Press, pp.207.

Chen C.Y., Roth J.A. and Eckenfelder W.W. (1980), Response of dissolved oxygen to changes in influent organic loading to activated sludge systems, *Water Research,* **14**(10), 1449–1457.

Comeau Y. (2008), Microbial Metabolism, In: M. Henze, M.C.M. van Loosdrecht, G.A. Ekama, D. Brdjanovic (eds.), Biological Wastewater Treatment, Principles, Modelling and Design, Chapter 2, ISBN: 9781843391883, IWA Publishing, London, UK, p. 9–32.

Crites R. and Tchobanoglous G. (1998), Small and decentralized wastewater management systems, McGraw-Hill, USA.

Eroglu V. (1991), Oxygen transfer and aeration in wastewater treatment, *Su Kirlenmesi Kontrolü Dergisi (SKKD),* **1**(2), 95–109 (in Turkish).

Fayolle Y., Cockx A., Gillot S., Roustan M. and Héduit A. (2007), Oxygen transfer prediction in aeration tanks using CFD, *Chemical Engineering Science,* **62**(24), 7163–7171.

Fingas M. and Fieldhouse B. (2009), Studies on crude oil and petroleum product emulsions: Water resolution and rheology, *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* **333,** 67–81.

Fouad M. and Bhargava R. (2005a), A simplified model for the steady-state biofilm-activated sludge reactor, *Journal of Environmental Management,* **74**(3), 245–253.

Fouad M. and Bhargava R. (2005b), Mathematical model for the biofilm–activated sludge reactor, *Journal of Environmental Engineering,* **131**(4), 557–562.

Ghangrekar M.M. and Kharagpur I.I.T. (2014), Industrial water pollution control, Aerobic secondary treatment of wastewater, Activated sludge process, pp.26., http://ghangrekar.com/wp-content/uploads/2014/04/19-Aerobic-treatment-process.pdf. Accessed in October 2014.

Giokas D.L., Kim Y., Paraskevas P.A., Paleologos E.K. and Lekkas, T.D. (2002), A simple empirical model for activated sludge thickening in secondary clarifiers, *Water Research,* **36**(13), 3245–3252.

Goto M. and Andrews J.F. (1985), On-line estimation of oxygen uptake rate in the activated sludge process, In: Instrumentationand Control of Water and Wastewater Treatment and Transport Systems, Proc. 4th IAWPRC Workshop, Houston Denver, 465–472.

Henderson M.A. (2002), Energy reduction methods in the aeration process at Perth Wastewater Treatment Plant, Strathclyde Uo, Glasgow, Scotland, pp.88., http://www.esru.strath.ac.uk/Documents/MSc_2002/henderson.pdf. Accessed in July 2015.

Henkel J. (2010), Oxygen transfer phenomena in activated sludge, Dissertation, Civil and Environmental Engineering, ISBN 978-3-940897-06-0, TU Darmstadt, Darmstadt, pp.194.

Hewings G.J.D., Changnon S. and Dridi C. (2002), Testing for the significance of extreme weather and climate event on the state economies, In: The Regional Economics Application Laboratory, REAL 00–T–6, S. Mathews, Urbana, IL, September, 2002, pp. 1–16.

Holmberg U. (1986), Adaptive dissolved oxygen control and on-line estimation of oxygen transfer and respiration rates, In: Proc. AIChE's Annual Meeting, Miami Beach.

Isazadeh S., Feng M., Urbina Rivas L.E. and Frigon D. (2014), New mechanistically based model for predicting reduction of biosolids waste by ozonation of return activated sludge, *Journal of Hazardous Materials,* **270**(2014), 160–168.

Kanat G. and Saral A. (2009), Estimation of biogas production rate in a thermophilic UASB reactor using artificial neural networks, *Environmental Modeling & Assessment,* **14,** 607–614.

Kumar K.S. (2011), Computer aided design of waste water treatment plant with activated sludge process, *International Journal of Engineering Science and Technology,* **3,** 3348–3356.

Liu G. and Wang J. (2015), Modeling effects of DO and SRT on activated sludge decay and production, *Water Research*, **80,** 169–178.

Loehr R.C. (2012), Pollution control for agriculture: Problems, processes, and applications, Second Edition, Academic Press, pp. 467.

Ma J., Liu J. and Liang S. (2013), Study on the changing law of dissolved oxygen and dissolved oxygen saturation in Baiyang Lake, *Nature Environment and Pollution Technology*, **12**(3), 407–412.

