Multivariable NNARMAX Model Identification of an AS-WWTP Using ARLS: Part 2 – Dynamic Modeling of the Secondary Settler and Clarifier

Vincent A. Akpan[1,*], Reginald A. O. Osakwe[2]

[1]Department of Physics Electronics, The Federal University of Technology, P.M.B. 704, Akure, Ondo State, Nigeria
[2]Department of Physics, The Federal University of Petroleum Resources, P.M.B. 1221, Effurun, Delta State, Nigeria

Abstract This paper presents the formulation and application of an online adaptive recursive least squares (ARLS) algorithm to the nonlinear model identification of the secondary settler/clarifier and the effluent tank of an activated sludge wastewater treatment (AS-WWTP). The performance of the proposed ARLS algorithm is compared with the so-called incremental backpropagation (INCBP) which is also an online identification. These algorithms are validated by one-step and five-step ahead prediction methods. The performance of the two algorithms is also assessed by using the Akaike's method to estimate the final prediction error (AFPE) of the regularized criterion. The validation results show the superior performance of the proposed ARLS algorithm in terms of much smaller prediction errors when compared to the INCBP algorithm.

Keywords Activated sludge wastewater treatment plant (AS-WWTP), Adaptive recursive least squares (ARLS), Artificial neural network (ANN), Benchmark simulation model No. 1 (BSM #1), Biological reactors, Effluent tank, incremental backpropagation (INCBP), Nonlinear model identification, Nonlinear neural network autoregressive moving average with exogenous input (NNARMAX) model, Secondary settler and clarifier

1. Introduction

Biological approach is the most widely used industrial technique in the wastewater treatment engineering community. The activated sludge process [1-5] is the most widely used system for biological wastewater treatment. Traditionally, activated sludge process (ASP) involve an anaerobic (anoxic) followed by an aerobic zone and a settler from which the major part of the biomass is recycled to the anoxic basin and this prevents washout of the process by decoupling the sludge retention time (SRT) from the hydraulic retention time (HRT). Activated sludge wastewater treatment plants (ASWWTP) are built to remove organic mater from wastewater where a bacterial biomass suspension (the activated sludge) is responsible for the removal of pollutants. Depending on the design and specific application, ASWWTP can achieve biological nitrogen removal, biological phosphorus removal and removal of organic carbon substances as well as the amount of dissolved oxygen [6]. Generally, an ASWWTP can generally be regarded as a complex system due to its highly nonlinear

dynamics, large uncertainty, multiple time scales in the internal processes as well as its multivariable structure [7-10]. A widely accepted biodegradation model is the activated sludge model no. 1 (ASM #1) which incorporate the basic biotransformation processes of an ASP [1, 2, 7].

Two of the critical points affecting classical biological reactors are controlling the sludge height in the secondary settler and the secondary settler efficiency, where the mixed liquor from the biological reactors is settled in order to obtain a clarified effluent prior to final discharge. The secondary settler parameters involved in the settling phase and those that highly affect the pollution removal process efficiency have been exploited by some authors [11-13]. A methodology to control the sludge blanket by measuring simple on-line (influent flow) parameters and off-line (a daily measured sludge volumetric index, SVI) parameters and manipulating the RAS recycle flow has been reported [14]. In the work of [14], the ability of fuzzy logic to integrate human knowledge was exploited to develop a fuzzy controller to maintain the secondary settler under stable conditions.

Biological nutrient removal (BNR) systems are reliable and effective in removing nitrogen and phosphorus. The process is based upon the principle that under specific conditions, microorganisms will remove more phosphorus and nitrogen than is required for biological activity.

However, several research papers have been published [8, 9], [15-19] concerning the BNR but the performance still depends on the biological activity and the process employed and it is an open intensive research field with emphasis on the effluent quality and operational cost.

The mixture of activated sludge and wastewater in the aeration tank together with the mixed liquor flows to a secondary settler and clarifier where the activated sludge is allowed to settle. The activated sludge is constantly growing and more is produced than can be returned for use in the aeration basin. Some of this sludge must be wasted to a sludge handling system for treatment and disposal (solids) or reuse (bio-solids). The volume of sludge returned to the aeration basins is normally 40 to 60% of the wastewater flow while the rest is wasted (solids) or reuse (bio-solids).

A number of factors affect the performance of the activated sludge system [20, 21]. These include the following: (*i*) temperature, (*ii*) return rates, (*iii*) amount of oxygen available, (*iv*) amount of organic matter available, (*v*) pH, (*vi*) waste rates, (*vii*) aeration time and (*viii*) wastewater toxicity. To obtain the desired level of performance in an activated sludge system, a proper balance must be maintained between the amount of food (organic matter), organisms (activated sludge), and dissolved oxygen (DO). The majority of problems with the activated sludge process result from an imbalance between these three items. The actual operation of an activated-sludge system is regulated by three factors: *1)* the quantity of air supplied to the aeration tank; *2)* the rate of activated-sludge recirculation; and *3)* the amount of excess sludge withdrawn form the system. Sludge wasting is an important operational practice because it allows the establishment of the desired concentration of mixed liquor soluble solids (MLSS), food to microorganisms ratio (F:M ratio), and the sludge age. The algorithm for computing these parameters are given in Appendices A, B and C. It should be mentioned here that air requirements in an activated sludge basin are governed by: *1)* BOD loading and the desired removal effluent, *2)* volatile suspended solids concentration in the aerator, and *3)* suspended solids concentration of the primary effluent (see Appendices A, B and C for details).

The aim of the paper is on the efficient modeling of a multivariable nonlinear neural network autoregressive moving average with exogenous input (NNARMAX) model that will capture the nonlinear dynamics of an AS-WWTP, trained with an online adaptive recursive least squares (ARLS) [22], for the purpose of developing an efficient nonlinear adaptive controller for the efficient control of the AS-WWTP process. Due to the multivariable nature of the AS-WWTP, the first paper (Part 1) considered the modeling of the five biological reactors including the influent tank [22] while the current paper (Part 2) is devoted to the modeling of the secondary settler and clarifier.

The paper is organized into six sections as follows. Section 2 presents the AS-WWTP process description while Section 3 highlights the operational considerations of the secondary settler and clarifier of the AS-WWTP process. Section 4 is devoted to the formulation of the neural network-based adaptive recursive least squares (ARLS) for model identification. Section 5 presents the neural network-based problem formulation for the secondary settler and the clarifier sections of the AS-WWTP process together with the simulation results and their discussion. The paper concludes in Section 6 with highlights on some major contributions in this paper with subsequent and future directions.

2. The AS-WWTP Process Description

Activated sludge wastewater treatment plants (WWTPs) are large complex nonlinear multivariable systems, subject to large disturbances, where different physical and biological phenomena take place. Many control strategies have been proposed for wastewater treatment plants but their evaluation and comparison are difficult [1-4]. This is partly due to the variability of the influent, the complexity of the physical and biochemical phenomena, and the large range of time constants (from a few minutes to several days) inherent in the activated sludge process. Additional complication in the evaluation is the lack of standard evaluation criteria.

With the tight effluent requirements defined by the European Union and to increase the acceptability of the results from wastewater treatment analysis, the generally accepted COST Actions 624 and 682 benchmark simulation model no. 1 (BSM1) model [1-4] is considered. The BSM1 model uses eight basic different processes to describe the biological behaviour of the AS-WWTP processes. The combinations of the eight basic processes results in thirteen different observed conversion rates as described in [1-4], [22, 23]. These components are classified into soluble components (S) and particulate components (X). The nomenclatures and parameter definitions used for describing the AS-WWTP in this work are given in Table 1. Moreover, four fundamental processes are considered: the growth and decay of biomass (heterotrophic and autotrophic), ammonification of organic nitrogen and the hydrolysis of particulate organics.

The activated sludge wastewater treatment plant considered here is strictly based on the benchmark simulation model no. 1 (BSM #1) proposed by the European Working Groups of COST Action 624 and 682 in conjunction with the International Water Association (IWA) Task Group on Benchmarking of Control Strategies for wastewater treatment plants (WWTPs) [1-4], [22, 23]. This implementation of the benchmark simulation model no. 1 (BSM #1) follows the methodology specified in [COST, 2008] especially from the viewpoint of control performances. The complete description of the conventional activated sludge wastewater treatment plant (AS-WWTP) based on the benchmark simulation model no. 1 (BSM1) used in this work is adapted from [1, 22, 23] together with the mathematical model of the benchmark simulation model no. 1 (BSM1) and

the MATLAB/Simulink programs that implements the mathematical model of the BSM #1 from [2-4], [23].

Table 1. The AS-WWTP Nomenclatures and Parameter Definitions

Parameters	Definition	Parameters	Definition
S_I*	Soluble inert organic matter	COD	Chemical oxygen demand
S_S*	Readily biodegradable substrate	BOD	Biochemical oxygen demand
X_I*	Particulate inert organic matter	IQ	Influent (inf) quality
X_S*	Slowly biodegradable substrate	EQ	Effluent (e) quality
X_{BH}*	Active heterotrophic biomass	QIN	Influent flow rate
X_{BA}*	Active autotrophic biomass	F_M_R	Food-to-microorganisms ratio
X_P*	Particulate products arising from biomass decay	Ntotal	Total nitrogen
S_O*	Soluble oxygen	AF1,AF2,AF3	Aeration control points for the aerated reactors
S_{NO}*	Nitrate and nitrite nitrogen	Qa1	Internal recycled nitrate (IRN) flow rates
S_{NH}*	Ammonia and ammonium nitrogen	Qa2	External recycled nitrate (ERN) flow rates
S_{ND}*	Soluble biodegradable organic nitrogen	Qf	Feed flow rates
X_{ND}*	Particulate biodegradable organic nitrogen	Qw	Waste activated sludge (WAS) flow rate
S_{ALK}*	Alkalinity	Qr	Recycled activated sludge (RAS) flow rates
TSS	Total soluble solids	Qe	Effluent flow rate
MA1, MA2	Mechanical aerators of the anaerobic and anoxic reactors	Qu	Sludge under flow rates
MLVSS	Mixed liquor volatile suspended solids	KLa	Mass transfer coefficient of the aerated reactors
IRN	Internal recycled nitrate	ERN	External recycled nitrate
Zf	Feed concentration	Ze	Effluent concentration
Zu	Settler underflow concentration	Zw	Waste activated sludge (WAS) concentration
Zr	Recycled activated sludge (RAS) concentration	PE	Pumping energy
AE	Aeration energy	DO	Dissolved oxygen
Za1	Internal recycled nitrate (IRN) concentration	Za2	External recycled nitrate (ERN) concentration

Note: (i) The numerical values of 1, 2, 3, 4, and 5 in front of each parameter correspond to the parameter description in the anaerobic, anoxic and the three aerated reactors respectively.

(ii) The *inf* and *E* (and sometimes *e)* refers to influent and effluent respectively.

(iii) Other parameters are introduced and defined as they are needed.

(iv) Notations with asterisk (*) are the state variables

Figure 1. The schematic of the AS-WWTP process

The schematic of a BNR-ASWWTP design with basic control strategies is shown in Fig.1 using the *Johannesburg* configuration [8, 9, 22] which consists of anaerobic, anoxic and aerobic zones as well as the secondary settler in a back-to-back scheme with multiple recycle streams. To ensure that plug flow conditions prevail in the bioreactors, the basins are usually partitioned such that back-mixing is minimized. The constructional features and nomenclature of the process together with the biological processes within the different zones of the reactors are detailed in [1-4], Appendices A and B of [22] as well as in Appendix C of [23]. The constructional features and the general characteristics of the secondary settler and clarifier and the criteria for evaluating and assessing the performances of the AS—WWTP control are given in Appendices A and B respectively.

3. The Operational Considerations of the Secondary Settler and Clarifier and Effluent Tank of the AS-WWTP

The activated sludge process generally requires more sampling and testing to maintain adequate process control than any of the other unit processes in the wastewater treatment system. During periods of operational problems, both the parameters tested and the frequency of testing may increase substantially. Process control testing may include the following [20, 21]: *1)* settleability testing to determine the settled sludge volume; *2)* suspended solids testing to determine influent and MLSS; *3)* RAS solids and WAS concentrations; *4)* determination of the volatile content of the mixed liquor suspended solids; *5)* DO and pH of the aeration tank; *6)* BOD and COD of the aeration tank influent and process effluent; and *7)* microscopic evaluation of the activated sludge to determine the predominant organism. To maintain the working organisms in the activated sludge process, it is necessary to ensure that a suitable environment is maintained by being aware of the many factors influencing the process and by monitoring them repeatedly. Control, in this case, can be defined as maintaining the proper solids (floc mass) concentration in the aerator for the incoming wastewater flow by adjusting the returned and waste activated sludge pumping rates and regulating the oxygen supply to maintain a satisfactory level of DO in the process.

Monitoring alkalinity in the aeration tank is essential to the control of the process. Insufficient alkalinity will reduce organism activity and may result in low effluent pH and, in some cases, extremely high chlorine demand in the disinfection process. However, the pH should also be monitored as part of the routine process control-testing schedule. Processes undergoing nitrification may show a significant decrease in effluent pH. Industrial waste discharges, septic wastes, or significant amounts of storm-water flows may produce wide variations in pH. Finally, because temperature directly affects the activity of the microorganisms, accurate monitoring of temperature can be helpful in identifying the causes of significant changes in organization populations or process performance.

The activated sludge process is an aerobic process that requires some DO be present at all times. The amount of oxygen required is dependent on the influent food (BOD), the activity of the activated sludge, and the degree of treatment desired. The mixed liquor suspended solids (MLSS), mixed liquor volatile suspended solids (MLVSS), and mixed liquor total suspended solids (MLTSS) needs to be determined and monitored, by adjusting the MLSS and MLVSS by increasing or decreasing the waste sludge rates. The MLSS or MLVSS can be used to represent the activated sludge or microorganisms present in the process. Process control calculations, such as sludge age and sludge volume index (SVI), cannot be calculated unless the MLSS is determined. The MLTSS is an important activated sludge control parameter. To increase the MLTSS, for example, the operator must decrease the waste rate or increase the mean cell residence time (MCRT). The MCRT must be decreased to prevent the MLTSS from changing when the number of aeration tanks in service is reduced.

The separation of solids and liquid in the secondary clarifier results in a blanket of solids. If solids are not removed from the clarifier at the same rate they enter, the blanket will increase in depth. If this occurs, the solids may carry over into the process effluent. The sludge blanket depth may be affected by other conditions, such as temperature variation, toxic wastes, or sludge bulking. The best sludge blanket depth is dependent upon such factors as hydraulic load, clarifier design, and sludge characteristics. The best blanket depth must be determined on an individual basis by experimentation. The sludge rate is also a critical control variable. The operator must maintain a continuous return of activated sludge to the aeration tank or the process will show a drastic decrease in performance. If the rate is too low, solids remain in the settling tank, resulting in solids loss and a septic return. If the rate is too high, the aeration tank can become hydraulically overloaded, causing reduced aeration time and poor performance. The return concentration is also important because it may be used to determine the return rate required to maintain the desired MLSS. Because the activated sludge contains living organisms that grow, reproduce, and produce waste matter, the amount of activated sludge is continuously increasing. If the activated sludge is allowed to remain in the system too long, the performance of the process will decrease. If too much activated sludge is removed from the system, the solids become very light and will not settle quickly enough to be removed in the secondary clarifier.

The biological oxygen demand (BOD) is a measure of the amount of food available to the microorganisms in a particular waste and it is measured by the amount of dissolved oxygen (DO) used up during a specific period of time (usually 5 days, BOD_5). The MLVSS is the organic matter in the mixed liquor suspended solids and is used to represent the amount of food (BOD) or COD) available per pound of MLVSS.

For the settling tank effluent, the DO levels of the activated sludge–settling tank must be maintained around 2 mg.L^1 as lower levels results in rising sludge. Normal pH in the activated sludge–settling tank must be maintained between 7 and 9. A decrease in pH indicates alkalinity deficiency and a lack of alkalinity prevents nitrification. On the other hand, for the settling tank effluent, an increase in BOD and TSS indicates decreasing treatment performance while a decrease indicates treatment performance is increasing. Moreover, an increase in Total Kjeldahl Nitrogen (TKN) indicates nitrification is increasing and a decrease indicates reduced nitrification.

The actual control of the sludge height and effluent quality of the AS-WWTP depends heavily on the different flow rates. Hydraulic overload of aeration and settling tanks, reduced aeration time, reduced settling time and loss of solid over time is an indication that the RAS return rate is too high. On the other hand, too low RAS returns rate results in septic return, solid build–up in settling tank, reduced MLVSS in aeration and loss of solid over weir. Moreover, the effects when the WAS flow rate is too high includes: reduced MLVSS, decreased sludge density, increased SVI, decreased MCRT, and increased F:M ratio; and vice versa when the WAS flow rate is too high. Finally, wasted energy, increased operating cost, rising solids and breakup of activated sludge are the effects when the aeration rate is too high while septic aeration tank, poor treatment performance and loss of nitrification results when the aeration rate is too low.

4. The Neural Network Identification Scheme and Validation Algorithms

4.1. Formulation of the Neural Network Model Identification Problem

The method of representing dynamical systems by vector difference or differential equations is well established in systems [24, 25] and control [26-28] theories. Assuming that a *p*-input *q*-output discrete-time nonlinear multivariable system at time k with disturbance $\tilde{d}(k)$ can be represented by the following Nonlinear AutoRegressive Moving Average with eXogenous inputs (NARMAX) model [24, 25]:

$$Y(k) = J\left[Y(k-1),\ldots,Y(k-n), \atop U(k-d),\ldots,U(k-d-m)\right] + \tilde{d}(k) \quad (1)$$

where $J(\bullet,\bullet)$ is a nonlinear function of its arguments, and $[U(k-d),\ldots,U(k-d-m)]$ are the past input vector, $[Y(k-1),\ldots,Y(k-n)]$ are the past output vector, $Y(k)$ is the current output, m and n are the number of past inputs and outputs respectively that define the order of the system, and d is time delay. The predictor form of (1) based on the information up to time $k-1$ can be expressed in the following compact form as [24, 25]:

$$\hat{Y}(k\,|\,k-1,\theta(k)) = J\left[\varphi(k,\theta(k)),\theta^T(k)\right] \quad (2)$$

where $\varphi(k,\theta(k)) = [Y(k-1),\ldots,Y(k-n),\ U(k-d),\ldots,$ $U(k-d-m),\ \varepsilon(k-1,\theta(k)),\ldots,\varepsilon(k-n,\theta(k))]^T$ is the regression (state) vector, $\theta(k)$ is an unknown parameter vector which must be selected such that $\hat{Y}(k\,|\,\theta(k)) \approx Y(k)$, $\varepsilon(k,\theta(k))$ is the error between (1) and (2) defined as

$$\varepsilon(k,\theta(k)) = Y(k) - \hat{Y}(k\,|\,\theta(k)) \quad (3)$$

where $k-1$ in $\hat{Y}(k\,|\,k-1,\theta(k))$ of (2) is henceforth omitted for notational convenience. Not that $\varepsilon(k,\theta(k))$ is the same order and dimension as $\hat{Y}(k\,|\,\theta(k))$.

Now, let Θ be a set of parameter vectors which contain a set of vectors such that:

$$\Theta:\ \theta(k) \in \mathbb{Q}_\theta \subset \mathfrak{R}^\nu \rightarrow \hat{\theta}(k) \quad (4)$$

where $\mathbb{Q}_\theta$ is some subset of $\mathfrak{R}^\nu$ where the search for $\hat{\theta}(k)$ is carried out; ν is the dimension of $\theta(k)$; $\hat{\theta}(k)$ is the desired vector which minimizes the error in (3) and is contained in the set of vectors $\Theta = \{\theta_1(k),\ldots,\theta_\tau(k)\}$; $\theta_1(k),\ldots,\theta_\tau(k)$ are distinct values of $\theta(k)$; and $\tau = 1,2,\ldots,\max_{iter}$ is the number of iterations required to determine the $\hat{\theta}(k)$ from the vectors in Θ.

Let a set of N input-output data pair obtained from prior system operation over NT period of time be defined:

$$Z^N = \{U(1),Y(1),\ldots,U(N),Y(N)\},\ \ N = 1,2,\ldots \quad (5)$$

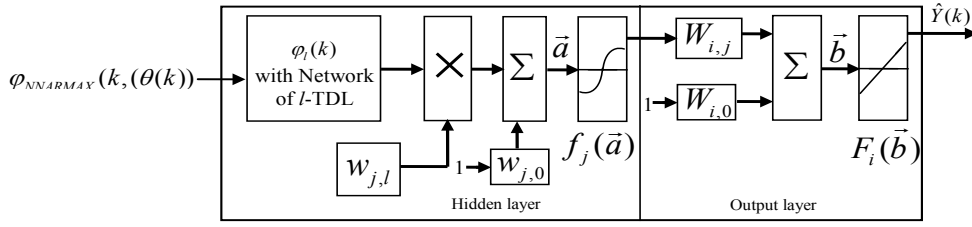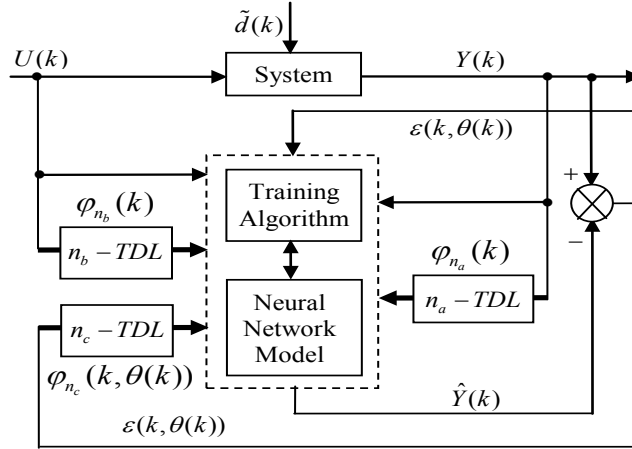where T is the sampling time of the system outputs. Then, the minimization of (3) can be stated as follows:

$$\hat{\theta}(k) = \arg\min_{\theta(k)} J(Z^N,\varphi(k,\theta(k)),\theta(k)) \quad (6)$$

where $J(Z^N,\varphi(k,\theta(k)),\theta(k))$ is formulated as a total square error (TSE) type cost function which can be stated as:

$$J(Z^N,\varphi(k,\theta(k)),\theta(k)) = \frac{1}{2N}\sum_{l=1}^N [\varepsilon(l,\theta(k))]^2 \quad (7)$$

The inclusion of $\theta(k)$ as an argument in $\varphi(k,\theta(k))$ is to account for the desired model $\hat{\theta}(k)$ dependency on $\tilde{d}(k)$. Thus, given as initial random value of $\theta(k)$, m, n and (5), the system identification problem reduces to the minimization of (6) to obtain $\hat{\theta}(k)$. For notational convenience, $J(\theta(k))$ shall henceforth be used instead of $J(Z^N,\varphi(k,\theta(k)),\theta(k))$.

4.2. Neural Network Identification Scheme

Figure 2.   Architecture of the dynamic feedforward NN (DFNN) model

Figure 3.   NN model identification based on the teacher-forcing method

The minimization of (6) is approached here by considering $\hat{\theta}(k)$ as the desired model of network and having the DFNN architecture shown in Fig. 2. The proposed NN model identification scheme based on the teacher-forcing method is illustrated in Fig. 3. Note that the "Neural Network Model" shown in Fig. 3 is the DFNN shown in Fig. 2. The inputs to NN of Fig. 3 are $\varphi_{l_m}(k) = [U(k-d),\dots,U(k-d-m)]$, $\varphi_{l_n}(k) = [Y(k-1),\dots,Y(k-n)]^T$ and $\varphi_{l_e}(k,\theta(k)) = [\varepsilon(k-1,\theta(k)),\dots,\varepsilon(k-n,\theta(k))]^T$ which are concatenated into $\varphi_l(k,\theta(k))$ as shown in Fig. 2. The output of the NN model of Fig. 3 in terms of the network parameters of Fig. 2 is given as:

$$\left.\begin{aligned}\hat{Y}(k\mid\hat{\theta}(k)) &= F_i\left(\sum_{j=1}^{n_h}W_{i,j}f_j(\vec{a})+W_{i,0}\right)\\ \vec{a} &= \sum_{l=1}^{n_\varphi}w_{j,l}\varphi_l(k,\theta(k))+w_{j,0}\end{aligned}\right\} \tag{8}$$

where n_h and n_φ are the number of hidden neurons and number of regressors respectively; i is the number of outputs, $w_{j,l}$ and $W_{i,j}$ are the hidden and output weights respectively; $w_{j,0}$ and $W_{i,0}$ are the hidden and output biases; $F_i(\vec{b})$ is a linear activation function for the output layer and $f_j(\vec{a})$ is an hyperbolic tangent activation function for the hidden layer defined here as:

$$f_j(\vec{a}) = 1 - \frac{2}{e^{2\cdot\vec{a}}+1} \tag{9}$$

Bias is a weight acting on the input and clamped to 1. Here, $\hat{\theta}(k)$ is a collection of all network weights and biases in (8) in term of the matrices $\mathbf{w} = \{w_{j,l}\ \ w_{j,0}\}$ and $\mathbf{W} = \{W_{i,j}\ \ W_{i,0}\}$. Equation (8) is here referred to as NN NARMAX (NNARMAX) model predictor for simplicity.

Note that $\tilde{d}(k)$ in (1) is unknown but is estimated here as a covariance noise matrix, $\Gamma[\theta(k)] = \mathbf{E}[\tilde{d}^T(k)\tilde{d}(k)]$. Using $\Gamma[\theta(k)]$, Equation (7) can be rewritten as:

$$J(\theta(k)) = \frac{1}{2N}\left(\begin{aligned}\sum_{l=1}^{N}\varepsilon[l,\theta(k)]^T\Gamma^{-1}[\theta(k)]\varepsilon[l,\theta(k)]\\ +\theta^T(k)D\theta(k)\end{aligned}\right) \tag{10}$$

where the second term in (10) is the regularization (weight decay) term [25, 27, 29] which has been introduced to reduce modeling errors, improve the robustness and performance of the two proposed training algorithms. $D = \alpha_d I = [\alpha_h\ \ \alpha_o]I$ is a penalty norm and also removes ill-conditioning, where I is an identity matrix, α_h and α_o are the weight decay values for the input-to-hidden and hidden-to-output layers respectively. Note that both $\hat{\Gamma}^{(j)}[\theta(k)]$ and D are adjusted simultaneously during network training with $\theta(k)$ and are used to update $\hat{\theta}(k)$ iteratively. The algorithm for estimating the covariance noise matrix and updating $\hat{\theta}(k)$ is summarized in Table 2. Note that this algorithm is

implemented at each sampling instant until $\hat{\Gamma}^{(j)}[\theta(k)]$ has reduced significantly as in step 7).

Table 2. Iterative algorithm for estimating covariance noise matrix

1) Given initial network weights $\theta(k) = \theta^{(0)}(k)$ and $j = j_{max}$.

2) For $k = 1$ to Number of Samples (N), Do,

3) Initialize $\Gamma^{(0)}[\theta(k)] = I$, Do,

4) Set $j = 1$

5) Train the network for τ iterations with a training algorithm using

$\Gamma^{(j-1)}[\theta(k)]$ to obtain $\hat{\theta}(k) \leftarrow \theta_\tau(k) \triangleq \theta^{(j)}(k)$.

6) Estimate the covariance matrix for the noise using

$$\hat{\Gamma}^{(j)}[\theta(k)] = \frac{1}{2N} \sum_{l=1}^{N} \varepsilon^T[\theta^{(j)}(k)]\varepsilon[\theta^{(j)}(k)]$$

7) If $\hat{\Gamma}^{(j)}[\theta(k)] < esp$, where esp is a convergence criteria.

Set $j \leftarrow j+1$ and Go To Step 4).

Else, set $\hat{\theta}(k) = \theta^{(j)}(k)$ and End *Set j*.

8) End *For k*.

Table 3. An algorithm for placing the roots of the time-varying filter of the NNARMAX model predictor with the unit circle for stability

1) Given network weights $\theta(k) = \theta^{(0)}(k)$, time-varying filter

$C(k, z^{-1}) = C^{(0)}(k, z^{-1})$ and regression vector $\varphi(k, \theta(k))$

2) Compute the roots of $C(k, z^{-1})$ as $C_{Roots}(k, z^{-1})$ and length

of $C_{Roots}(k, z^{-1})$ as l_{CRoots}.

3) Compute the absolute value of

$$C_{Roots}(k, z^{-1}) = abs(C_{Roots}(k, z^{-1}))$$

4) For $i = 1$ to l_{CRoots},

if $abs(C_{Roots}^{(i)}(k, z^{-1})) > 1$

$$C_{Roots}^{(i)}(k, z^{-1}) = \frac{1}{C_{Roots}^{(i)}(k, z^{-1})}$$

End if, End for

5) Compute the $C(k, z^{-1})$ using the real root from Step 4).

4.3. Formulation of the Neural Network-Based ARLS Algorithm

Unlike the BP which is a steepest descent algorithm, the ARLS and MLMA algorithms proposed here are based on the Gauss-Newton method with the typical updating rule [23, 25, 27, 28]:

$$\hat{\theta}(k) = \theta_\tau(k) + \Delta\theta_\tau(k) \qquad (11)$$

where

$$\Delta\theta_\tau(k) = -R[\theta_\tau(k)]^{-1} G[\theta_\tau(k)] \qquad (12)$$

$\theta_\tau(k)$ denotes the value of $\theta(k)$ at the current iterate τ, $\Delta\theta_\tau(k)$ is the search direction, $G[\theta_\tau(k)]$ and $R[\theta_\tau(k)]$ are the Jacobian (or gradient matrix) and the Gauss-Newton Hessian matrices evaluated at $\theta(k) = \theta_\tau(k)$.

As mentioned earlier, due to the model $\theta(k)$ dependency on the regression vector $\varphi(k, \theta(k))$, the NNARMAX model predictor depends on a *posteriori* error estimate using the feedback as shown in Fig. 2. Suppose that the derivative of the network outputs with respect to $\theta(k)$ evaluated at $\theta(k) = \theta_\tau(k)$ is given as

$$\psi[k, \theta_\tau(k)] = \frac{d\hat{Y}(k \mid \theta(k))}{d\theta(k)} \qquad (13)$$

The derivative of (13) is carried out in a BP fashion for the input-to-hidden layer and for the hidden-to-output layer respectively for the two-layer DFNN of Fig. 2. Thus, the derivative of the NNARMAX model predictor can be expressed as

$$\left.\begin{aligned}\psi[k, \theta_\tau(k)] &= \frac{\partial\hat{Y}(k \mid \theta(k))}{\partial\theta(k)} \\ &- \frac{\partial\hat{Y}(k \mid \theta(k))}{\partial\varepsilon(k-1, \theta(k))}\frac{\partial\hat{Y}(k-1 \mid \theta(k))}{\partial\theta(k)} \\ &- \cdots - \frac{\partial\hat{Y}(k \mid \theta(k))}{\partial\varepsilon(k-n_c, \theta(k))}\frac{\partial\hat{Y}(k-n_c \mid \theta(k))}{\partial\theta(k)}\end{aligned}\right\} \quad (14)$$

Thus, Equation (14) can be expressed equivalently as

$$\left.\begin{aligned}\psi[l, \theta(k)] &= \frac{d\hat{Y}(k \mid \theta(k))}{d\theta(k)} - C_1(k)\psi[k-1, \theta(k)] - \\ &\cdots - C_n(k)\psi[k-n, \theta(k)]\end{aligned}\right\} \quad (15)$$

By letting $C(k, z^{-1}) = I + C_1(k)z^{-1} + \cdots + C_n(k)z^{-n_c}$, then (15) can be reduced to the following form

$$\psi[k, \theta(k)] = \frac{1}{C(k, z^{-1})}\frac{d\hat{Y}(k \mid \theta(k))}{d\theta(k)} \qquad (16)$$

As it can be seen from (16), the gradient is calculated by filtering the partial derivatives with the time-varying filter $1/C(k, z^{-1})$ which depends on the prediction error based on the predicted output. Equation (16) is the only component

that actually impedes the implementation of the NN training algorithms depending on its computation.

Due to the feedback signals, the NNARMAX model predictor may be unstable if the system to be identified is not stable since the roots of (16) may, in general, not lie within the unit circle. The approach proposed here to iteratively ensure that the predictor becomes stable is summarized in the algorithm of Table 3. Thus, this algorithm ensures that roots of $C(k, z^{-1})$ lies within the unit circle before the weights are updated by the training algorithm proposed in the next sub-section.

4.3.1. The Adaptive Recursive Least Squares (ARLS) Algorithm

The proposed ARLS algorithm is derived from (11) with the assumptions that: *1)* new input-output data pair is added to Z^N progressively in a first-in first-out fashion into a sliding window, *2)* $\hat{\theta}(k)$ is updated after a complete sweep through Z^N, and *3)* all Z^N is repeated τ times. Thus, Equation (10) can be expressed as [23], [30]:

$$J(\theta(k)) = \frac{1}{2N} \sum_{l=1}^{N} \left(\begin{array}{c} \pi^{N-l} \varepsilon^T[l, \theta_\tau(k)] \Gamma^{-1}[\theta(k)] \varepsilon[l, \theta_\tau(k)] \\ + \theta^T(k) D \theta(k) \end{array} \right) (17)$$

$\pi \in (0,1)$ is the exponential forgetting and resetting parameter for discarding old information as new data is acquired online and progressively added to the set Z^N.

Assuming that $\theta(k-1)$ minimized (17) at time $k-1$; then using (17), the updating rule for the proposed ARLS algorithm can be expressed from (11) as:

$$\hat{\theta}(k) = \theta_\tau(k \mid k-1) - R[\theta_\tau(k \mid k-1)]^{-1} G[\theta_\tau(k \mid k-1)] (18)$$

where $G[\theta_\tau(k)]$ and $R[\theta_\tau(k)]$ given respectively as:

$$G[\theta_\tau(k)] = -\frac{1}{N} \left(\begin{array}{c} \psi[l, \theta_\tau(k-1)] \cdot \Gamma^{-1}[\theta(k)] \cdot \varepsilon[l, \theta_\tau(k-1)] \\ + D \theta_\tau(k-1) \end{array} \right)$$

$$R[\theta_\tau(k)] = R[l, \theta_\tau(k-1)] + \\ \frac{1}{N} \left(\begin{array}{c} \psi^T[l, \theta_\tau(k)] \cdot \Gamma^{-1}[\theta(k)] \cdot \psi[l, \theta_\tau(k)] \\ - R[l, \theta_\tau(k-1)] + D \end{array} \right) (19)$$

where $\psi[l, \theta_\tau(k)]$ is computed according to (16).

In order to avoid the inversion of $R[\theta_\tau(k)]$, Equation (19) is first computed as a covariance matrix estimate, $P(k)$, as

$$P(k) = \frac{1}{N} R[l, \theta_\tau(k)]^{-1}, \qquad l = 1, 2, \ldots, N \qquad (20)$$

Then, by using the following matrix inversion lemma:

$$[A + BCD]^{-1} = A^{-1} - A^{-1} B [DA^{-1} B + C^{-1}]^{-1} DA^{-1}$$

By setting $A = \pi R[l, \theta_\tau(k)]^{-1}$, $B = D^{-1}$ and $C = 1$, Equation (20) can also be expressed equivalently as

$$P(k) = \frac{1}{\pi} P(k-1) - \Lambda(k) \psi^T[l, \theta_\tau(k-1)] P(k-1) \\ + \beta I - \delta' P^2(k-1) \qquad (21)$$

where $\Lambda(k)$ is the adaptation factor given by

$$\Lambda(k) = \frac{\alpha P(k-1) \psi[l, \theta_\tau(k-1)]}{\left[\Gamma^{-1}[\theta(k)] + \psi^T[l, \theta_\tau(k-1)] P(k-1) \psi[l, \theta_\tau(k-1)] \right]}$$

and I is an identity matrix of appropriate dimension, α, β, δ' and π are four design parameters are selected such that the following conditions are satisfied [22], [23], 30]:

$$0 < \gamma < \alpha < 1, \qquad \beta > 0, \qquad \delta' > 0, \\ (\gamma - \alpha)^2 + 4\beta\delta' < (1-\alpha)^2 \qquad (22)$$

where $\alpha \in [0.1, 0.5]$ in $\Lambda(k)$ adjusts the gain of the (21), $\delta' \in [0, 0.01]$ is a small constant that is inversely related to the maximum eigenvalue of $P(k)$, $\pi \in [0.9, 0.99]$ is the exponential forgetting factor which is selected such that $\gamma \triangleq \frac{1-\pi}{\pi}$ and $\beta \in [0, 0.01]$ is a small constant which is related to the minimum $\overline{e}_{\min}$ and maximum $\overline{e}_{\max}$ eigenvalues of (21) given respectively as [22], [23], [30], [31]:

$$\overline{e}_{\min} = \left((\alpha - \gamma) / 2\delta' \right) \left(-1 + \sqrt{1 + \left(4\beta\delta' / (\alpha - \gamma)^2 \right)} \right) \\ \overline{e}_{\max} = \left(\gamma / 2\delta' \right) \left(1 + \sqrt{1 + \left(4\beta\delta' / \gamma^2 \right)} \right) \qquad (23)$$

The values of α, β, δ' and π in (22) is selected such that $\overline{e}_{\max} / \overline{e}_{\min} \simeq 10^4$ while the initial value of $P(k)$, that is $P(0)$, is selected such that $\overline{e}_{\min} I < P(0) < \overline{e}_{\max} I$ [27].

Thus, the ARLS algorithm updates based on the exponential forgetting and resetting method is given from (18) as

$$\hat{\theta}(k) = \theta_\tau(k-1) + \\ \Lambda(k) \Gamma^{-1}[\theta(k)][Y(k) - \hat{Y}(k \mid \theta_\tau(k-1))] \qquad (24)$$

where the second term in (20) is $\Delta\theta_\tau(k)$. Note that after $\hat{\theta}(k)$ has been obtained, the algorithm of Table 2 is implemented the conditions in Step *7)* of the Table 2 algorithm is satisfied.

4.4. Proposed Validation Methods for the Trained Multivariable NNARMAX Model

Network validations are performed to assess to what extend the trained network captures and represents the operation of the underlying system dynamics [25, 29].

The first test involves the comparison of the predicted outputs with the true training data and the evaluation of their corresponding one-step ahead prediction errors using (3).