Makinia J. (2010), Mathematical modelling and computer simulation of activated sludge systems, IWA Publishing, London, UK, pp. 319.

Makinia J. and Wells S.A. (1999), Improvements in modeling dissolved oxygen in activated sludge systems, Civil and Environmental Engineering Faculty Publications and Presentations, Proceedings of the 8th International Association on Water Quality Conference on Design, Operation, and Economics of Large Wastewater Treatment Plants Paper 121, pp. 9.

Makinia J. and Wells S.A. (2000), A general model of the activated sludge reactor with dispersive flow – I. Model development and parameter estimation, *Water Research,* **34**(16), 3987–3996.

Makinia J., Wells S.A. and Zima P. (2005), Temperature modeling in activated sludge systems: A case study, *Water Environment Research,* **77**(5), 525-532.

Martinez W.L., Martinez A. and Solka J.(2004), Exploratory data analysis with MATLAB®, CRC Press, pp. 424.

Martinez W.L. and Martinez A.R. (2001), Computational statistics handbook with MATLAB®, CRC Press, pp. 616.

Mulas M. (2006), Modelling and control of activated sludge processes, Dottorato di Ricerca in Ingegneria Industriale, Università degli Studi di Cagliari, XVIII Ciclo.

Muslu Y. (1996a), Wastewater Treatment, Vol. 1. ITU Publishing Corp., Istanbul, pp. 429 (in Turkish).

Muslu Y. (1996b), Wastewater Treatment, Vol. 2. ITU Publishing Corp., Istanbul, pp. 447 (in Turkish).

Nasr M.S., Moustafa M.A., Seif H.A. and El Kobrosy G. (2012), Application of artificial neural network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT, *Alexandria Engineering Journal,* **51**(1), 37–43.

NPTEL (National Programme on Technology Enhanced Learning) (2015), E-Learning Courses from the IITs IISc, Civil

Engineering, Water Wastewater Engineering, Design of Completely Mixed Activated Sludge System, http://nptel.ac.in/courses/105104102/Lecture%2026.htm. Accessed in August 2015.

Nuhoglu A., Keskinler B. and Yildiz E. (2005), Mathematical modelling of the activated sludge process—the Erzincan case, *Process Biochemistry,* **40**(7), 2467–2473.

Ong S.K. (2005), CE 428 Water and Wastewater Treatment Design, Spring 2005, Department of Civil, Construction and Environmental Engineering, Iowa State University, Ames, IA, USA, pp. 3.

Orhon D. and Artan N. (1994), Modelling of activated sludge systems, Technomic Publishing Co. Inc., Lancaster, PA, USA, pp. 589.

Painmanakul P., Wachirasak J., Jamnongwong M. and Hebrard G. (2009), Theoretical Prediction of volumetric mass transfer coefficient ($k_La$) for designing an aeration tank, *Engineering Journal,* **13**(3), 13–28.

Pamukoglu M.Y. and Kargi F. (2007), Mathematical modeling of copper (II) ion inhibition on COD removal in an activated sludge unit, *Journal of Hazardous Materials,* **146**(1), 372–377.

Qasim S.R. (1998), Wastewater Treatment Plants: Planning, Design, and Operation, Second Edition, CRC Press, pp. 1126.

Reinius L.G. and Hultgren J. (1988), Evaluation of the efficiency of a new aeration system at Henriksdal Sewage Treatment Plant, *Water Science Technology,* **20**(4-5), 85–92.

Roman M.D. and Mureşan M.V. (2014), Analysis of oxygen requirements and transfer efficiency in a wastewater treatment plant, *International Journal of Latest Research in Science and Technology,* **3**(2), 30–33.

Shammas N.K., Wang L.K., Pereira N.C. and Hung Y-T. (2009), Biological treatment processes, Volume 8, Springer Science Business Media, Humana Press, pp. 840.

Sharma K., Derlon N., Hu S. and Yuan Z. (2014), Modeling the pH effect on sulfidogenesis in anaerobic sewer biofilm, *Water Research,* **49**(2014), 175–185.

Singh K.P., Basant N., Malik A. and Jain G. (2010), Modeling the performance of “up-flow anaerobic sludge blanket” reactor based wastewater treatment plant using linear and nonlinear approaches—a case study, *Analytica Chimica Acta,* **658**(1), 1–11.