The second validation test is the Akaike's final prediction error (AFPE) estimate [25, 29] based on the weight decay

parameter D in (10). A smaller value of the AFPE estimate indicates that the identified model approximately captures all the dynamics of the underlying system and can be presented with new data from the real process. Evaluating the $\varepsilon(k, \hat{\theta}(k))$ portion of (3) using the trained network with $\theta(k) = \hat{\theta}(k)$ and taking the expectation $\mathbf{E}\{J(Z^N, \hat{\theta}(k))\}$ with respect to $\varphi(k)$ and $\tilde{d}(k)$ leads to the following AFPE estimate [23, 30, 31]:

$$\hat{F}(Z^N \hat{\theta}(k)) \approx \frac{N + p_a}{N - p_b} J(Z^N \hat{\theta}(k)) + \gamma \qquad (25)$$

where

$$p_a = \mathbf{tr}\left\{ V(\hat{\theta}(k)) \left[V(\hat{\theta}(k)) + D \right]^{-1} V(\hat{\theta}(k)) \left[V(\hat{\theta}(k)) + D \right]^{-1} \right\}$$

and $\mathbf{tr}\{\cdot\}$ is the trace of its arguments and it is computed as the sum of the diagonal elements of its arguments, $p_b = tr\{ V(\hat{\theta}^*)[V(\hat{\theta}^*) + (1/N) D]^{-1} \}$ and $\boldsymbol{\gamma}$ is a positive quantity that improves the accuracy of the estimate and can be computed according to the following expression:

$$\gamma = \frac{\hat{\theta}(k)^T D}{N^2} \left(R[\hat{\theta}(k)] + \frac{D}{N} \right)^{-1} R[\hat{\theta}(k)] \left(R[\hat{\theta}(k)] + \frac{D}{N} \right)^{-1} D\hat{\theta}(k)$$

The third method is the K-step ahead predictions [10] where the outputs of the trained network are compared to the unscaled output training data. The K-step ahead predictor follows directly from (8) and for $\varphi(k) = \hat{\varphi}(k + K)$ and $\theta(k) = \hat{\theta}(k)$, takes the following form:

$$\hat{Y}((k + K) \mid k, \hat{\theta}) = \hat{J}(Z^N, \hat{\varphi}(k + K), \hat{\theta}(k)) \qquad (26)$$

where

$$\hat{\varphi}(k + K) = [U((k + K - 1) \mid \hat{\theta}), \ldots, \ U((k + K - m) \mid \hat{\theta}),$$

$$\hat{Y}((k + K - 1) \mid \hat{\theta}), \ldots, \hat{Y}((k + K + 1 - \min(k, n)) \mid \hat{\theta}),$$

$$Y((k + K - 1) \mid \hat{\theta}), \ldots, Y((k + K - \max(n - k, 0) \mid \hat{\theta})]^T$$

The mean value of the K-step ahead prediction error (MVPE) between the predicted output and the actual training data set is computed as follows:

$$MVPE = mean\left(\sum_{k=m+K}^{N} \frac{Y(k) - \hat{Y}((k + K) \mid k, \hat{\theta})}{Y(k)} \right) \times 100\% \quad (27)$$

where $Y(k)$ corresponds to the unscaled output training data and $\hat{Y}((k + K) \mid k, \hat{\theta})$ the K-step ahead predictor output.

5. Integration and Formulation of the NN-Based AS-WWTP Problem

5.1. Selection of the Manipulated Inputs and Controlled Outputs of the AS-WWTP Process

This section concentrates on the nonlinear model identification, predictions and control of the sludge height in

the secondary settler and clarifier as well as the quality of the discharged effluent. The proposed scheme for the ANN-based nonlinear model identification and prediction, fuzzy rule-based decision system and intelligent adaptive control strategy for the for the sludge height in the secondary settler and clarifier as well the quality of the discharged effluent in the effluent tank of the AS-WWTP process is illustrated in Fig. 4.

Case I: Nonlinear Model Identification and Prediction of the Sludge Height in the Secondary Settler and Clarifier

The objectives here is to maintain the quality and regulate the quantity of the disposed and recycled sludge by manipulating the feed flow rate (Q_{F2} = 36892 $m^3.d^{-1}$) depending on the feed flow decision system, the RAS recycled flow rates (Q_{R2} = 18446 $m^3.d^{-1}$ and Q_{R3} = 18446 $m^3.d^{-1}$), and the WAS flow rate (Q_W = 385 $m^3.d^{-1}$) using TSS2, TSS3, TSS4, TSS5, TSS6, TSS7, TSS8, TSS9, and TSS10 as inputs with the following additional as inputs parameters: depending on the feed flow decision system using SI_s, SND_s, SNH_s, SNO_s, XND_s, XI_s, XBH_s, XBA_s, XP_s, SS_s, XS_s, SN_s, SALK_s, TSS_s as inputs.

Case II: Nonlinear Model Identification and Prediction of the Effluent Quality in the Effluent Tank

The objectives here is to maintain the quality and regulate the quantity of the discharged effluent by manipulating the effluent flow rate (Q_E = 18061 $m^3.d^{-1}$) and the feed flow rate (Q_{F3} = 36892 $m^3.d^{-1}$) depending on the feed flow decision system using SI_e, SND_e, SNH_e, SNO_e, XND_e, XI_e, XBH_e, XBA_e, XP_e, SS_e, XS_e, SN_e, SALK_e, TSS_e as inputs.

The European Union regulations and the COST 624 restrictions on effluent quality (EQ) stipulates that SNH_e = 4 gm^{-3}, N_{total_e} = 17 gm^{-3}, TSS_E = 30 gm^{-3}, BOD_5_e = 2 gm^{-3}, $COD_{total,5}\_e$ = 48.2 gm^{-3}, and EQ = 7550 kgd^{-1}. These six parameters: SNH_e, N_{total_e}, TSS_e, BOD_5_e, $COD_{total,5}\_e$, and EQ have also been chosen as the constraint decision variables in this work.

The ANN identification and prediction #1 block accepts its inputs from the effluent tank to generate the NN model and compute the predicted values of all the inputs. The constrained multi-objective optimization and decision block (Fig. 4) uses the predicted values to compute: SNH_e, N_{total_e}, TSS_E, BOD_5_e, $COD_{total,5}\_e$, and EQ. Based on the restricted effluent quality, genetic algorithm is used to perform the multi-objective optimization problem to determine the optimal values of the five constraint decision variables which forms the reference (set-point) values for the controlling and regulating the effluent quality and flow rate by the ANN identification, prediction and control #2 block (Fig. 4). This block also gives the final predicted values of

the six parameters describing the effluent quality (SNH_e, N_total_e, TSS_E, BOD_5_e, COD_total,5_e, and EQ) which would be used in future work to the compute the control signals to manipulate the following flow rates: Q_{inf}, Q_E, Q_{A1}, Q_{A2}, Q_{F1}, Q_{F2}, Q_{R1}, Q_{R2}, Q_{R3}, and Q_W as well as Q_{F12} and Q_{R13}.
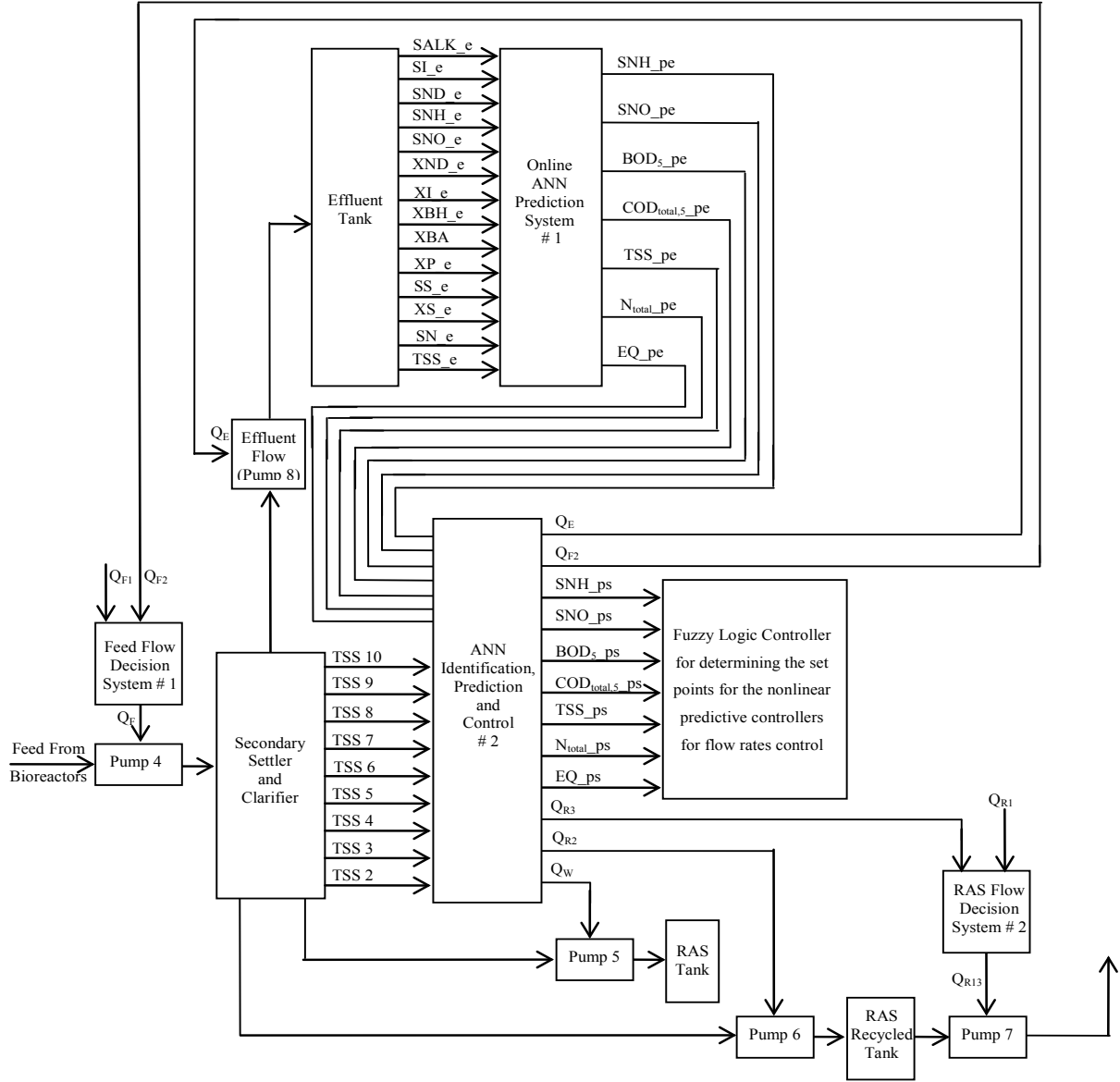

Figure 4. The proposed scheme for the multivariable neural network-based NNARMAX model identification and prediction, fuzzy rule-based decision system and intelligent adaptive control strategy for the sludge height in the secondary settler and clarifier as well the quality of the discharged effluent in the effluent tank of the AS-WWTP process

5.2. Formulation of the AS-WWTP Model Identification Problem

5.2.1. Statement of the AS-WWTP Neural Network Model Identification Problem

The activated sludge wastewater treatment plant model defined by the benchmark simulation model no. 1 (BSM1) is described by eight coupled nonlinear differential equations given in Appendix A. The BSM1 model consist of thirteen states defined in Table 1 but they are redefined here for the effluent tank (with E and subscript e and pe for inputs and outputs respectively) as follows: E_{SI_e}, E_{SS_e}, E_{XI_e}, E_{XS_e}, E_{XBH_e}, E_{XBA_e}, E_{XP_e}, E_{SO_e}, E_{SNO_e}, E_{SNH_e}, E_{SND_e}, E_{XND_e}, E_{SALK_e} and for the secondary settler (with SS and subscript s and ps for inputs and outputs respectively) as follows SS_{SI_p}, SS_{SS_p}, SS_{XI_p}, SS_{XS_e}, SS_{XBH_p}, SS_{XBA_p}, SS_{XP_p}, SS_{SO_p}, SS_{SNO_p}, SS_{SNH_p}, SS_{SND_p}, SS_{XND_p}, SS_{SALK_p}. Out of thirteen states, only four states are measurable namely: E_{SS_e}, SS_{SS_p} (readily biodegradable substrate), E_{XBH_e}, SS_{XBH_p} (active heterotrophic biomass), E_{SO_e}, SS_{SO_p} (oxygen) and E_{SNO_e},

SS_{SNO_p} (nitrate and nitrite nitrogen). An additional important parameter E_{TSS_e} and SS_{TSS_p} are used to assess the amount of soluble solids in the effluent tank and secondary settler/clarifier respectively.

As highlighted above, the main objective here is on the efficient neural network model identification to obtain a multivariable NNARMAX model equivalent of the secondary settler/clarifier and the effluent tank of the activated sludge wastewater treatment plant (AS-WWTP) with a view in using the obtained model together with those from [22] to design a multivariable intelligent adaptive predictive control for the AS-WWTP process in our future work. Thus, from the discussions so far, the measured inputs that influence the behaviour of the AS-WWTP shown in Fig. 4 and Fig. 5 are:

$$
\begin{aligned}
U(k) = \Big[& E_{SI\_e}(k) \;\; E_{SND\_e}(k) \;\; E_{SNH\_e}(k) \;\; E_{SNO\_e}(k) \\
& E_{XND\_e}(k) \;\; E_{XI\_e}(k) \;\; E_{XBH\_e}(k) \;\; E_{XBA\_e}(k) \\
& E_{XP\_e}(k) \;\; E_{SS\_e}(k) \;\; E_{XS\_e}(k) \;\; E_{SN\_e}(k), E_{SALK\_e}(k) \\
& E_{TSS\_e}(k); \;\; SS_{TSS2\_s}(k) \;\; SS_{TSS3\_s}(k) \;\; SS_{TSS4\_s}(k) \\
& SS_{TSS5\_s}(k) \;\; SS_{TSS6\_s}(k) \;\; SS_{TSS7\_s}(k) \;\; SS_{TSS8\_s}(k) \\
& SS_{TSS9\_s}(k) \;\; SS_{TSS10\_s}(k) \;\; SS_{SI\_s}(k) \;\; SS_{SND\_s}(k) \\
& SS_{SNH\_s}(k) \;\; SS_{SNO\_s}(k) \;\; SS_{XND\_s}(k) \;\; SS_{XI\_s}(k) \\
& SS_{XBH\_s}(k) \;\; SS_{XBA\_s}(k) \;\; SS_{SS\_s}(k) \;\; SS_{XP\_s}(k) \\
& SS_{XS\_s}(k) \;\; SS_{SN\_s}(k) \;\; SS_{SALK\_s}(k) \;\; SS_{TSS\_s}(k) \Big]^T
\end{aligned}
\tag{28}
$$

Furthermore, based on the discussions thus far, the output parameters that capture the behaviour of the AS-WWTP are defined here as:

$$
\begin{aligned}
Y(k) = \Big[& E_{SNH\_pe}(k) \;\; E_{SNO\_pe}(k) \;\; E_{BOD\_pe}(k) \;\; E_{COD\_pe}(k) \\
& E_{TSS\_pe}(k) \;\; E_{N_{total}\_pe}(k) \;\; E_{EQ\_pe}(k) \;\; E_{SI\_pe}(k) \\
& E_{SND\_pe}(k) \;\; E_{XND\_pe}(k) \;\; E_{XI\_pe}(k) \;\; E_{XBH\_pe}(k) \\
& E_{XBA\_pe}(k) \;\; E_{XP\_pe}(k) \;\; E_{SS\_pe}(k) \;\; E_{XS\_pe}(k) \\
& E_{SN\_pe}(k) \;\; E_{SALK\_pe}(k); \;\; SS_{QE\_ps}(k) \;\; SS_{QF2\_ps}(k) \\
& SS_{SNH\_ps}(k) \;\; SS_{SNO\_ps}(k) \;\; SS_{BOD\_ps}(k) \;\; SS_{COD_{total}\_ps}(k) \\
& SS_{TSS\_ps}(k) \;\; SS_{N_{total}\_ps}(k) \;\; SS_{EQ\_ps}(k) \;\; SS_{QR2\_ps}(k) \\
& SS_{QR3\_ps} \;\; SS_{QW\_ps}(k) \;\; SS_{SI\_ps}(k) \;\; SS_{SND\_ps}(k) \\
& SS_{XP\_ps}(k) \;\; SS_{XND\_ps}(k) \;\; SS_{XI\_ps}(k) \;\; SS_{XBH\_ps}(k) \\
& SS_{XBA\_ps}(k) \;\; SS_{SS\_ps}(k) \;\; SS_{XS\_ps}(k) \;\; SS_{SN\_ps}(k) \\
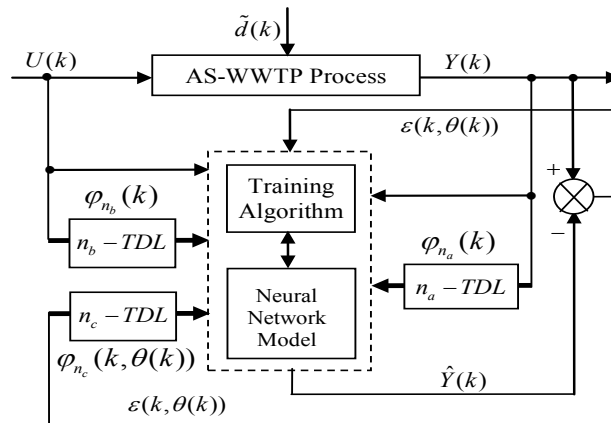& SS_{SALK\_ps}(k) \Big]^T
\end{aligned}
\tag{29}
$$

Figure 5. The neural network model identification scheme for AS-WWTP based on NNARMAX model

Although, the multivariable system is formulated as 23–input 19–output control problem, but the neural network model identification is a much more complicated multiple–input multiple–output (MIMO) problem since all the 14 states of the secondary settler and the effluent tank must be predicted at each sampling instant in order to obtain a reasonable approximate model that describes and captures the system's dynamics at each time instant; thus, making the total inputs 37 and total outputs 41. Additional complexity arises from the number of past inputs and outputs in the regression matrix that defines the system. The neural network identification scheme used here is shown in Fig. 5 and is based on the NNARMAX model predictor discussed in Section 4. The input vector to the neural network (NN) consists of the regression vectors which are concatenated into $\varphi_{NNARMAX}(k, \theta(k))$ for the NNARMAX models predictors discussed in Section 4 and defined here as Equation (30), (31), (32) and (33) below.

The outputs of the neural network for the secondary settler with the clarifier and the effluent tank are the predicted values of the 13 states each together with the amount of total soluble solids (*TSS*), thus resulting in fourteen states; which together with the outputs of blocks #1 and #2 in Fig. 4 gives 41 outputs to be predicted at each sampling time instant as given in (34) below.

$$
\begin{aligned}
\varphi_{n_a}(k) = \big[& E_{SNH\_pe}(k-n_a) \ \ E_{SNO\_pe}(k-n_a) \ \ E_{BOD\_pe}(k-n_a) \ \ E_{COD\_pe}(k-n_a) \ \ E_{TSS\_pe}(k-n_a) \ \ E_{N_{total}\_pe}(k-n_a) \\
& E_{EQ\_pe}(k-n_a) \ \ E_{SI\_pe}(k-n_a) \ \ E_{SND\_pe}(k-n_a) \ \ E_{XND\_pe}(k-n_a) \ \ E_{XI\_pe}(k-n_a) \ \ E_{XBH\_pe}(k-n_a) \\
& E_{XBA\_pe}(k-n_a) \ \ E_{XP\_pe}(k-n_a) \ \ E_{SS\_pe}(k-n_a) \ \ E_{XS\_pe}(k-n_a) \ \ E_{SN\_pe}(k-n_a) \ \ E_{SALK\_pe}(k-n_a); \\
& SS_{QE\_ps}(k-n_a) \ \ SS_{QF2\_ps}(k-n_a) \ \ SS_{SNH\_ps}(k-n_a) \ \ SS_{SNO\_ps}(k-n_a) \ \ SS_{BOD\_ps}(k-n_a) \ \ SS_{COD_{total}\_ps}(k-n_a) \\
& SS_{TSS\_ps}(k-n_a) \ \ SS_{N_{total}\_ps}(k-n_a) \ \ SS_{EQ\_ps}(k-n_a) \ \ SS_{QR2\_ps}(k-n_a) \ \ SS_{QR3\_ps}(k-n_a) \ \ SS_{QW\_ps}(k-n_a) \\
& SS_{SI\_ps}(k-n_a) \ \ SS_{SND\_ps}(k-n_a) \ \ SS_{XP\_ps}(k-n_a) \ \ SS_{XND\_ps}(k-n_a) \ \ SS_{XI\_ps}(k-n_a) \ \ SS_{XBH\_ps}(k-n_a) \\
& SS_{XBA\_ps}(k-n_a) \ \ SS_{SS\_ps}(k-n_a) \ \ SS_{XS\_ps}(k-n_a) \ \ SS_{SN\_ps}(k-n_a) \ \ SS_{SALK\_ps}(k-n_a) \big]^T
\end{aligned} \tag{30}
$$

$$
\begin{aligned}
\varphi_{n_b}(k) = \big[& E_{SI\_e}(k-n_b) \ \ E_{SND\_e}(k-n_b) \ \ E_{SNH\_e}(k-n_b) \ \ E_{SNO\_e}(k-n_b) \ \ E_{XND\_e}(k-n_b) \ \ E_{XI\_e}(k-n_b) \\
& E_{XBH\_e}(k-n_b) \ \ E_{XBA\_e}(k-n_b) \ \ E_{XP\_e}(k-n_b) \ \ E_{SS\_e}(k-n_b) \ \ E_{XS\_e}(k-n_b) \ \ E_{SN\_e}(k-n_b) \\
& E_{TSS\_e}(k-n_b); \ SS_{TSS2\_s}(k-n_b) \ \ SS_{TSS3\_s}(k-n_b) \ \ SS_{TSS4\_s}(k-n_b) \ \ SS_{TSS5\_s}(k-n_b) \ \ SS_{TSS6\_s}(k-n_b) \\
& SS_{TSS7\_s}(k-n_b) \ \ SS_{TSS8\_s}(k-n_b) \ \ SS_{TSS9\_s}(k-n_b) \ \ SS_{TSS10\_s}(k-n_b) \ \ SS_{SI\_s}(k-n_b) \ \ SS_{SND\_s}(k-n_b) \\
& SS_{SNH\_s}(k-n_b) \ \ SS_{SNO\_s}(k-n_b) \ \ SS_{XND\_s}(k-n_b) \ \ SS_{XI\_s}(k-n_b) \ \ SS_{XBH\_s}(k-n_b) \ \ SS_{XBA\_s}(k-n_b) \\
& SS_{XP\_s}(k-n_b) \ \ SS_{SS\_s}(k-n_b) \ \ SS_{XS\_s}(k-n_b) \ \ SS_{SN\_s}(k-n_b) \ \ SS_{TSS\_s}(k-n_b) \big]^T
\end{aligned} \tag{31}
$$

$$
\begin{aligned}
\varphi_{n_c}(k) = \big[& \varepsilon_{E_{SNH\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{SNO\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{BOD\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{COD\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{TSS\_pe}}(k-n_c, \theta(k)) \\
& \varepsilon_{E_{N_{total}\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{EQ\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{SI\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{SND\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{XND\_pe}}(k-n_c, \theta(k)) \\
& \varepsilon_{E_{XI\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{XBH\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{XBA\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{XP\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{SS\_pe}}(k-n_c, \theta(k)) \\
& \varepsilon_{E_{XS\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{SN\_pe}}(k-n_c, \theta(k)) \ \ \varepsilon_{E_{SALK\_pe}}(k-n_c, \theta(k)); \ \ \varepsilon_{SS_{QE\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{QF2\_ps}}(k-n_c, \theta(k)) \\
& \varepsilon_{SS_{SNH\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{SNO\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{BOD\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{COD_{total}\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{TSS\_ps}}(k-n_c, \theta(k)) \\
& \varepsilon_{SS_{N_{total}\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{EQ\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{QR2\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{QR3\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{QW\_ps}}(k-n_c, \theta(k)) \\
& \varepsilon_{SS_{SI\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{SND\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{XP\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{XND\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{XI\_ps}}(k-n_c, \theta(k)) \\
& \varepsilon_{SS_{XBH\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{XBA\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{SS\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{XS\_ps}}(k-n_c, \theta(k)) \ \ \varepsilon_{SS_{SN\_ps}}(k-n_c, \theta(k)) \\
& \varepsilon_{SS_{SALK\_ps}}(k-n_c, \theta(k)) \big]^T
\end{aligned} \tag{32}
$$

$$
\varphi_{NNARMAX}(k, \theta(k)) = \big[\varphi_{n_a}(k) \ \ \varphi_{n_b}(k) \ \ \varphi_{n_c}(k, \theta(k)) \big] \tag{33}
$$

$$\hat{Y}(k) = \left[\hat{y}_{SNH\_pe}(k) \quad \hat{y}_{SNO\_pe}(k) \quad \hat{y}_{BOD\_pe}(k) \quad \hat{y}_{COD\_pe}(k) \quad \hat{y}_{TSS\_pe}(k) \quad \hat{y}_{N_{total}\_pe}(k) \quad \hat{y}_{EQ\_pe}(k) \quad \hat{y}_{SI\_pe}(k) \quad \hat{y}_{SND\_pe}(k) \right.$$

$$\hat{y}_{XND\_pe}(k) \quad \hat{y}_{XI\_pe}(k) \quad \hat{y}_{XBH\_pe}(k) \quad \hat{y}_{XBA\_pe}(k) \quad \hat{y}_{XP\_pe}(k) \quad \hat{y}_{SS\_pe}(k) \quad \hat{y}_{XS\_pe}(k) \quad \hat{y}_{SN\_pe}(k) \quad \hat{y}_{SALK\_pe}(k);$$

$$\hat{y}_{QE\_ps}(k) \quad \hat{y}_{QF2\_ps}(k) \quad \hat{y}_{SNH\_ps}(k) \quad \hat{y}_{SNO\_ps}(k) \quad \hat{y}_{BOD\_ps}(k) \quad \hat{y}_{COD_{total}\_ps}(k) \quad \hat{y}_{TSS\_ps}(k) \quad \hat{y}_{N_{total}\_ps}(k) \quad \hat{y}_{EQ\_ps}(k) \tag{34}$$

$$\hat{y}_{QR2\_ps}(k) \quad \hat{y}_{QR3\_ps} \quad \hat{y}_{QW\_ps}(k) \quad \hat{y}_{SI\_ps}(k) \quad \hat{y}_{SND\_ps}(k) \quad \hat{y}_{XP\_ps}(k) \quad \hat{y}_{XND\_ps}(k) \quad \hat{y}_{XI\_ps}(k) \quad \hat{y}_{XBH\_ps}(k)$$

$$\left. \hat{y}_{XBA\_ps}(k) \quad \hat{y}_{SS\_ps}(k) \quad \hat{y}_{XS\_ps}(k) \quad \hat{y}_{SN\_ps}(k) \quad \hat{y}_{SALK\_ps}(k) \right]^T$$

Since disturbances play an important role in the evaluation of controller performances, three influent disturbance data are defined for different weather conditions, namely: dry-weather data, rain weather data, and storm weather data. The data for these three influent disturbances are provided by the European COST Actions for evaluating controller performances [2-4]. In this study, the dry weather influent data is used in order to measure how well the trained neural network mimic the dynamics of the AS-WWTP process to meet the control requirement specified above. The dry weather data contains two weeks of influent data at 15 minutes sampling interval. Although, disturbances $\tilde{d}(k)$ affecting the AS-WWTP are incorporated into dry-weather data provided by the COST Action Group, additional sinusoidal disturbances with non-smooth nonlinearities are introduced for closed-loop performances evaluation based on an updated neural network model at each sampling time instants.
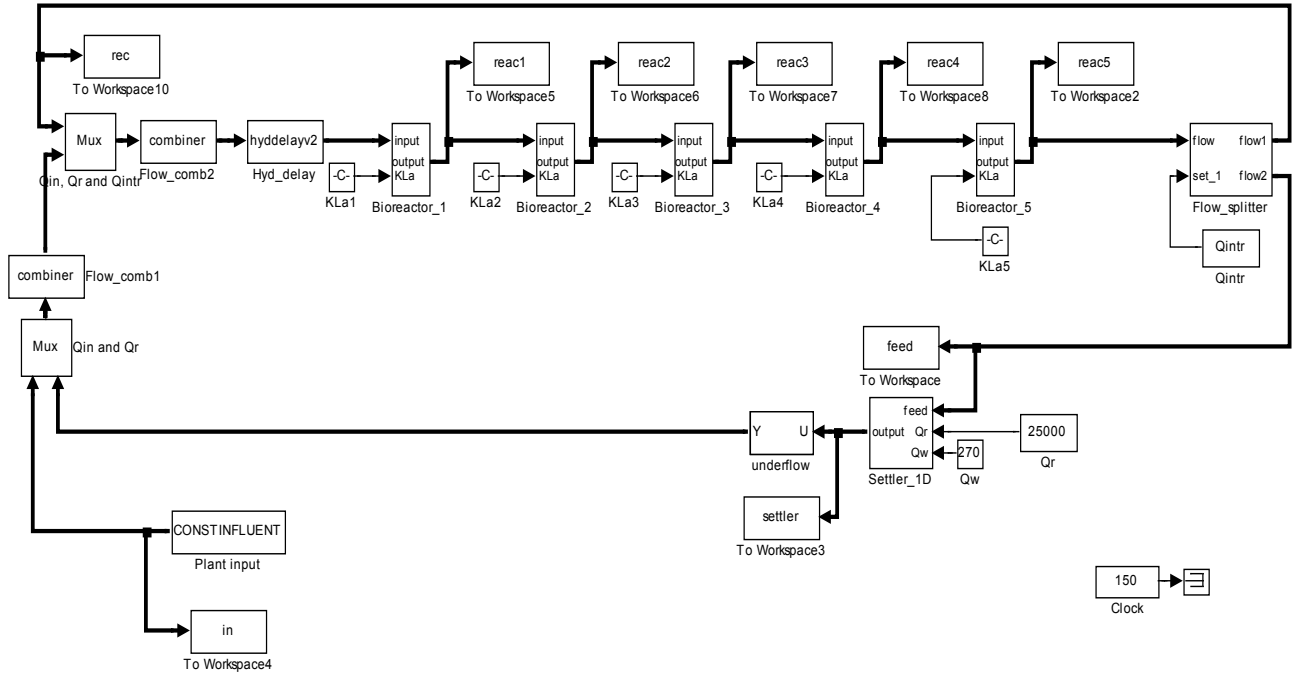

Figure 6. Open-loop steady-state benchmark simulation model No.1 (BSM1) with constant influent

5.2.2. Experiment with the BSM1 for AS-WWTP Process Neural Network Training Data Acquisition

For the efficient control of the activated sludge wastewater treatment plant (AS-WWTP) using neural network, a neural network (NN) model of the AS-WWTP process is needed which requires that the NN be trained with dynamic data obtained from the AS-WWTP process. In other to obtain dynamic data for the NN training, the validated and generally accepted COST Actions 624 benchmark simulation model no. 1 (BSM1) is implemented and simulated using MATLAB and Simulink as shown in Fig. 6. The BSM1 process model for the AS-WWTP process is given in Appendix A.

A two-step simulation procedure defined in the simulation benchmark [2]–[4], [23] is used in this study. The first step is the steady state simulation using the constant influent flow (CONSTINFLUENT) for 150 days as shown and implemented in Fig. 6. Note that each simulation sample period indicated by the "*Clock*" of the AS-WWTP Simulink model in Fig. 6 corresponds to one day. In the second step, starting from the steady state solution obtained with the CONSTINFLUENT data and using the dry-weather influent weather data (DRYINFLUENT) as inputs, the AS-WWTP process is then simulated for 14 days using the same Simulink model of Fig. 6 but by replacing the CONSTINFLUENT influent data with the DRYINFLUENT influent data. This second simulation generates 1345 dynamic data in which is used for NN training while the 130 first day dry-weather data samples provided by the COST

Actions 624 and 682 is used for the trained NN validation.

5.2.3. The Incremental or Online Back-Propagation (INCBP) Algorithm

In order to investigate the performance of the ARLS, the so-called incremental (or online) back-propagation (INCBP) algorithm is used for this purpose. The incremental or online back-propagation (INCBP) algorithm was originally proposed by [32] which has been modified in [23] is used in this paper. The incremental back-propagation (INCBP) algorithm is easily derived by setting the covariance matrix $P(k) = \mu I$ on the left hand side of (20) in Section 4.3.1 under the formulation of the ARLS algorithm; that is:

$$\mu_\tau I = \frac{1}{k} \sum_{t=1}^{k} R[t, \theta_\tau(k)]^{-1} \tag{35}$$

where μ is the step size and I is an identity matrix of appropriate dimension. Next, the basic back-propagation given from [27] as:

$$\hat{\theta}(k) = \theta_\tau(k) - \mu_\tau \left. \frac{dJ[\theta(k)]}{d\theta(k)} \right|_{\theta(k)=\theta_\tau(k)} \tag{36}$$

is used to update the algorithm in (35). Finally, all that is required is to specify a suitable step size μ and carry out the recursive computation of the gradient given by (36).

5.2.4. Scaling the Training Data and Rescaling the Trained Network that Models the AS-WWTP Process

Due to the fact the input and outputs of a process may, in general, have different physical units and magnitudes; the scaling of all signals to the same variance is necessary to prevent signals of largest magnitudes from dominating the identified model. Moreover, scaling improves the numerical robustness of the training algorithm, leads to faster convergence and gives better models. The training data are scaled to unit variance using their mean values and standard deviations according to the following equations:

$$U^{(s)}(k) = \frac{U(k) - \overline{U}(k)}{\sigma_{U(k)}} \text{ and } Y^{(S)}(k) = \frac{Y(k) - \overline{Y}(k)}{\sigma_{Y(k)}} \tag{37}$$

where $\overline{U}(k)$, $\overline{Y}(k)$ and $\sigma_{U(k)}$, $\sigma_{Y(k)}$ are the mean and standard deviation of the input and output training data pair; and $U^{(S)}(k)$ and $Y^{(S)}(k)$ are the scaled inputs and outputs respectively. Also, after the network training, the joint weights are rescaled according to the expression

$$\hat{Y}(k, \hat{\theta}(k)) = \hat{Y}(k, \hat{\theta}(k))\sigma_{Y(k)} + \overline{Y}(k) \tag{38}$$

so that the trained network can work with other unscaled validation data and test data not used for training. However, for notational convenience, $U(k) = U^{(S)}(k)$ and $Y(k) = Y^{(S)}(k)$ shall be used.

5.2.5. Training the Neural Network that Models the Secondary Settler/Clarifier and Effluent Tank

The NN input vector to the neural network (NN) is the NNARMAX model regression vector $\varphi_{NNARMAX}(k, \theta(k))$ defined by (33). The input $\varphi_{n_c}(k, \theta(k))$, that is the initial error estimates $\varepsilon(k, \theta(k))$ given by (32), is not known in advance and it is initialized to small positive random matrix of dimension n_c by n_c. The outputs of the NN are the predicted values of $\hat{Y}(k)$ given by (34).

For assessing the convergence performance, the network was trained for $\tau = 100$ epochs (number of iterations) with the following selected parameters: $p = 37$, $q = 41$, $n_a = 2$, $n_b = 2$, $n_c = 2$, $n_\varphi = 238$ (NNARMAX), $n_h = 5$, $n_o = 41$, $\alpha_h = 1e-6$ and $\alpha_o = 1e-5$. The details of these parameters are discussed in Section 3; where p and q are the number of inputs and outputs of the system, n_a, n_b and n_c are the orders of the regressors in terms of the past values, n_φ is the total number of regressors (that is, the total number of inputs to the network), n_h and n_o are the number of hidden and output layers neurons, and α_h and α_o are the hidden and output layers weight decay terms. The four design parameters for adaptive recursive least squares (ARLS) algorithm defined in (22) are selected to be: α=0.5, β=5e-3, δ'=1e-5 and π=0.99 resulting to γ=0.0101. The initial values for $\bar{e}_{min}$ and $\bar{e}_{max}$ in (23) are equal to 0.0102 and 1.0106e+3 respectively and were evaluated using (23). Thus, the ratio of $\bar{e}_{min}/\bar{e}_{max}$ from (23) is 9.9018e+4 which imply that the parameters are well selected. Also, $\mu = 1e-3$ is selected to initialize the INCBP algorithm given in (36).

The 1345 dry-weather training data is first scaled using equation (37) and the network is trained for $\tau = 100$ epochs using the proposed adaptive recursive least squares (ARLS) and the incremental back-propagation (INCBP) algorithms proposed in Sections 4.3 and 5.2.3. After network training, the trained network is again rescaled respectively according to (38), so that the resulting network can work or be used with unscaled AS-WWTP data. Although, the convergence curves of the INCBP and the ARLS algorithms for 100 epochs each are not shown but the minimum performance indexes for both algorithms are given in the third rows of Tables 4 and 5 for the secondary settler/clarifier and effluent tank. As one can observe from these Tables, the ARLS has smaller performance index when compared to the INCBP which is an indication of good convergence property of the ARLS at the expense of higher computation time when compared the small computation time used by the INCBP for 100 epochs as evident in the first rows of Tables 4 and 5.

The total square error (TSE) discussed in subsection 4.1, for the network trained with the INCBP and the ARLS algorithms are given in the second rows of Tables 4 and 5. Again, the ARLS algorithm also has smaller mean square

errors and minimum performance indices when compared to the INCBP algorithm. The small values of the total square error (TSE) and the minimum performance indices indicate that ARLS performs better than the INCBP for the same number of iterations (epochs). These small errors suggest that the ARLS model approximates better the secondary settler/clarifier and effluent tank of the AS-WWTP process giving smaller errors than the INCBP model.

5.3. Validation of the Trained NNARMAX Model of the AS-WWTP Process

According to the discussion on network validation in Section 4.4, a trained network can be used to model a process once it is validated and accepted, that is, the network demonstrates its ability to predict correctly both the data that were used for its training and other data that were not used during training. The network trained by the INCBP and the proposed ARLS algorithms has been validated with three different methods by the use of scaled and unscaled training data as well as with the 130 dry-weather data reserved for the validation of the trained network for the secondary settler/clarifier and effluent tank of the AS-WWTP process.

The plots shown in (a) to (c) of Fig. 7, Fig. 8 and Fig. 9 corresponds to results obtained for the secondary settler and clarifier while those shown in (d) and (e) of Fig. 7, Fig. 8 and Fig. 9 are the results obtained for the effluent tank. The output parameters obtained in each of (a) to (e) were previously defined in Section 5.2.1 during the problem formulation but they are redefined here again as they appear in the next few figures (Fig. 7, Fig. 8 and Fig. 9) as follows:

(a) is from the secondary settler and clarifier for Q_{F2}, Q_W, Q_{R2} and Q_{F3};

(b) is from the secondary settler and clarifier for QE_ps, E_Q_ps, Ntotal_ps and TSS_ps;

(c) is from the secondary settler and clarifier for COD_ps, BOD_ps, SNO_ps and SNH_ps;

(d) is from the effluent tank for TSS_pe, BOD_pe, COD_pe and Ntotal_pe;

(e) is from the effluent tank for EQ_pe, SNO_pe and SNH_pe.