Szilveszter S., Ráduly B., Ábrahám B., Lányi S. and Niculae D.R. (2010), Mathematical models for domestic biological wastewater treatment process, *Environmental Engineering and Management Journal,* **9**(5), 629-636.

Toprak H. (2000), Operating of Activated Sludge Process, Dokuz Eylul University Publishing Corp., Izmir (in Turkish).

US EPA (United States Environmental Protection Agency) (1977), Aerobic Biological Wastewater Treatment Facilities, Process Control Manual, Municipal Operations Branch, Office of Water Program Operations, Washington DC, pp. 333.

Varolleghem P. (1994), On-line modelling of activated sludge process: Development of an adaptive sensor, Ph.D. Thesis, Faculty of Applied Biological Sciences, Section Environmental Technology, The University of Gent, Belgium, pp. 201.

von Sperling M. (2007), Activated sludge and aerobic biofilm reactors, Biological wastewater treatment, Vol. 5, IWA Publishing, pp. 322.

Wang L.K., Pereira N.C., Hung Y-T. and Shammas N.K. (2009), Advanced biological treatment processes, Handbook of Environmental Engineering, Vol. 8, Springer Science Business Media, Humana Press, pp. 818.

Yetilmezsoy K. and Abdul-Wahab S.A. (2012), A prognostic approach based on fuzzy-logic methodology to forecast $PM_{10}$ levels in Khaldiya residential area, Kuwait, *Aerosol and Air Quality Research,* **12,** 1217–1236.

Yetilmezsoy K. (2005), Optimisation using prediction models: air cyclones' body diameter/pressure drop, *Filtration & Separation,* **42**(10), 32–35.

Yetilmezsoy K.(2006), Determination of optimum body diameter of air cyclones using a new empirical model and a neural network approach, *Environmental Engineering Science,* **23**(4), 680–690.

Yetilmezsoy K. (2007), A new empirical model for the determination of the required retention time in hindered settling, *Fresenius Environmental Bulletin,* **16**(6), 674–684.

Yetilmezsoy K. (2010), Modeling studies for the determination of completely mixed activated sludge reactor volume: steady-state, empirical and ANN applications, *Neural Network World,* **20**(5), 559–589.

Yetilmezsoy K. (2012), Composite desirability function-based empirical modeling for packed tower design in physical ammonia absorption, *Asia-Pacific Journal of Chemical Engineering,* **7**(6), 795–813.

Yetilmezsoy K. (2012), Integration of kinetic modeling and desirability function approach for multi-objective optimization of UASB reactor treating poultry manure wastewater, *Bioresource Technology,* **118,** 89–101.

Yetilmezsoy K. (2016), A new simple model for the prediction of waste sludge flow rate in the steady-state completely mixed activated sludge process, *Environmental Engineering and Management Journal,* **15**(12), 2613–2630.

Yetilmezsoy K. and Abdul-Wahab S.A. (2014), A composite desirability function-based modeling approach in predicting mass condensate flux of condenser in seawater greenhouse, *Desalination,* **344,** 171–180.

Yetilmezsoy K., Demirel S. and Vanderbei R.J. (2009), Response surface modeling of Pb (II) removal from aqueous solution by Pistacia vera L.: Box–Behnken experimental design, *Journal of Hazardous Materials,* **171**(1), 551–562.

Yetilmezsoy K., Fingas M. and Fieldhouse B.(2011), An adaptive neuro-fuzzy approach for modeling of water-in-oil emulsion formation,*Colloids and Surfaces A: Physicochemical and Engineering Aspects,* **389**(1), 50-62.

Yetilmezsoy K., Ozgun H., Dereli R.K., Ersahin M.E. and Ozturk I. (2015), Adaptive neuro-fuzzy inference-based modeling of a full-scale expanded granular sludge bed reactor treating corn processing wastewater, *Journal of Intelligent & Fuzzy Systems,* **28**(4), 1601–1616.

Yetilmezsoy K. and Sakar S. (2008), Development of empirical models for performance evaluation of UASB reactors treating poultry manure wastewater under different operational conditions, *Journal of Hazardous Materials,* **153**(1-2), 532–543.

Yetilmezsoy K. and Sapci-Zengin Z. (2009), Stochastic modeling applications for the prediction of COD removal efficiency of UASB reactors treating diluted real cotton textile wastewater, *Stochastic Environmental Research and Risk Assessment,* **23**(1), 13–26.