The secondary settler and the clarifier depends on the following input parameters TSS2_s, TSS3_s, TSS4_s, TSS5_s, TSS6_s, TSS7_s, TSS8_s, TSS9_s, and TSS10_s together with the additional input parameter SI_s, SND_s, SNH_s, SNO_s, XND_s, XI_s, XBH_s, XBA_s, XP_s, SS_s, XS_s, SN_s, SALK_s, TSS_s as inputs to compute the outputs parameters given in (a), (b) and (c) above.

The effluent tank depends on the following input parameters SI_e, SND_e, SNH_e, SNO_e, XND_e, XI_e, XBH_e, XBA_e, XP_e, SS_e, XS_e, SN_e, SALK_e, TSS_e as inputs to compute the outputs parameters given respectively in steps (d) and (e) above.

In addition to the outputs given in (a) to (e) above, the following outputs parameters: SI_ps, SND_ps, SNH_ps, SNO_ps, XND_ps, XI_ps, XBH_ps, XBA_ps, XP_ps, SS_ps, XS_ps, SN_ps, SALK_ps, TSS_ps for the secondary settler and clarifier as well as SI_pe, SND_pe, SNH_pe, SNO_pe, XND_pe, XI_pe, XBH_pe, XBA_pe, XP_pe, SS_pe, XS_pe, SN_pe, SALK_pe, TSS_pe for the effluent tank are also predicted that are all used to compute the respective constraint parameters for the flow rates decision variables in order manipulating the pumps. The results for these additional outputs are not shown here for space economy rather only the results for the actual outputs are shown.

Table 4(a). Secondary settler and clarifier with the flow rates parameters predictions

	Q_{F2}		Q_W		Q_{R2}	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	4.7268e+00	1.2418e+01	4.1808e+00	1.3494e+01	4.3680e+00	1.3042e+01
Total square error (TSE)	3.7785e+03	1.3846e-02	8.4725e+01	1.3300e-01	8.9394e-01	1.0859e-02
Minimum performance index	1.1111e-01	1.3642e-04	2.7379e-01	2.7277e-05	4.9970e-01	2.4429e-05
Mean error of one step ahead prediction of training data	2.3598e+01	2.2086e-01	4.8445e+01	6.4563e-02	1.2542e-01	1.4129e-03
Mean error of one step prediction of test data	3.0642e-02	4.8566e-04	3.0274e-01	7.4624e-05	2.4421e-01	2.9394e-05
Mean value of 5-step ahead prediction error	2.0799e-02	1.0294e-03	4.2539e+00	5.3471e-01	8.1505e-04	3.7579e-05
Akaike's final prediction error (AFPE) estimate	7.5675e+06	8.0059e+03	3.4868e+03	3.4542e-01	2.7653e-01	1.3230e-05

Table 4(b). Secondary settler and clarifier with the flow rates parameters predictions

	Q_{R3}		Q_E		EQ_ps	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	4.4460e+00	2.8361e+01	4.4616e+00	1.2745e+01	4.4928e+00	1.2542e+01
Total square error (TSE)	2.0723e-01	1.3010e-02	6.1591e+03	1.4480e-02	2.4746e+03	1.2639e-03
Minimum performance index	2.3132e-01	2.2653e-05	5.7763e-01	9.3172e-05	3.2264e-01	3.4483e-03
Mean error of one step ahead prediction of training data	3.2182e-01	1.0311e-03	9.6652e+02	1.3650e-01	1.0871e+03	6.9857e-01
Mean error of one step prediction of test data	4.9750e-01	1.1728e-05	2.9026e-01	5.3270e-04	2.2114e-01	7.8258e-04
Mean value of 5-step ahead prediction error	2.5628e-03	2.4933e-05	3.5867e+00	5.1073e-00	3.0270e+01	4.8898e-01
Akaike's final prediction error (AFPE) estimate	1.3640e-01	1.3078e-05	4.0143e+07	6.3631e-03	3.7055e+06	3.9143e-04

Vincent A. Akpan *et al.*: Multivariable NNARMAX Model Identification of an AS-WWTP
Using ARLS: Part 2 – Dynamic Modeling of the Secondary Settler and Clarifier

Table 4(c). Secondary settler and clarifier with the flow rates parameters predictions

	N_{total_ps}		TSS_ps		COD_ps	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	4.4772e+00	1.2776e+01	4.2276e+00	1.2652e+01	4.3212e+00	1.3010e+01
Total square error (TSE)	1.0855e+00	9.1302e-02	1.6997e+00	1.0189e-02	1.4965e+00	9.4150e-01
Minimum performance index	5.5744e-01	4.8806e-04	5.5744e-01	3.4972e-03	5.7086e-01	1.3648e-03
Mean error of one step ahead prediction of training data	2.6441e-01	7.0259e-03	2.7052e+00	1.2117e-01	4.1079e+00	4.0013e-02
Mean error of one step prediction of test data	1.8333e-02	1.8520e-04	4.9481e-01	4.5158e-03	3.4363e-01	7.1299e-04
Mean value of 5-step ahead prediction error	1.5775e+00	6.1830e-01	1.1093e+01	1.4722e-01	4.7815e+00	6.3471e-01
Akaike's final prediction error (AFPE) estimate	2.2534e+00	3.0280e-03	9.6502e+00	5.8161e-02	2.6921e+01	6.2357e-02

Table 4(d). Secondary settler and clarifier with the flow rates parameters predictions

	BOD_ps		SNO_ps		SNH_ps	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	4.3212e+00	1.2823e+01	4.2900e+00	1.2901e+01	4.3056e+00	1.2558e+01
Total square error (TSE)	6.6523e-02	6.1962e-03	9.2767e-01	8.6691e-02	3.4690e+00	8.9834e-02
Minimum performance index	7.1825e-02	7.9210e-04	1.9957e-02	2.8498e-04	2.3132e-01	2.2653e-05
Mean error of one step ahead prediction of training data	6.2113e-03	1.7525e-04	2.3800e-01	9.3384e-03	1.3259e+00	2.3595e-04
Mean error of one step prediction of test data	5.8925e-02	5.2515e-04	7.3932e-02	3.2977e-05	2.4330e-01	5.2090e-05
Mean value of 5-step ahead prediction error	1.0057e+00	6.1215e-01	6.2721e+00	9.9042e-01	1.7461e+01	1.3057e+01
Akaike's final prediction error (AFPE) estimate	1.4761e-03	1.6177e-05	1.7277e-01	2.6034e-03	3.7425e+00	2.2511e-03

Table 5(a). Effluent tank and the effluent constrained parameters predictions

	TSS_pe		BOD_pe		COD_pe		N_{total_pe}	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	4.2900e+00	1.2714e+01	4.3992e+00	1.2730e+01	4.4148e+00	1.2511e+01	4.5084e+00	2.8829e+01
Total square error (TSE)	1.4828e+00	8.0397e-02	1.2636e-01	6.4199e-03	2.2213e-01	1.0684e-01	1.8858e+00	3.4063e-02
Minimum performance index	1.0197e-01	1.7543e-04	2.1038e-01	1.9652e-04	4.6426e-01	4.1338e-04	4.4253e-01	8.0309e-05
Mean error of one step ahead prediction of training data	7.1218e-01	2.9174e-02	9.9223e-02	2.9070e-05	2.1703e+00	1.5154e-02	6.1978e-01	1.0336e-02
Mean error of one step prediction of test data	3.0642e-02	4.8566e-04	3.0274e-01	7.4624e-05	3.0642e-02	4.8566e-04	3.0274e-01	7.4624e-05
Mean value of 5-step ahead prediction error	3.7343e-01	1.9528e+00	1.3726e+01	6.3707e-01	6.5585e+00	3.9099e-01	6.9951e+00	4.0417e-01
Akaike's final prediction error (AFPE) estimate	3.5342e+00	5.6495e-03	4.7602e-03	4.3409e-06	2.6288e+01	2.1616e-02	3.2223e+00	5.8032e-04

Table 5(b). Effluent tank and the effluent constrained parameters predictions

	EQ_pe		SNO_pe		SNH_pe	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	4.4148e+00	1.2730e+01	4.4460e+00	1.2636e+01	4.4616e+00	1.2948e+01
Total square error (TSE)	1.1869e+03	5.1786e+02	1.7746e+00	1.8578e-02	3.4100e+00	5.6248e-02
Minimum performance index	4.3682e-01	2.2938e-03	2.9654e-01	4.2429e-05	2.2740e-01	6.5602e-05
Mean error of one step ahead prediction of training data	7.7791e+02	1.2262e+01	2.5907e+00	2.2112e-03	1.0252e+00	7.4323e-03
Mean error of one step prediction of test data	3.0642e-02	4.8566e-04	3.0274e-01	7.4624e-05	2.0392e-01	2.5707e-04
Mean value of 5-step ahead prediction error	1.0844e+01	5.5355e+00	4.7470e+00	9.5292e-01	5.2291e+01	2.6106e+01
Akaike's final prediction error (AFPE) estimate	8.3834e+06	4.2942e+04	2.7096e+00	3.8483e-04	3.3786e+00	9.6484e-04

5.3.1. Validation by the One-Step Ahead Predictions
       Simulation

In the one-step ahead prediction method, the errors obtained from one-step ahead output predictions of the trained network are assessed. In Fig. 7(a)–(e) the graphs for the one-step ahead predictions of the scaled training data (blue-) against the trained network output predictions (red --*) using the neural network models trained by INCBP and ARLS algorithms respectively are shown for 100 epochs.

The mean value of the one-step ahead prediction errors are given in the fourth rows of Tables 4 and 5 respectively. It can be seen in the figures that the network predictions of the training data closely match the original training data. Although, the scaled training data prediction errors by both algorithms are small, the ARLS algorithm appears to have a much smaller error when compared to the INCBP algorithm as shown in the fourth rows of Tables 4 and 5. These small one-step ahead prediction errors are indications that the networks trained using the ARLS captures and approximate the nonlinear dynamics of the secondary settler/clarifier and effluent tank of the AS-WWTP process to a high degree of accuracy. This is further justified by the small mean values of the TSE obtained for the networks trained using the proposed ARLS algorithms for the process as shown in the second rows of Tables 4 and 5.

Furthermore, the suitability of the INCBP and the proposed ARLS algorithms for neural network model identification for use in the real AS-WWTP industrial environment is investigated by validating the trained network with the 130 unscaled dynamic data obtained for dry-weather as provided by the COST Action Group. Graphs of the trained network predictions (red --*) of the validation (test) data with the actual validation data (blue -) using the INCBP and the proposed ARLS algorithms are shown in Fig. 8(a)–(e) for the secondary settler/clarifier and effluent tank of the AS-WWTP process based on the selected process parameters.

The almost identical prediction of these data proves the effectiveness of the proposed approaches. The prediction accuracies of the unscaled test data by the networks trained using the INCBP and the proposed ARLS algorithm evaluated by the computed mean prediction errors shown in the fifth rows of Tables 4 and 5. Again, one can observe that although the validation data prediction errors obtained by both algorithms are small, the validation data predictions errors obtained with the model trained by the proposed ARLS algorithm appears much smaller when compared to those obtained by the model trained using the INCBP algorithm. These predictions of the unscaled validation data given in Fig. 8(a)–(e) as well as the mean value of the one step ahead validation (test) prediction errors in the fifth rows of Tables 4 and 5 verifies the neural network ability to model accurately the dynamics for the secondary settler/clarifier and effluent tank of the AS-WWTP process based on the dry-weather influent data using the proposed ARLS training algorithm.
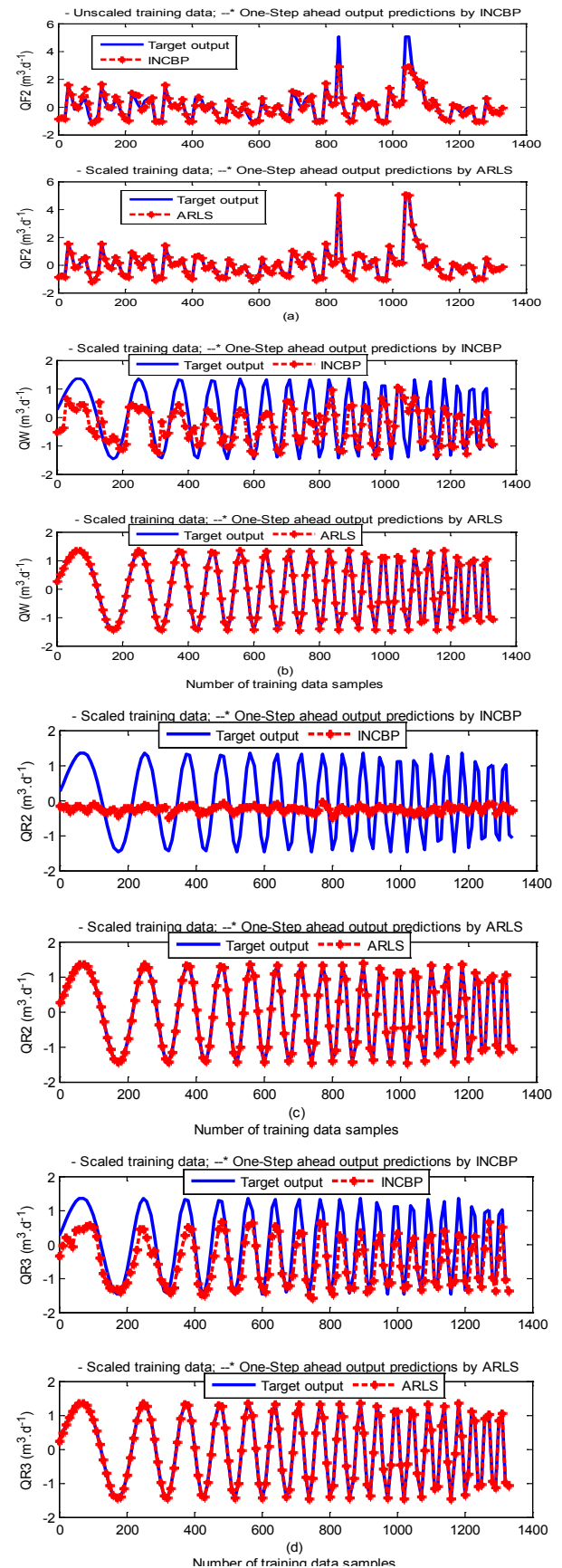

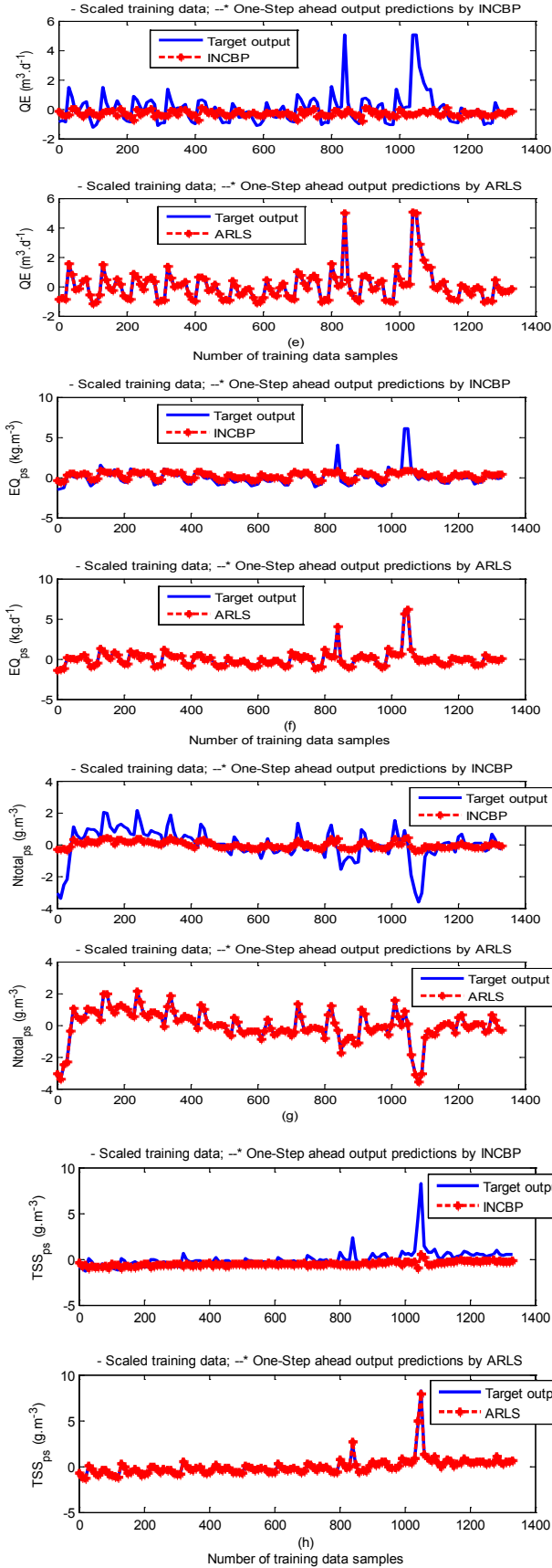**Figure 7(a).** One-step ahead prediction of scaled Q_{F2}, Q_W, Q_{R2} and Q_{F3} training data

Figure 7(b). One-step ahead prediction of scaled Q_{E_ps}, EQ_ps, N_{total_ps} and TSS_ps training data
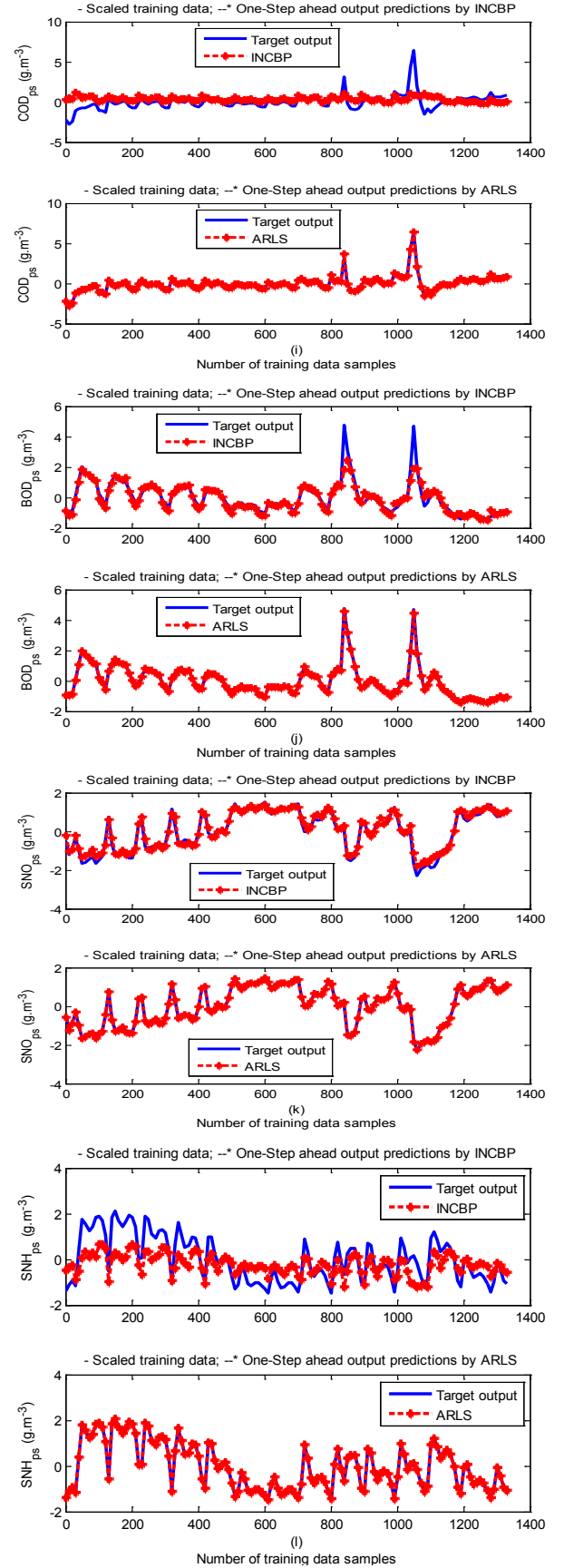
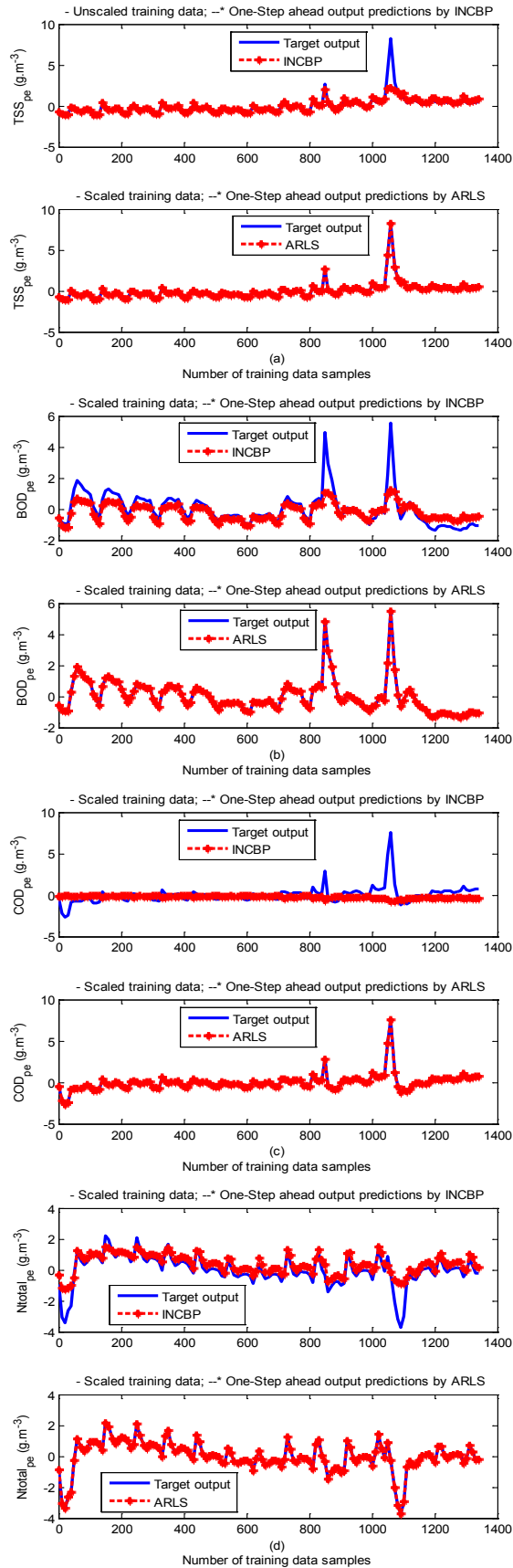**Figure 7(c).** One-step ahead prediction of scaled COD_ps, BOD_ps, SNO_ps and SNH_ps training data

Figure 7(d). One-step ahead prediction of scaled TSS_pe, BOD_pe, COD_pe and N_total_pe training data

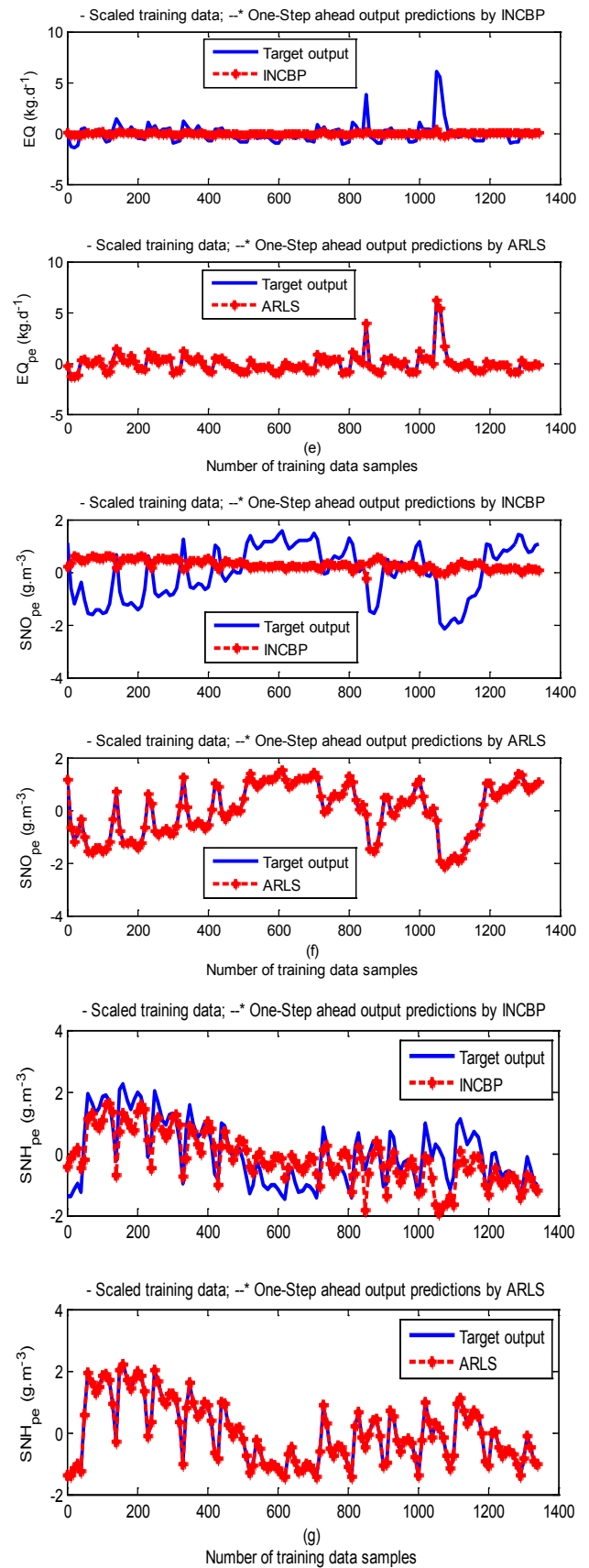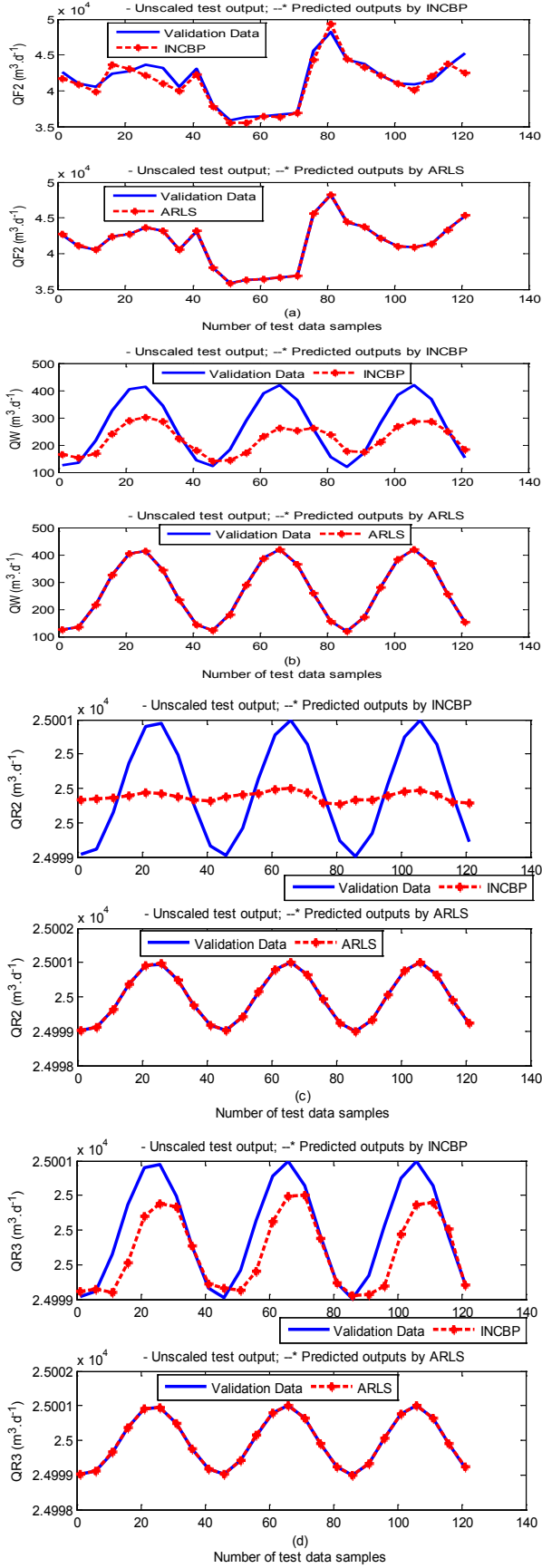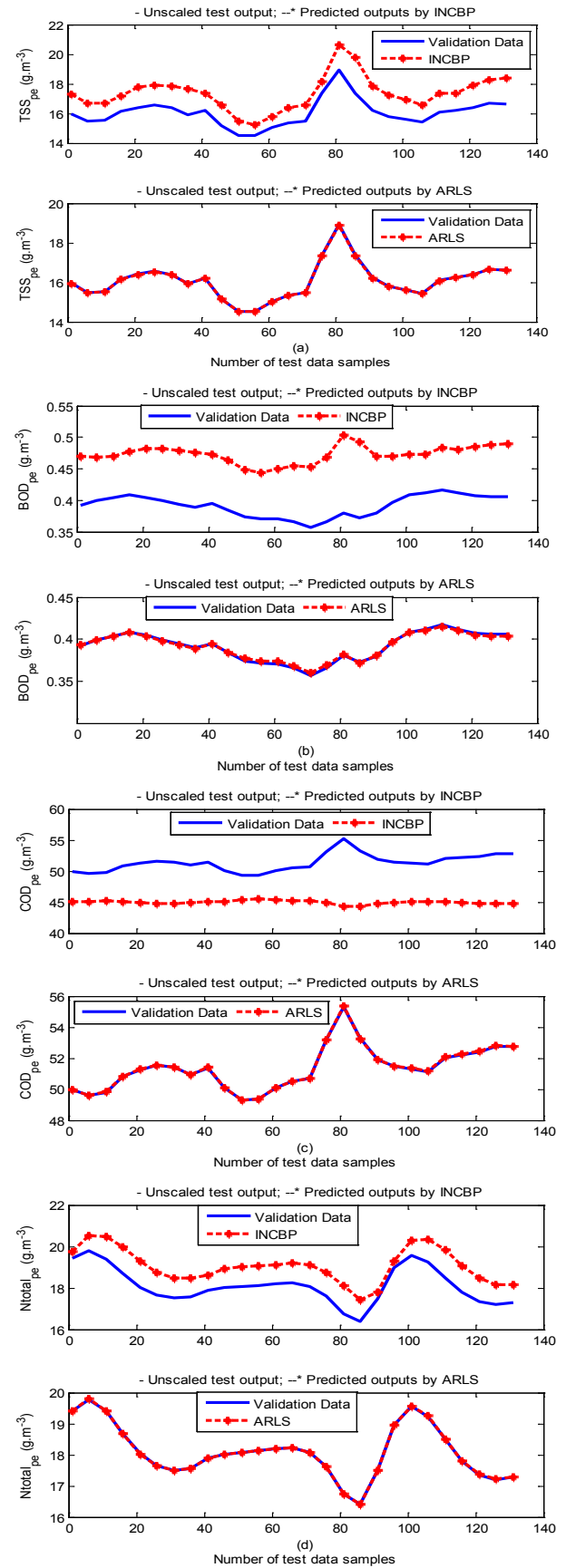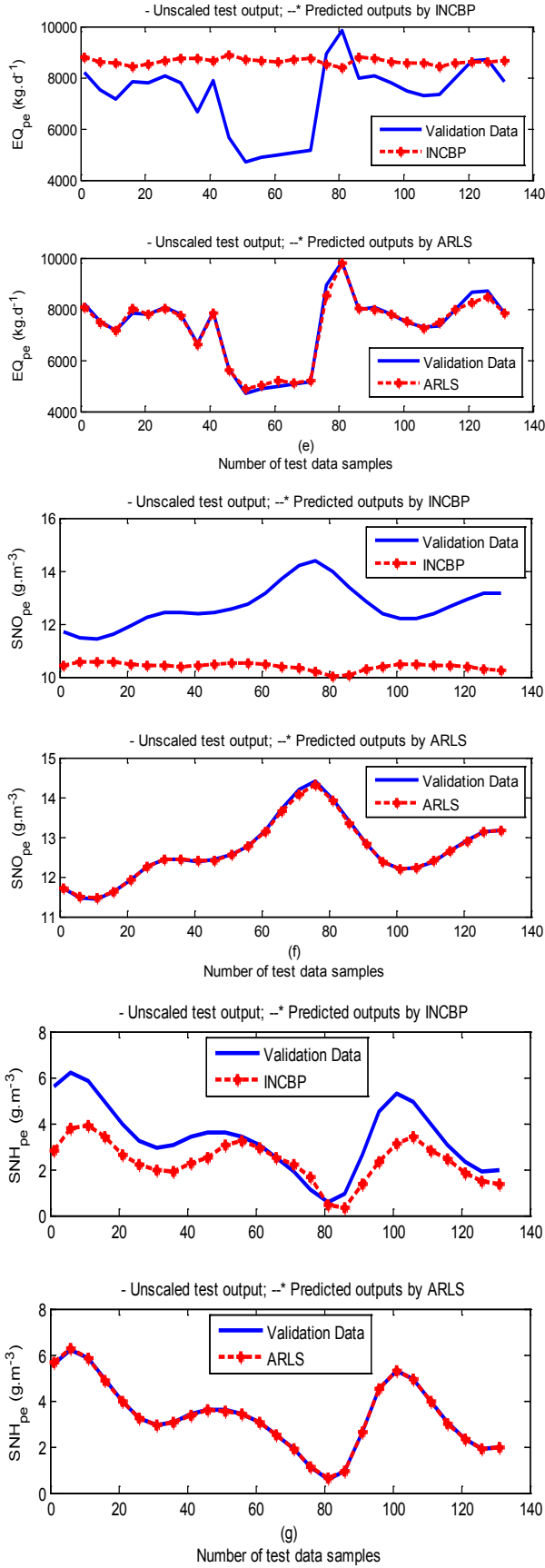**Figure 7(e).** One-step ahead prediction of scaled EQ_pe, SNO_pe and SNH_pe training data

Figure 8(a). One-step ahead prediction of unscaled Q_{F2}, Q_W, Q_{R2} and Q_{F3} validation data

Figure 8(b). One-step ahead prediction of unscaled Q_{E_ps}, EQ_ps, N_{total_ps} and TSS_ps validation data

Figure 8(c). One-step ahead prediction of unscaled COD_ps, BOD_ps, SNO_ps and SNH_ps validation data

Figure 8(d). One-step ahead prediction of unscaled TSS_pe, BOD_pe, COD_pe and N_total_pe validation data

Figure 8(e). One-step ahead prediction of unscaled EQ_pe, SNO_pe and SNH_pe validation data

Figure 9(a). Five-step ahead prediction of unscaled Q_{F2}, Q_W, Q_{R2} and Q_{F3} training data
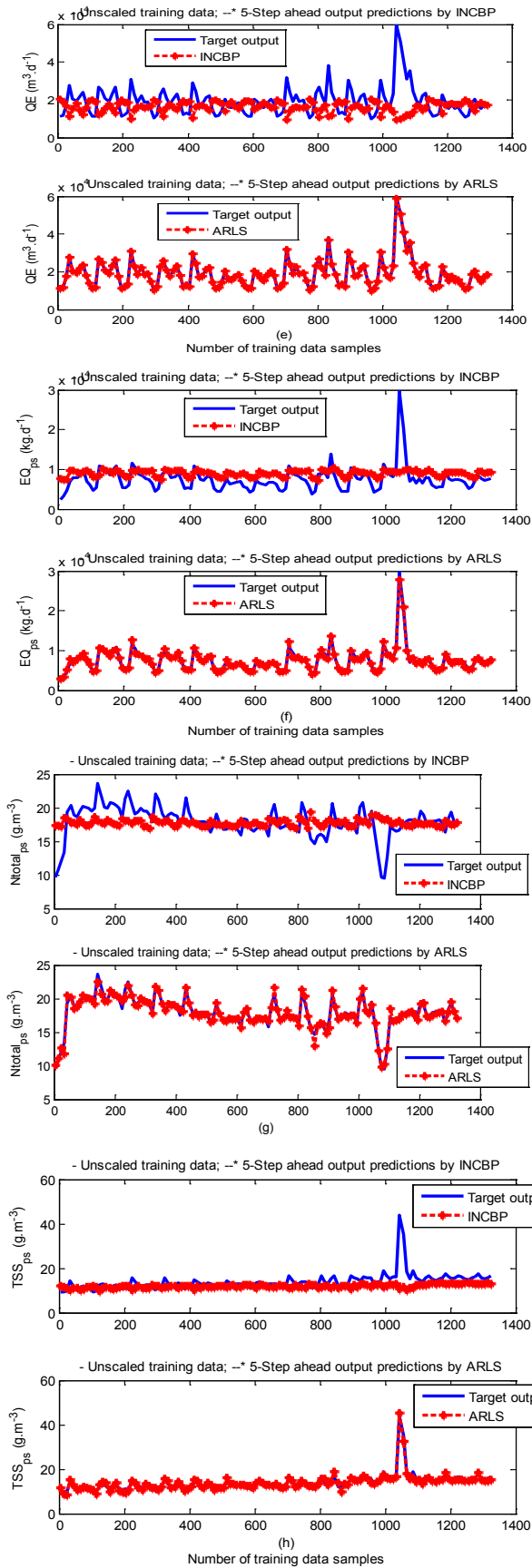
Figure 9(b). Five-step ahead prediction of unscaled Q_E_ps, EQ_ps, N_{total}_ps and TSS_ps training data
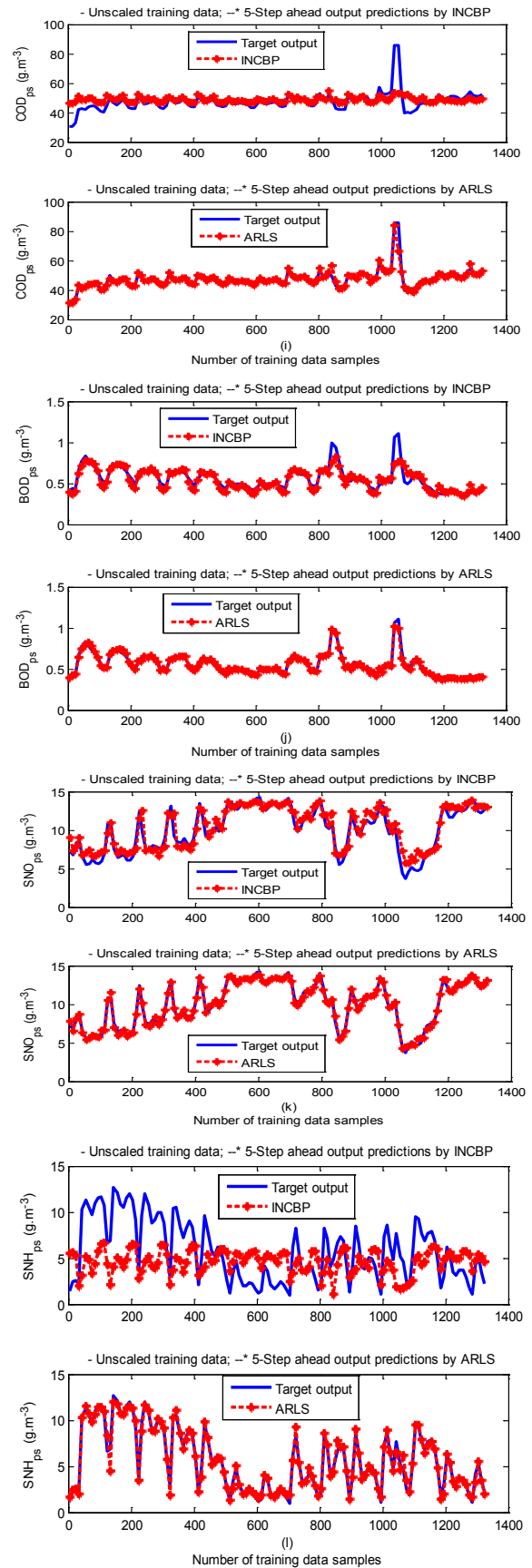
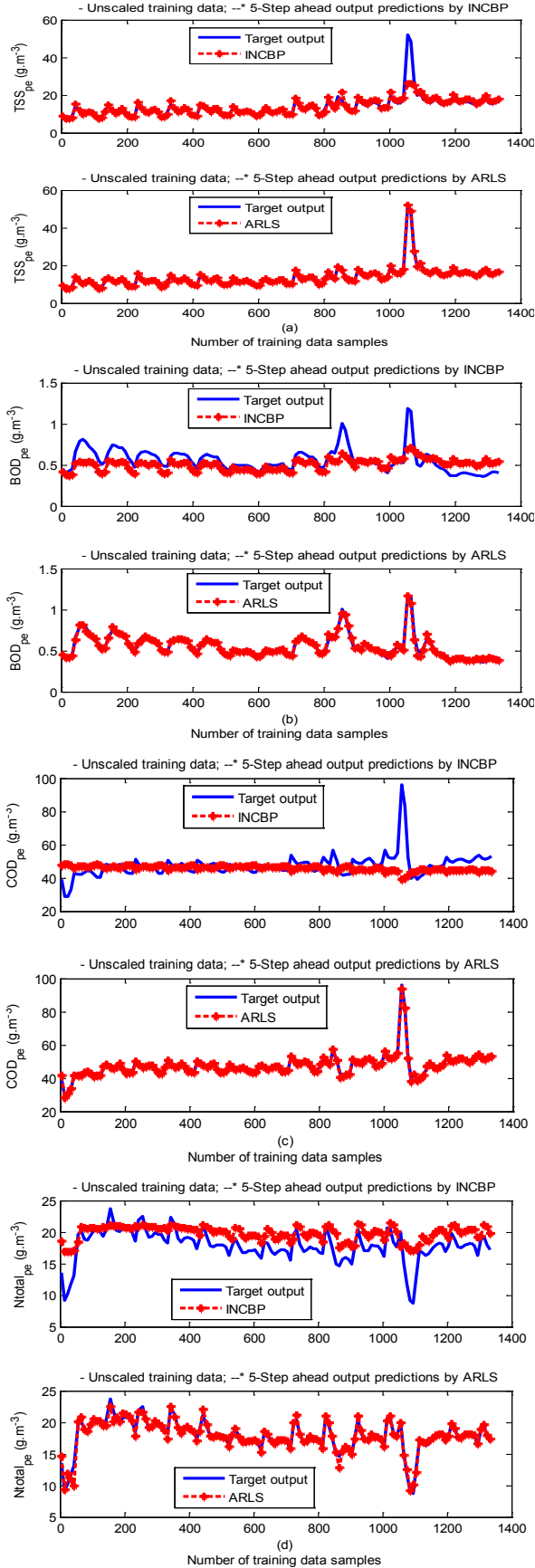**Figure 9(c).** Five-step ahead prediction of unscaled COD_ps, BOD_ps, SNO_ps and SNH_ps training data

Figure 9(d). Five-step ahead prediction of unscaled TSS_pe, BOD_pe, COD_pe and N_total_pe training data
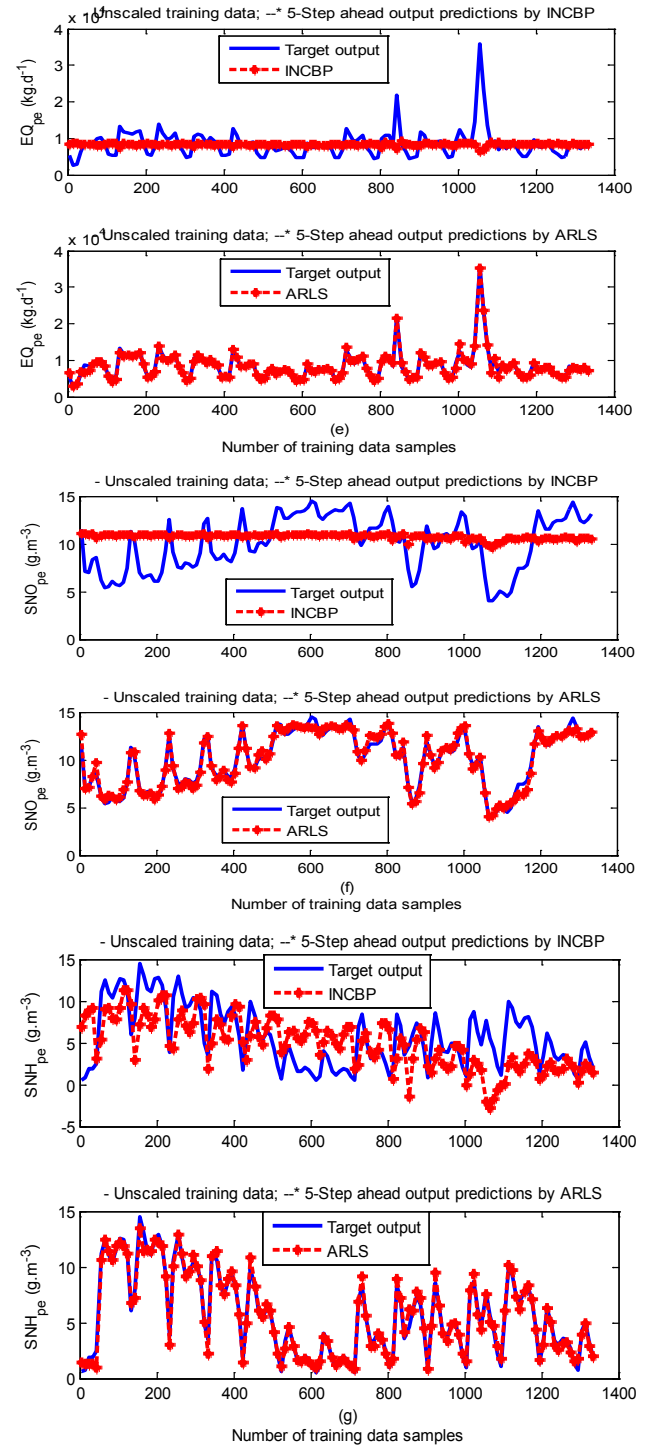

Figure 9(e). Five-step ahead prediction of unscaled EQ_pe, SNO_pe and SNH_pe training data

5.3.2. *K*–Step Ahead Prediction Simulations for the AS-WWTP Process

The results of the *K*-step ahead output predictions (red --*) using the *K*-step ahead prediction validation method discussed in Section 4.4 for 5-step ahead output predictions ($K = 5$) compared with the unscaled training data (blue -) are shown in Fig. 9(a) to Fig. 9(e) for the networks trained using the INCBP and the proposed ARLS. Again, the value $K = 5$

is chosen since it is a typical value used in most model predictive control (MPC) applications. The comparison of the 5-step ahead output predictions performance by the network trained using the INCBP and the proposed ARLS algorithms indicate the superiority of the proposed ARLS over the so-called INCBP algorithm.

The computation of the mean value of the K-step ahead prediction error (MVPE) using (27) is given in the sixth rows of Tables 4 and 5 by the network trained using INCBP and the proposed ARLS algorithms respectively. The small mean values of the 5-step ahead prediction error (MVPE) are indications that the trained network approximates the dynamics for the secondary settler/clarifier and effluent tank of the AS-WWTP process to a high degree of accuracy with the networks of both algorithms but with the network based on the ARLS algorithm giving much smaller distant prediction errors.

5.3.3. Akaike's Final Prediction Error (AFPE) Estimates

The implementation of the AFPE algorithm discussed in Section 4.4 and defined by (25) for the regularized criterion for the network trained using the INCBP and the proposed ARLS algorithms with multiple weight decay gives their respective AFPE estimates which are defined in the seventh rows of Tables 4 and 5 respectively. These relatively small values of the AFPE estimate indicate that the trained networks capture the underlying dynamics for the secondary settler/clarifier and effluent tank of the AS-WWTP and that the network is not over-trained [29]. This in turn implies that optimal network parameters have been selected including the weight decay parameters. Again, the results of the AFPE estimates computed for the networks trained using the proposed ARLS algorithm are much smaller when compared to those obtained using INCBP algorithm.

6. Conclusions

This paper presents the formulation of an advanced online nonlinear adaptive recursive least squares (ARLS) model identification algorithm based on artificial neural networks for the nonlinear model identification of a AS-ASWWTP process. The mathematical model of the process obtained from COST Actions 632 and 624 has been simulated in open-loop to generate the training data while the *First Day* dry–weather data, also provided by the COST Actions group, is used as the validation (test) data. In order to investigate the performance of the proposed ARLS algorithm, the incremental backpropagation (INCBP), which is also an online algorithm, is implemented and its performance compared with proposed ARLS. The results from the application of these algorithms to the modeling of the secondary settler and the clarifier units of the AS-WWTP biological reactors as well as the validation results show that the neural network-based ARLS outperforms the INCBP algorithm with much smaller predictions error and good tracking abilities with high degree of accuracy and that the

proposed ARLS model identification algorithm can be used for the AS-ASWWTP process in an industrial environment.

The next aspect of the work is on the development of an adaptive fuzzy rule-based logic decision system which would produce the set points that can be used for the development of an intelligent multivariable nonlinear adaptive model-based predictive control algorithm for the efficient control of the complete AS-WWTP by manipulating the pumps based on the decision parameters discussed in this work.

Appendix

Appendix A: Computation of the Decision Parameters and Food-to-Microorganism Ratio

```
% Compute the influent Quality Index
% Beta factor specification
BTSS = 2; BCOD = 1; BNK_j = 20; BNO = 20;  BBOD = 2;
% Compute the decision functions
SNO_inf  = SNO_inf_tr;
TKN_inf = SNH_inf_tr + SND_inf_tr + XND_inf_tr +
          i_XB*(XBH_inf_tr + XBA_inf_tr) +
          i_XP*(XP_inf_tr + XI_inf_tr);
TSS_inf = 0.75*(XS_inf_tr + XI_inf_tr + XBH_inf_tr +
                XBA_inf_tr + XP_inf_tr);
BOD_inf = 0.65*(SS_inf_tr + XS_inf_tr +
          0.94*(XBH_inf_tr + XBA_inf_tr));
COD_inf = SS_inf_tr + SI_inf_tr + XS_inf_tr + XI_inf_tr
          + XBH_inf_tr + XBA_inf_tr + XP_inf_tr;
Q_inf   = Qin_tr;
% Compute total nitrogen
Ntotal_inf = TKN_inf + SNO_inf;
% Compute influent quality
inf_qua = (BNK_j*TKN_inf + BTSS*TSS_inf +
          BCOD*COD_inf + BBOD*BOD_inf +
          BNO*SNO_inf).*Q_inf/1000;
% Compute the total sludge
tot_slu = TSS_r1_tr*VOL1 + TSS_r2_tr*VOL2 +
          TSS_r3_tr*VOL3 + TSS_r4_tr*VOL4 +
          TSS_r5_tr*VOL5;
% Compute food-to-microorganisms ratio
F_to_M = Qin_tr.*BOD_inf./tot_slu;
% Compute the amount of microorganisms (MLVSS)
MLVSS  = BOD_inf./F_to_M;
%=========================================
Startindex  = max(find(t <= starttime));
stopindex   = min(find(t >= stoptime));
time        = t(startindex:stopindex);
feedx       = feed(startindex:stopindex,:);
recx        = rec(startindex:stopindex,:);
settlerx    = settler(startindex:stopindex,:);
inx         = in(startindex:stopindex,:);
[n,m]       = size(reac1);
BODinfluent= 0.65.*(inx(n,2) + inx(n,4) +
             0.92.*(inx(n,5) + inx(n,6)));
```

```
sludge      = reac1(n,14)*VOL1 + reac2(n,14)*VOL2 +
              reac3(n,14)*VOL3 + reac4(n,14)*VOL4 +
              reac5(n,14)*VOL5;
F_to_M      = inx(n,15).*BODinfluent/sludge
```

Appendix B: General Characteristics of the Secondary Settler and Clarifier

Appendix B.1: Constructional Features

Table B.1. The double-exponential settling velocity function parameters with their definition, units and values

Parameters Description	Parameters Symbol	Value	Units
Maximum settling velocity	v_0'	250.0	m.(day)$^{-1}$
Maximum Vesilind settling velocity	v_0	474	m.(day)$^{-1}$
Hindered zone settling parameter	r_h	0.000576	m^3.(g SS)$^{-1}$
Flocculant zone settling parameter	r_p	0.00286	m^3.(g SS)$^{-1}$
Non-settleable fraction	f_{ns}	0.00228	Dimensionless

The secondary settler and clarifier is modelled as a ten layers non-reactive unit, that is, no biological reactions takes place in the secondary settler. The 6th layer, counting from the top, is the feed layer as can be seen in Fig. 1. The settler has an area (A) of $1,500\,m^2$. The height (z_m) of each layer (m) is $0.4\,m$, for a total height of $4\,m$. Therefore, the settler has total volume equal to $6,000\,m^3$.

The solid flux due to gravity is

$$J_s = v_s(X)X \tag{B.1}$$

where X is the total sludge concentration and $v_s(X)$ is a double-exponential settling velocity function defined as:

$$v_s(X) =$$
$$\max\left[0, \min\left\{v_0', \; v_0\left(e^{r_h(X-X_{\min})} - e^{r_p(X-X_{\min})}\right)\right\}\right] \tag{B.2}$$

where $X_{\min} = f_{ns}X_f$. The parameter values for the non-exponential settling velocity function (B.2) are given in Table B.1. Thus, the mass balances for the sludge are expressed as:

For the feed layer $(m = 6)$:

$$\frac{dX_m}{dt} =$$
$$\frac{\frac{Q_f X_f}{A} + J_{clar,m+1} - \left(v_{up} - v_{dn}\right)X_m - \min\left(J_{s,m}, J_{s,m-1}\right)}{z_m} \tag{B.3}$$

For the intermediate layers below the feed layer $(m = 2)$ to $(m = 5)$:

$$\frac{dX_m}{dt}$$
$$= \frac{v_{dn}\left(X_{m+1} - X_m\right) + \min\left(J_{s,m}, J_{s,m+1}\right) - \min\left(J_{s,m}, J_{s,m-1}\right)}{z_m} \tag{B.4}$$

For the bottom layer $(m = 1)$:

$$\frac{dX_1}{dt} = \frac{v_{dn}\left(X_2 - X_1\right) + \min\left(J_{s,2}, J_{s,1}\right)}{z_1} \tag{B.5}$$

For the intermediate clarification layers above the feed layer $(m = 7)$ to $(m = 9)$:

$$\left.\begin{aligned} \frac{dX_m}{dt} &= \frac{v_{up}\left(X_{m-1} - X_m\right) + J_{clar,m+1} - J_{clar,m}}{z_m} \\[6pt] J_{clar,j} &= \begin{cases} \min\left(v_{s,j}X_j, v_{s,j-1}X_{j-1}\right), & \text{if } X_{j-1} > X_t \\ v_{s,j}X_j, & \text{if } X_{j-1} \le X_t \end{cases} \end{aligned}\right\} \tag{B.6}$$

For the top layer $(m = 10)$:

$$\left.\begin{aligned} \frac{dX_{10}}{dt} &= \frac{v_{up}\left(X_9 - X_{10}\right) - J_{clar,10}}{z_{10}} \\[6pt] J_{clar,10} &= \begin{cases} \min\left(v_{s,10}X_{10}, v_{s,9}X_9\right), & \text{if } X_9 > X_t \\ v_{s,10}X_{10}, & \text{if } X_9 \le X_t \end{cases} \end{aligned}\right\} \tag{B.7}$$

The threshold concentration $X_t = 3,000\,g.m^{-3}$.

For the soluble components (including dissolved oxygen), each layer represents a completely mixed volume and the concentrations of soluble components are given accordingly as:

For the feed layer $(m = 6)$:

$$\frac{dZ_m}{dt} = \frac{\frac{Q_f X_f}{A} - \left(v_{dn} + v_{up}\right)Z_m}{z_m} \tag{B.8}$$

For the intermediate layers below the feed layer $(m = 1)$ to $(m = 5)$:

$$\frac{dZ_m}{dt} = \frac{v_{dn}\left(Z_{m+1} - Z_m\right)}{z_m} \tag{B.9}$$

For the intermediate layers above the feed layer $(m = 7)$ to $(m = 10)$:

$$\frac{dZ_m}{dt} = \frac{v_{up}\left(Z_{m-1} - X_m\right)}{z_m} \tag{B.10}$$

where

$$v_{dn} = \frac{Q_u}{A} = \frac{Q_r + Q_w}{A} \quad and \quad v_{up} = \frac{Q_e}{A} \tag{B.11}$$

The concentrations in the recycle and waste flow are equal to those of the first layer $(m=1)$, that is, $Z_u = Z_1$.

The sludge concentration from the concentrations in Unit 5 of Fig .1 can be computed from:

$$\left.\begin{aligned} X_f &= \frac{1}{fr_{COD-SS}}\left(X_{S,5} + X_{P,5} + X_{I,5} + X_{B,H,5} + X_{B,A,5}\right) \\ &= 0.75\left(X_{S,5} + X_{P,5} + X_{I,5} + X_{B,H,5} + X_{B,A,5}\right) \end{aligned}\right\} \text{(B.12)}$$

Since $fr_{COD-SS} = 4/3 = 0.75$. The same calculation is applied for X_u in the settler underflow and X_e at the plant exit.

To calculate the distribution of particulate concentrations in the recycled and waste flows, their ratios with respect to the total solid concentration are assumed to remain constant across the settler:

$$\frac{X_{S,5}}{X_f} = \frac{X_{S,u}}{X_u} \quad \text{and} \quad X_{S,u} = \frac{X_{S,5}}{X_f}X_u \qquad \text{(B.13)}$$

Equation (B.13) is also true for $X_{P,u}$, $X_{I,u}$, $X_{B,H,u}$, $X_{B,A,u}$ and $X_{ND,u}$. The assumption made here means that the dynamics of the fractions of particulate concentrations in the inlet of the settler will be directly propagated to the settler underflow, without taking into account the normal retention time in the settler [1-4].

Appendix B.2: Sludge Age

Appendix B.2.1: Sludge Age Based on Total Amount of Biomass

In the steady-state case, the sludge age calculation is based on the total amount of biomass present in the system that is in the reactor and settler:

$$Sludge_{Age} = \frac{TX_a + TX_s}{\phi_e + \phi_w} \qquad \text{(B.14)}$$

where TX_a is the total amount of biomass present in the reactor and it is expressed as:

$$TX_a = \sum_{i=1}^{n}\left(X_{B,H,i} + X_{B,A,i}\right)V_i, \qquad n=5 \quad \text{(B.15)}$$

TX_s is the total amount of biomass present in the effluent and it is expressed as:

$$TX_s = \sum_{j=1}^{m}\left(X_{B,H,i} + X_{B,A,i}\right)z_j A, \qquad m=10 \quad \text{(B.16)}$$

ϕ_e is the loss rate of biomass in the effluent and it is expressed as:

$$\phi_e = \left(X_{B,H,m} + X_{B,A,m}\right)Q_e, \qquad m=10 \quad \text{(B.17)}$$

ϕ_w is the loss rate of biomass in the waste flow and it is expressed as:

$$\phi_w = \left(X_{B,H,u} + X_{B,A,u}\right)Q_w \qquad \text{(B.18)}$$

Appendix B.2.2 Sludge Age Based on Total Amount of Solids

In an actual AS-WWTP, the sludge age is measured based on the total amount of solid present in the system. Thus,

$$Sludge_{Age} = \frac{TX_{fa} + TX_{fs}}{\phi_{fe} + \phi_{fw}} \qquad \text{(B.19)}$$

where TX_{fa} is the total amount of solids present in the reactor and can be expressed as:

$$TX_{fa} = \sum_{i=1}^{n} X_{f,i}V_i, \qquad n=5 \qquad \text{(B.20)}$$

where

$$X_{f,i} = \frac{1}{fr_{COD-SS}}\left(X_{S,i} + X_{P,i} + X_{I,i} + X_{B,H,i} + X_{B,A,i}\right) \text{(B.21)}$$

TX_{fs} is the total amount of solids present in the settler and can be expressed as:

$$TX_{fs} = \sum_{j=1}^{m} X_{f,i}z_j A, \qquad m=10 \quad \text{(B.22)}$$

where

$$X_{f,j} = \frac{1}{fr_{COD-SS}}\left(X_{S,j} + X_{P,j} + X_{I,j} + X_{B,H,j} + X_{B,A,j}\right) \text{(B.23)}$$

ϕ_{fe} is the loss rate of solids in the effluent and can be expressed as:

$$\phi_{fe} = X_{f,m}Q_e, \qquad m=10 \qquad \text{(B.24)}$$

where

$$X_{f,m} = \frac{1}{fr_{COD-SS}}\begin{pmatrix} X_{S,m} + X_{P,m} + X_{I,m} \\ + X_{B,H,m} + X_{B,A,m} \end{pmatrix} \text{(B.25)}$$

ϕ_w is the loss rate of solids in the waste flow and can be expressed as:

$$\phi_{fw} = X_{f,u}Q_w \qquad \text{(B.26)}$$

where $X_{f,u} = \frac{1}{fr_{COD-SS}}\begin{pmatrix} X_{S,u} + X_{P,u} + X_{I,u} \\ + X_{B,H,u} + X_{B,A,u} \end{pmatrix}$

Appendix C: Criteria for Evaluating and Assessing the Performances of the AS-WWTP Control

Appendix C.1: Influent Quality (IQ)

As a check on the IQ calculation, an influent quality index (IQ) can be calculated by applying the above equations to the influent file but the BOD coefficient must be changed from 0.25 to 0.65. It is defined as:

$$I.Q. = \frac{1}{T \cdot 1000} \int\limits_{t=7days}^{t=14days} \begin{pmatrix} B_{SS} \cdot SS_0(t) + B_{COD} \cdot COD_0(t) \\ + B_{NKj} \cdot S_{NKj,0}(t) + B_{NO} \cdot S_{NO,0}(t) \\ + B_{BOD} \cdot BOD_{5,0}(t) \end{pmatrix} \quad (C.1)$$
$$\times 3 * Q_{in\_tr}(t)dt$$

where the composition variables are calculated as follows:

$$S_{NKj,0} = S_{NH,0} + S_{ND,0} + X_{ND,0} + i_{XB}(X_{B,H,0} + X_{B,A,0})$$
$$+ i_{XP}(X_{P,0} + X_{i,0})$$

$$SS_0 = 0.75 \cdot \left(X_{S,0} + X_{I,0} + X_{B,H,0} + X_{B,A,0} + X_{P,0} \right)$$

$$BOD_0 = 0.65 \cdot \left(S_{S,0} + X_{S,0} + (1 - f_P) \cdot (X_{B,H,0} + X_{B,A,0}) \right)$$

$$COD_0 = S_{S,0} + S_{I,0} + X_{S,0} + X_{I,0} + X_{B,A,0} + X_{B,A,0} + X_{P,0}$$

Appendix C.2: Effluent Quality and Constraints

Appendix C.2.1: Effluent Quality (EQ)

The effluent quality (EQ), in kg pollution unit/d, is averaged over the period of observation $T(d)$ (i.e. the second week or 7 last days for each weather file) based on a weighting of the effluent loads of compounds that have a major influence on the quality of the receiving water and that are usually included in regional legislation. It is defined as [2]–[4]:

$$E.Q. =$$

$$\frac{1}{T \cdot 1000} \int\limits_{t=7days}^{t=14days} \begin{pmatrix} B_{TSS} \cdot TSS_e(t) + B_{COD} \cdot COD_e(t) \\ + B_{NKj} \cdot S_{NKj,e}(t) + B_{NO} \cdot S_{NO,e}(t) \\ + B_{BOD} \cdot BOD_{5,e}(t) \end{pmatrix} \cdot Q_e(t)dt \quad (C.2)$$

where the composition variables are calculated according to the following equations:

$$S_{NKj,e} = S_{NH,e} + S_{ND,e} + X_{ND,e} + i_{XB}(X_{B,H,e} + X_{B,A,e})$$
$$+ i_{XP}(X_{P,e} + X_{I,e})$$

$$TSS_e = 0.75 \cdot \left(X_{S,e} + X_{I,e} + X_{B,H,e} + X_{B,A,e} + X_{P,e} \right)$$

$$BOD_e = 0.25 \cdot \left(S_{S,e} + X_{S,e} + (1 - f_P) \cdot (X_{B,H,e} + X_{B,A,e}) \right)$$

$$COD_e = S_{S,e} + S_{I,e} + X_{S,e} + X_{I,e} + X_{B,H,e} + X_{B,A,e} + X_{P,e}$$

$$Ntotal_e = S_{NKj,e} + S_{NO,e}$$

where B_i are weighting factors for the different types of pollution to convert them into pollution units (g/m^3) and were chosen to reflect these calculated fractions as follows: $B_{TSS} = 2$, $B_{COD} = 1$, $B_{NKj} = 20$, $B_{NO} = 20$ and $B_{BOD5} = 2$.

The major operating cost in biological nutrient removal process as well as nitrogen removing ASPs is blower energy consumption. If the DO set-point is reduced by a control strategy, significant energy saving can be achieved. Operational issues are considered through three items: sludge production, pumping energy and aeration energy (integrations performed on the final 7 days of weather

simulations (i.e. from day 22 to day 28 of weather file simulations, $T = 7 days$).

Appendix C.2.2: Constraints on the Effluent Quality

The flow average values of the effluent concentrations over the three test periods (dry, rain and storm weather: 7 days for each) should be constrained for the five effluent components within the following limit: total nitrogen ($N_{total} = S_{NO,e} + S_{NKj,e} < 18 g \cdot m^{-3}$),   total   COD ($COD_e < 100 g \cdot m^{-3}$),   ammonia   ($S_{NH,e} < 4 g \cdot m^{-3}$), suspended solids   ($TSS_e < 30 g \cdot m^{-3}$)   and BOD_5 ($BOD_e < 10 g \cdot m^{-3}$).

Appendix C.3: The Sludge production to be disposed (kg/d)

This is the sludge production, P_{sludge}, is calculated from the total solid flow from wastage and solid accumulated in the system over the period of time considered ($t_f = 7$ for each weather file). The amount of solids in the system at time t is given by:

$$M(TSS_{system})(t) = M(TSS_{reactor})(t) + M(TSS_{ssttler})(t) \quad (C.3)$$

where $M(TSS_{reactor})(t)$ is the amount of solids in the reactor given by:

$$M(TSS_{reactor})(t) = 0.75 \cdot \sum_{i=1}^{i=5} \begin{pmatrix} X_{S,i} + X_{I,i} + X_{B,H,i} \\ + X_{B,A,i} + X_{P,i} \end{pmatrix} \cdot V_i \quad (C.4)$$

$M(TSS_{settler})(t)$ is the amount of solids in the settler given by:

$$M(TSS_{settler})(t) = 0.75 \sum_{j=1}^{j=7} \begin{pmatrix} X_{S,j} + X_{I,j} + X_{B,H,j} \\ + X_{B,A,j} + X_{P,j} \end{pmatrix} z_i A \quad (C.5)$$

$\Delta M(TSS)_{system}$ the change in system sludge mass from the end of day 7 to the end of day 14 given by:

$$\Delta M(TSS)_{system} = M(TSS_{system})_{14days} - M(TSS_{system})_{7days}$$

and $M(TSS_{watse})(t)$ the amount of waste sludge is given by:

$$M(TSS_{watse})(t)$$
$$= 0.75 \int\limits_{t=7days}^{t=14days} \begin{pmatrix} X_{S,w} + X_{I,w} + X_{B,H,w} \\ + X_{B,A,w} + X_{P,w} \end{pmatrix} \cdot Q_w(t)dt \quad (C.6)$$

So that the total sludge to be disposed becomes:

$$P_{sludge} = \frac{1}{T} \left(\Delta M(TSS_{system})(t) + M(TSS_{watse})(t) \right) \quad (C.7)$$

Appendix C.4: The Total Sludge production (kg/d)

The total sludge production takes into account the sludge to be disposed and the sludge lost to the weir and is calculated as follows:

$$P_{total\_sludge} = P_{sludge} + M(TSS_e) \quad (C.8)$$

where

$$M(TSS_e) = \frac{0.75}{T} \int_{t=7days}^{t=14days} \begin{pmatrix} X_{S,e} + X_{I,e} + X_{B,H,e} \\ + X_{B,A,e} + X_{P,e} \end{pmatrix} \cdot Q_e(t) dt$$

Appendix C.5: The Pumping Energy (PE)

The pumping energy in kWh/d is calculated as follows:

$$PE = \frac{0.04}{T} \int_{t=7days}^{t=14days} \left(Q_a(t) + Q_r(t) + Q_w(t) \right) \cdot dt \quad (C.9)$$

where $Q_a(t)$ is internal recycle flow rate at time $t (m^3 d^{-1})$, $Q_r(t)$ is the return sludge recycle flow rate at time $t (m^3 d^{-1})$, and $Q_w(t)$ is the waste sludge flow rare at time $t (m^3 d^{-1})$.

Appendix C.6: The Aeration Energy (AE)

The aeration energy (AE) in kWh/d takes into account the plant peculiarities (type of diffuser, bubble size, depth of submersion, etc,) and is calculated from the $k_l a$ in the three aerated tanks according to the following relation, valid for Degrémont DP230 porous disks at an immersion depth of 4m:

$$AE = \frac{24}{T} \int_{t=7days}^{t=14days} \sum_{i=3}^{i=5} \begin{pmatrix} 0.4032 \left(k_l a\right)_i^2 \\ + 7.4408 \left(k_l a\right)_i \end{pmatrix} \cdot dt \quad (C.10)$$

where $k_l a$ is in h^{-1} and where i is the compartment number.

The increase in capacity which could be obtained using the proposed control strategy should be evaluated. This factor is relative to investment costs if the plant would simply be extended to deal with increased load. This is expressed by the relative increase in the influent flow rate, a, which can be applied while maintaining the reference effluent quality index (EQ_{ref}) for the three weather conditions ($T = 7$ days for each). EQ_{ref} is calculated from the above equation in open loop. $Q_{0,i}(t) = a * Q_{0,i}(t)$ with $i = 1$ for dry weather, $i = 2$ for storm weather and $i = 3$ for rain weather. Operation variables such as Q_w, Q_r and $k_L a$ in compartments 3 and 4 remains unchanged.

REFERENCES

[1] M. Henze, P. Harremoës, J. Jansen and E. Arvin, "Wastewater Treatment," *Biological and Chemical Processes*, 2nd ed., Berlin: Springer Verlag, 1996.

[2] J.B. Coop. (2000, Sept.). The COST Simulation Benchmark: Description and Simulation Manual (a product of COST Actions 624 & 628). [Online]. Available: http://www.ensic. inpl-nancy.fr/COSTWWTP/.

[3] Working Groups of COST Actions 632 and 624. (2000, Sept.). IWA Task Group on Benchmarking of Control Strategies for WWTPs. [Online]. Available: http://www.ensic-inpl-nancy.f r/benchmarkWWTP/Bsm1/Benchmark1.htm.

[4] Working Groups of COST Actions 632 and 624. (Apr., 2008). IWA Task Group on Benchmarking of Control Strategies for WWTPs: http://www.ensic.inplnancy.fr/benchmarkWWTP/ Bsm1/Benchmark1.htm.

[5] W. Zarrad, J. Harmand, M. Devisscher and J.P. Steyer, "Comparison of advanced control strategies for improving the monitoring of activated sludge processes," *Control Engineering Practice*, vol. 12, pp. 323-333, 2004.

[6] K.V. Gernaey, M.C. van Loosdrecht, M. Henze, M. Lind and S.B. Jørgensen, "Activated sludge wastewater treatment plant modeling and simulation: state of the art," *Environmental Modelling & Software*, vol. 19, pp. 763-783, 2004.

[7] R. Piotrowski, M.A. Brdys, K. Konarczak, K. Duzinkiewic and W. Chotkowski, "Hierarchical dissolved oxygen control for activated sludge processes", *Control Engineering Practice*, vol. 16, pp. 114-131, 2008.

[8] T.T. Lee, F.Y. Wang, A. Islam and R.B. Newell, "Generic distributed parameter model control of a biological nutrient removal (BNR) activated sludge process," *Journal of Process Control*, vol. 9, pp. 505-525, 1999.

[9] T.T. Lee, F.Y. Wang and R.B. Newell, "Advances in distributed parameter approach to the dynamics and control of activated sludge processes for wastewater treatment", *Water Research*, vol. 40, pp. 853-869, 2006.

[10] I.Y. Smets, J.V. Haegebaert, R. Carrette and J.F. Van Impe, "Linearization of the activated sludge model ASM1 for fast and reliable predictions", *Water Research*, vol. 37, pp. 1831-1851, 2003.

[11] J. Wanner, Activated sludge bulking and foaming control. Lancaster: Technomic Publishing Company, 1994.

[12] G.A. Ekama, J.L. Barnard, F.W. Gunthert, P. Krebs, J.A. McCorquodale and D.S. Parker, "Secondary settling tanks: theory, modeling, design and operation," International Association on Water Quality, London, UK, Sci. and Tech. Rep., No. 6, 1997.

[13] A. Vanderhasselt and P.A. Vanrolleghem, "Estimation of sludge sedimentation parameters from single batch settling curves," *Water Research,* vol. 34, no. 2, pp.395-406, 2000.

[14] A. Traoré, S. Grieu, F. Thiery, M. Polit and J. Colprim, "Control of sludge height in a secondary settler using fuzzy algorithms," *Computers & Chemical Engineering*, vol. 30, pp. 1235-1242, 2006.

[15] M. A. Steffens and P. A. Lant, "Multivariable control of nutrient-removing activated sludge systems", *Water Research*, vol. 33, no. 12, pp. 2864 – 2878, 1999.

[16] T. T. Lee, F. Y. Wang and R. B. Newell, "Robust multivariable control of complex biological process", *Journal of Process Control*, vol. 14, pp. 193 – 209, 2004

[17] A. Stare, N. Hvala and D. Vrecko, "Modeling, identification, and validation of models for predictive control ammonia control in a wastewater treatment plant – A case study", *ISA Transactions*, vol. 45, no. 2, pp. 159 – 174, 2006.

[18] M. Ekman, B. Björlenius and M. Andersson, "Control of the aeration volume in an activated sludge process using supervisory control strategies", *Water Research*, vol. 40, pp. 1668 – 1676, 2006.

[19] A. Stare, D. Vrečko, H. Hvala and S. Strmčnik, "Comparison of control strategies for nitrogen removal in an activated sludge process in terms of operating costs: A simulation study", *Water Reseach*, vol. 41, pp. 2004 – 2014, 2007.

[20] N.P. Cheremisinoff, *Biotechnology for Waste and Wastewater Treatment*. New Jersey, USA: Noyes Publications, 1996.

[21] F.R. Spellman, *Handbook of Water and Wastewater Treatment Plant Operations*. Boca Raton, Florida: CRC Press LLC, 2003.

[22] V. A. Akpan and R. A. O. Osakwe, "Multivariable NNARMAX model identification of an AS-WWTP using ARLS: Part 1 – Dynamic modeling of the biological reactors", American Journal of Intelligent Systems, vol. x, issue xx, 26 pages, 2014. (Submitted)

[23] V. A. Akpan (Jul., 2011): *Development of new model adaptive predictive control algorithms and their implementation on real-time embedded systems*, Ph.D. Dissertation, 517 pages. [Online] Available: http://invenio. lib.auth.gr/record/127274/files/GRI-2011-7292.pdf.

[24] G. C. Goodwin, K.S. Sin, *Adaptive Filtering, Prediction and Control*, Prentice-Hall, 1984.

[25] L. Lung, *System Identification: Theory for the User*. 2nd ed., Prentice-Hall, 1999.

[26] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks". *IEEE Trans. Neural Networks,* vol. 1, no. 1, pp. 4 – 27, 1990.

[27] M. Nørgaard, O. Ravn, N. K. Poulsen, and L.K. Hansen, "*Neural Networks for Modelling and Control of Dynamic Systems: A Practitioner's Handbook"*. London: Springer-Verlag, 2000.

[28] R. Chiong, *Intelligent Systems for Automated Learning and Adaptation: Emerging Trends and Applications*, Information Science Reference, 2010, chapter 4.

[29] J. Sjöberg and L. Ljung, Overtraining, regularization, and searching for minimum in neural networks, *International Journal of Control,* vol. 62, 1995, pp.: 1391-1408.

[30] V. A. Akpan and G. D. Hassapis, "Training dynamic feedforward neural networks for online nonlinear model identification and control applications". *International Reviews of Automatic Control: Theory & Applications*, vol. 4, no. 3, pp. 335 – 350, 2011.

[31] M. Salgado, G. Goodwin, R. Middleton, Modified least squares algorithm incorporating exponential forgetting and resetting, *Int. J. Control*, vol. 47, no. 2, pp. 477-491, 1988.

[32] J. Hertz, A. Krough and R. G. Palmer, "*An Introduction to the Theory of Neural Computation*", Lecture Notes, vol. 1, Redwood City, California: Addison-Wesley, 1991.